

PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK PENGELOMPOKKAN SAHAM BERDASARKAN *PRICE EARNING RATIO* DAN *PRICE TO BOOK VALUE*

Derrick Tohendry^a, Deny Jollyta^b

^a Institut Bisnis dan Teknologi Pelita Indonesia, derrick@student.pelitaindonesia.ac.id

^b Institut Bisnis dan Teknologi Pelita Indonesia, deny.jollyta@lecturer.pelitaindonesia.ac.id

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 9 April 2023

Revisi Akhir: 30 April 2023

Diterbitkan Online: 31 April 2023

KATA KUNCI

Saham, PER, PBV, *K-Means*, *Elbow*

KORESPONDENSI

E-mail: deny.jollyta@lecturer.pelitaindonesia.ac.id

ABSTRACT

Saham merupakan salah satu instrument investasi yang memiliki risiko tinggi namun tingkat imbal hasil yang didapatkan cukup tinggi. Banyak masyarakat yang mulai melakukan investasi ke saham. Kebanyakan orang membeli saham tanpa melakukan analisa apapun sehingga banyak terjadi setelah dibeli harga saham malah turun maupun ketika menjual saham harganya malah naik. Untuk meminimalisir risiko ini maka diperlukan pengelompokan saham berdasarkan rasio keuangan *Price Earning Ratio* (PER) dan *Price to Book Value* (PBV). Rasio PER dan PBV sering digunakan oleh investor untuk menilai sebuah perusahaan murah atau mahal. Data yang akan digunakan pada penelitian ini diambil dari broker mirae asset sekuritas. Pengelompokan ini akan menggunakan algoritma *K-Means*. Selain itu metode *Elbow* juga akan digunakan untuk mengevaluasi jumlah *cluster* yang akan dipakai pada *K-Means*. Algoritma *K-Means* dipilih karena cukup mudah untuk dipahami dan sangat umum digunakan sebagai teknik *clustering* atau pengelompokan. Algoritma *K-Means* yang digunakan akan ditulis dengan Bahasa pemrograman *Python* melalui *Google Colaboratory*. Hasil dari Penelitian ini adalah 3 kelompok saham yang terdiri dari kelompok saham murah, kelompok saham sedang, dan kelompok saham mahal. Kelompok saham murah cenderung memiliki risiko yang rendah, kelompok saham sedang cenderung memiliki risiko yang sedang, dan kelompok saham mahal cenderung memiliki risiko yang tinggi.

1. PENDAHULUAN

Teknologi saat ini telah menjadi bagian besar dari kehidupan manusia. Dengan berkembangnya teknologi, seluruh kegiatan manusia sangat bergantung dengan teknologi mulai dari belajar mengajar hingga berinvestasi. Investasi merupakan penundaan konsumsi sekarang untuk dimasukkan ke aktiva produktif selama periode waktu yang tertentu. Terdapat berbagai produk investasi yang dilegalkan di Indonesia yaitu saham, reksadana, dan obligasi. Saham termasuk instrument yang *high risk* dan *high gain*. Saham diperdagangkan di Bursa Efek Indonesia. Terdapat ratusan perusahaan yang dapat diperdagangkan oleh pelaku pasar

Dalam menilai sebuah perusahaan, investor dibantu oleh rasio keuangan. Rasio keuangan yang digunakan pada Penelitian ini adalah *Price Earning Ratio* (PER) dan *Price to Book Value* (PBV). Rasio ini sangat sering digunakan untuk menilai sebuah saham tersebut murah atau mahal.

Dalam menentukan keputusan investasi saham, investor pemula cenderung memilih saham secara acak dan tidak melakukan analisa yang menyebabkan ketika saham tersebut dibeli harganya turun dan ketika dijual harganya naik. Salah satu cara untuk menyelesaikan permasalahan ini adalah dengan melakukan pengelompokan saham yang mempunyai nilai murah

atau mahal dengan menggunakan teknik *clustering*. Hasil dari pengelompokan data ini berupa beberapa kelompok yang terdiri dari kelompok saham dengan nilai murah, sedang, dan mahal.

Dalam penelitian ini menggunakan algoritma *K-Means*. Algoritma *K-Means* merupakan salah satu dari metode pengelompokan data nonhierarki (sekatan) yang dapat mempartisi data kedalam bentuk dua kelompok ataupun lebih. Algoritma *K-Means* adalah algoritma terbaik, sederhana dan lebih cepat dibandingkan dengan algoritma *clustering* lainnya, karena itu algoritma *K-Means* lebih cocok dan sesuai untuk mengekstraksi data pasar modal. [1]

Penelitian yang menggunakan algoritma *K-Means* dalam bidang obat-obatan pada rumah sakit Pekanbaru telah diteliti oleh Gustientiedina[2], penelitian ini mempunyai tujuan untuk mengetahui klasterisasi data obat pada rumah sakit Pekanbaru. Data yang digunakan untuk penelitian ini berasal dari Laporan Pemakaian dan Lembar Permintaan Obat (LPLPO) RSUD Arifin Achmad Pekanbaru tahun 2017 dengan data atribut yaitu nama obat dan jumlah pemakaian. Penelitian ini menghasilkan 3 kelompok data. *Cluster* obat pemakaian sedikit (C1), *cluster* obat pemakaian sedang (C2), dan *Cluster* obat pemakaian tinggi (C3). Iterasi yang dilakukan sebanyak 4 kali.

2. TINJAUAN PUSTAKA

1. Saham

Saham merupakan selembar kertas yang menunjukkan hak pemegang saham tersebut yaitu pemodal atau pemilik kertas tersebut untuk mendapatkan prospek maupun keuntungan dari organisasi yang menerbitkan sekuritas tersebut. [3] Saham juga dapat diartikan sebagai bukti dari penyertaan dana baik dari masyarakat maupun institusi pada sebuah perusahaan yang membutuhkan modal. Dengan bukti tersebut pemegang saham berhak atas aset-aset perusahaan.

Dalam melakukan investasi saham terdapat keuntungan dan kerugian yang dapat diperoleh. Keuntungan yang bisa diperoleh investor yaitu : *dividen* dan *capital gain*. *Dividen* merupakan pembagian keuntungan yang diberikan perusahaan dan berasal dari keuntungan yang dihasilkan perusahaan. Sedangkan *capital gain* merupakan selisih antara harga beli dan jual saham dimana harga jual lebih tinggi dari harga beli. Lalu kerugian yang dapat diperoleh yaitu *capital loss* dan likuidasi. *Capital loss* merupakan selisih antara harga beli dan jual saham dimana harga beli lebih tinggi dari harga jual. Kemudian likuidasi merupakan kondisi ketika perusahaan dinyatakan bangkrut oleh pengadilan dan semua *asset* dijual untuk membayar kewajiban dari perusahaan tersebut.

2. Price Earning Ratio (PER)

Price Earning Ratio menjelaskan valuasi perusahaan dimana harga saham dibandingkan dengan laba per lembar saham. Semakin tinggi PER, maka semakin mahal juga valuasi perusahaan tersebut. Begitu juga sebaliknya semakin rendah PER, maka semakin murah valuasi perusahaan tersebut. PER juga mencerminkan perkembangan (*growth*) saham tersebut. [4] *Price Earning Ratio* (PER) juga dapat diartikan sebagai rasio yang membandingkan harga saham dengan *earning* perusahaan atau pendapatan perusahaan. Semakin tinggi rasio PER sebuah perusahaan maka investor akan menilai bahwa perusahaan itu cukup mahal sebaliknya bila rasio PER cukup rendah maka perusahaan tersebut dinilai oleh investor cukup murah.

3. Price to Book Value (PBV)

Price to Book Value (PBV) merupakan gambaran yang memberitahukan mengenai hubungan antar harga saham dengan nilai buku per lembar saham. Rasio ini dapat digunakan untuk menentukan nilai suatu saham, karena nilai pasar saham harus mencerminkan nilai bukunya.[5]

Price to Book Value dapat diartikan sebagai rasio yang membandingkan harga saham dengan nilai buku dari perusahaan tersebut. Rasio PBV membantu investor dalam menilai apakah sebuah perusahaan *overvalued* atau *undervalued*. Apabila rasio PBV dari sebuah perusahaan memiliki nilai kurang dari 1, maka harga saham tersebut dinilai murah dan sebaliknya bila rasio PBV lebih dari 1, maka harga saham tersebut dinilai mahal.

4. Clustering

Clustering merupakan adalah suatu proses yang memisahkan himpunan objek data menjadi beberapa bagian yang disebut *cluster*, dimana setiap objek dalam suatu *cluster* memiliki karakteristik yang mirip satu sama lain dan berbeda dengan objek dalam *cluster* lainnya. Pembagian *cluster* dilakukan dengan menggunakan algoritma *clustering*, bukan secara manual. Oleh karena itu, *clustering* memiliki nilai tambah karena dapat menemukan kelompok atau grup yang tidak terdeteksi sebelumnya dalam data.[6]

Clustering adalah suatu proses yang membagi sekelompok objek data ke dalam beberapa *cluster* atau kelompok yang memiliki tingkat kesamaan yang tinggi, sementara objek-objek dalam kelompok yang berbeda memiliki perbedaan yang

signifikan. Tujuan dari proses *clustering* ini adalah untuk mengelompokkan objek-objek yang serupa menjadi satu kelompok dan membedakannya dari kelompok lainnya.[7]

5. K-Means

K-Means merupakan Algoritma klasterisasi *K-Means* memiliki konsep dasar yang sederhana yaitu dengan cara meminimalkan jumlah kesalahan kuadrat (*Sum of Squared Error/SSE*) antara data objek dengan sejumlah *k centroid*. [7] *K-Means* adalah algoritma *unsupervised learning* yang berfungsi untuk mengelompokkan data ke dalam beberapa *cluster*. Algoritma ini dapat digunakan untuk data yang tidak memiliki label kategori dan termasuk dalam metode *non-hierarchy*. Tujuan dari metode *Clustering* adalah untuk mengelompokkan data ke dalam kelompok-kelompok yang memiliki karakteristik yang sama, tetapi berbeda dengan kelompok lain. Sebagai contoh, teknik pengambilan sampel *Cluster Sampling* menggunakan kelompok sebagai unit populasi, sedangkan klasterisasi atau *clustering* merupakan suatu masalah dalam teknik *unsupervised learning*. [8]

Tahapan dari *K-Means* sebagai berikut

1. Pilih jumlah *cluster* *k*
2. Inisialisasi *k* pusat *cluster* secara *random*
3. Menghitung jarak setiap data ke pusat dengan rumus *Euclidean distance* sebagai berikut

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (2.1)$$

Dimana $D(i,j)$ = Jarak data ke *i* ke pusat *cluster* *j*

X_{ki} = Data ke *i* pada atribut data ke *k*

X_{kj} = Titik pusat ke *j* pada atribut ke *k*

4. Jika ada data yang berpindah maka hitung ulang jika tidak maka proses perhitungan dapat dihentikan.

6. Min-max normalization

Data yang digunakan dalam penelitian ini akan distandarasi dengan menggunakan metode *min-max normalization*. *Min-max normalization* merupakan metode normalisasi dengan strategi linier untuk mentransformasikan rentang nilai data ke dalam rentang yang baru. Tujuannya adalah untuk mencapai keseimbangan antara nilai perbandingan antara data sebelum dan sesudah proses normalisasi. Melalui metode ini, data akan diubah menjadi seimbang di antara nilai 0 hingga 1. Metode *Min-max normalization* merupakan salah satu metode mengubah data yang kompleks dengan tidak menghilangkan isi, sehingga lebih mudah diolah. [9]

Berikut merupakan rumus dari *min-max normalization* :

$$z = \frac{x - \min(x)}{[\max(x) - \min(x)]} \quad (2.2)$$

Dimana Z = hasil normalisasi,

X = nilai *x* (asli),

$\min(x)$ = nilai minimal untuk variabel *x*,

$\max(x)$ = nilai maksimal untuk variabel *x*.

7. Elbow

Metode *Elbow* adalah salah satu teknik untuk menentukan jumlah *cluster* yang optimal dengan cara membandingkan persentase hasil perbandingan antara jumlah *cluster* yang membentuk siku pada suatu titik. Metode *Elbow* untuk menentukan jumlah *cluster* yang paling optimum atau yang terbaik. Langkah-langkah metode *elbow*:

1. Menginisialisasi nilai awal *k*
2. Meningkatkan nilai *k*.
3. Menghitung nilai *Sum of Square Error* untuk setiap nilai *k*.
4. Menganalisis *Sum of Square Error* untuk nilai *k* yang mengalami penurunan secara signifikan.

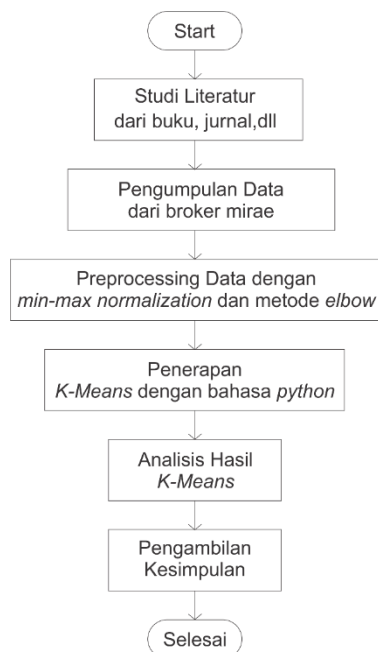
5. Menetapkan nilai k pada titik dimana terdapat siku pada grafik hasil analisis *Sum of Square Error*. Berikut Merupakan rumus dari metode *elbow* :

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|^2 \quad (2.3)$$

Dimana K = Jumlah *Cluster*
 X_i = data ke- i
 C_k = *centroid cluster*

3. METODOLOGI

Penelitian ini dilakukan pada perusahaan – perusahaan yang terdaftar pada Bursa Efek Indonesia. Dengan mengambil rasio keuangan berupa *Price Earning Ratio* (PER) dan *Price to Book Value* (PBV).



Gambar 1. Metodologi Penelitian

Gambar 1 diatas merupakan kerangka penelitian yang digunakan oleh peneliti dalam penelitian ini. Kerangka penelitian ini yang akan menjadi pedoman bagi peneliti dalam melakukan penelitian ini. Kerangka penelitian ini terdiri dari 6 tahap.

Berikut merupakan Langkah – Langkah yang akan dilakukan pada Penelitian ini

1. Studi Literatur

Tahap ini merupakan tahap pertama dalam penelitian ini, dimana peneliti mengumpulkan seluruh informasi yang berkaitan dengan penerapan algoritma *K-Means*. Informasi yang telah dikumpulkan dapat berasal dari berbagai sumber seperti buku, jurnal, artikel, dan skripsi. Dengan studi literatur yang dilakukan, peneliti mendapatkan berbagai informasi serta pengetahuan dalam mengerjakan penelitian ini. Serta peneliti juga dapat mengetahui persamaan dan perbedaan terhadap penelitian yang dilakukan peneliti dengan peneliti lainnya.

2. Pengumpulan Data

Tahap selanjutnya merupakan tahap pengumpulan data. Pada penelitian ini pengumpulan data sangat penting karena merupakan titik tumpu dari penelitian ini. Tanpa data penelitian ini tidak dapat dilaksanakan. Pengumpulan data ini harus dilakukan dengan benar, jika terdapat data yang salah ataupun

keliru maka akan membuat penelitian ini menjadi sulit dan bahkan mustahil untuk dilakukan dengan benar dan tepat. Selain itu hasil yang didapatkan dari penelitian dengan data yang salah akan menjadi salah juga.

3. Preprocessing Data

Sebelum melakukan proses *Data Mining* terdapat proses yang perlu dilakukan terhadap data yang telah dikumpulkan sebelumnya yaitu *preprocessing* data. Proses ini akan memproses data terlebih dahulu sebelum digunakan dalam proses *Data Mining*. Tahapan *Preprocessing* data yang akan digunakan dalam penelitian ini adalah normalisasi data dengan menggunakan metode *min-max normalization*. Dengan menggunakan metode normalisasi data ini akan menyamakan skala atribut data kedalam *range* yang lebih kecil seperti dari 0 sampai 1. Sehingga setelah data telah disesuaikan ke rentang nilai terbaru akan mempermudah dalam memproses data tersebut. Perhitungan *min-max normalization* menggunakan rumus 2.2.

Tahap selanjutnya yaitu menentukan jumlah *cluster* yang ideal dengan menggunakan metode *elbow*. Metode ini akan melihat persentase hasil dari perbandingan antara jumlah *cluster* yang akan membentuk siku pada suatu titik. Perhitungan metode *elbow* menggunakan rumus 2.3.

4. Penerapan K-Means

Tahap ini merupakan proses penerapan algoritma *K-Means* pada data yang telah melewati tahap *preprocessing* data yaitu *min-max normalization* dan metode *elbow*. Algoritma *K-Means* ini akan menghasilkan *cluster* data yang akan dianalisis pada tahap berikutnya.

Tahapan dari algoritma *K-Means* adalah sebagai berikut :

- Pilih jumlah *cluster* k ,
- Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara *random*. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka *random*,
- Alokasikan semua data/ objek ke *cluster* terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Perhitungan jarak dengan menggunakan rumus *Euclidean distance* 2.1.
- Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
- Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi

5. Analisis Hasil K-Means

Tahap ini merupakan tahap analisis hasil dari penerapan *K-Means*. Setelah hasil didapatkan, *cluster-cluster* yang didapatkan dari penelitian ini akan dideskripsikan berdasarkan atribut yang terbentuk. Kemudian melihat pola hubungan yang terbentuk dari hasil *cluster*.

6. Pengambilan Kesimpulan

Tahap terakhir adalah pengambilan kesimpulan. Pada tahap ini telah diperoleh hasil dari penelitian menggunakan algoritma *K-Means*. Peneliti akan menyampaikan keunggulan dan kelemahan dari penelitian. Kesimpulan yang ditulis harus

menjawab rumusan masalah yang telah disampaikan pada bab sebelumnya. Selain itu peneliti juga akan memberikan saran bagi peneliti selanjutnya yang akan mengangkat masalah ini sebagai topik penelitiannya agar dapat dijalankan dengan baik.

4. HASIL DAN PEMBAHASAN

Penelitian ini akan menggunakan Bahasa pemrograman *Python* untuk menjalankan algoritma *K-Means* yang tulis pada *Google Colaboratory* untuk mempermudah proses pembangunan program. Data yang digunakan akan berada dalam file dengan ekstensi *.csv* (excel).

1. Sampel Data

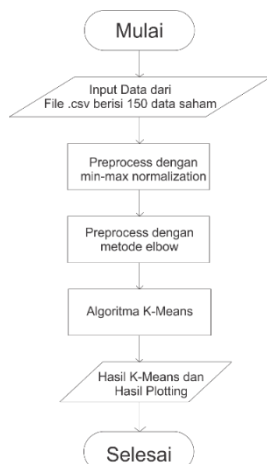
Sampel dari penelitian ini adalah 150 perusahaan yang terdaftar pada Bursa Efek Indonesia untuk yaitu sebagai berikut: Tabel 1 dibawah merupakan tabel yang berisi sampel data yang akan digunakan pada penelitian ini dengan jumlah 150 data.

Tabel 1 : Sampel Data

No	Kode	Nama Perusahaan	PER	PBV
1	AALI	Astra Agro Lestari Tbk.	9,8	0,8
2	ACES	Ace Hardware Indonesia Tbk.	22,3	1,9
3	ADHI	Adhi Karya (Persero) Tbk.	90,8	0,4
4	ADMR	Adaro Minerals Indonesia Tbk.	12,7	9,9
5	ADRO	Adaro Energy Indonesia Tbk.	3,3	1,4
...	...	...	...	...
146	UNTR	United Tractors Tbk.	5,8	1,5
147	UNVR	Unilever Indonesia Tbk.	30	32,2
148	WIRG	Wir Asia Tbk.	111,3	8,8
149	WMUU	Widodo Makmur Unggas Tbk.	12,1	1,1
150	YELO	Yelooo Integra Datanet Tbk.	22,4	15,6

2. Implementasi Sistem

Berikut merupakan *flowchart* yang akan digunakan dalam membangun sistem dengan bahasa *Python*.



Gambar 2. Flowchart Program

Gambar 2. Diatas merupakan *flowchart program* yang akan dibuat. Program ini dimulai dengan menginput data yang berisi 150 data saham yang berasal dari tabel 1. Kemudian melakukan *preprocess* dengan menggunakan *min-max normalization* dan evaluasi *cluster* dengan metode *elbow*. Setelah *cluster* terbaik ditemukan maka selanjutnya penerapan algoritma *K-Means*. Kemudian hasil *K-Means* didapatkan beserta dengan gambar plotting kelompoknya.

Selanjutnya merupakan implementasi sistem yang dibangun dengan menggunakan *Google Colaboratory*

1. Import Data

```

[1] import pandas as pd
    from google.colab import files
    file1= files.upload()

Choose Files skripsi.csv
• skripsi.csv(text/csv) - 6020 bytes, last modified: 1/20/2023 - 100% done
Saving skripsi.csv to skripsi.csv

[2] import numpy as np
    from sklearn.cluster import KMeans
    import matplotlib.pyplot as plt
    import seaborn as sns

[3] df = pd.read_csv('skripsi.csv')

[4] df.head()

```

	Kode	Nama Perusahaan	PER	PBV
0	AALI	Astra Agro Lestari Tbk.	9.8	0.8
1	ACES	Ace Hardware Indonesia Tbk.	22.3	1.9
2	ADHI	Adhi Karya (Persero) Tbk.	90.8	0.4
3	ADMR	Adaro Minerals Indonesia Tbk.	12.7	9.9
4	ADRO	Adaro Energy Indonesia Tbk.	3.3	1.4

Gambar 3. Import Data

Gambar 3. diatas merupakan Langkah pertama yang dilakukan, yaitu mengimport data dengan cara *upload* data. Data yang *diupload* berupa file *excel* dengan format *.csv*. setelah data *diimport* selanjutnya menampilkan Sebagian data yang telah *diimport*.

2. Min-max normalization

```

[95] df = df[['Nama Perusahaan', 'PER', 'PBV']]

[96] df.head()

[97] from sklearn import preprocessing

[98] # min-max normalization
    minmax_scaler = preprocessing.MinMaxScaler()
    minmax = minmax_scaler.fit_transform(df.drop('Nama Perusahaan', axis=1))

[99] df2 = pd.DataFrame(minmax, index=df.index, columns=df.columns[1:])

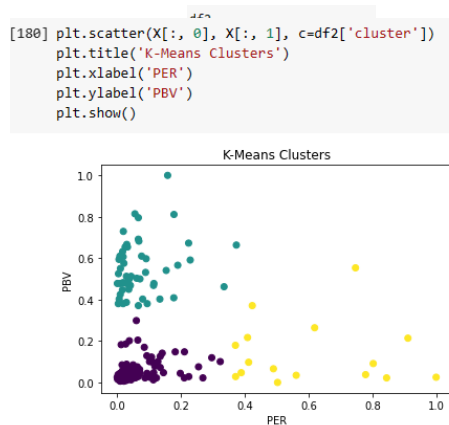
```

	Nama Perusahaan	PER	PBV
0	Astra Agro Lestari Tbk.	9.8	0.8
1	Ace Hardware Indonesia Tbk.	22.3	1.9
2	Adhi Karya (Persero) Tbk.	90.8	0.4
3	Adaro Minerals Indonesia Tbk.	12.7	9.9
4	Adaro Energy Indonesia Tbk.	3.3	1.4

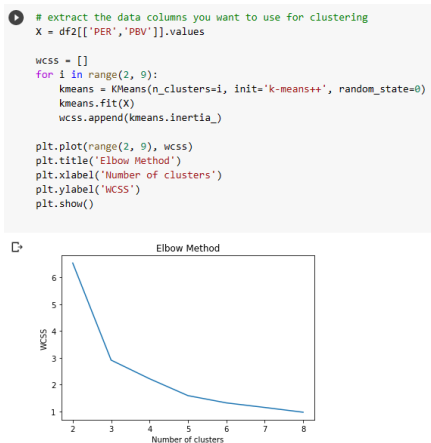
Gambar 4. Min-max normalization

Gambar 4. merupakan Langkah selanjutnya yaitu melakukan normalisasi data dengan *min-max normalization*. Dan menampilkan sebagian data yang telah dinormalisasi

Gambar 5. Tampilan Data setelah Normalisasi
Gambar 5. diatas merupakan tampilan data yang telah melalui tahap *preprocessing min-max normalization*



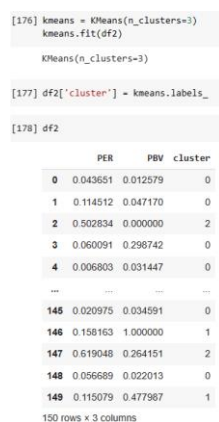
3. Metode Elbow



Gambar 6. Metode Elbow

Gambar 6. diatas merupakan Langkah berikutnya, yaitu menggunakan metode *elbow* untuk menentukan jumlah *cluster* yang ideal. Dapat dilihat dari grafik terdapat siku yang cukup tajam terbentuk pada k uji 3. Sehingga k uji 3 merupakan jumlah *cluster* yang ideal untuk digunakan pada *K-Means*.

4. K-Means Clustering



Gambar 7. K-Means Clustering

Gambar 7. diatas merupakan Langkah selanjutnya, yaitu menggunakan algoritma *K-Means* untuk melakukan *clustering* pada data saham.

Gambar 8. Plotting Cluster K-Means

Gambar 8. diatas merupakan plotting hasil dari *K-Means Clustering* dimana terdapat 3 kelompok data yang terbentuk dengan warna yang berbeda. Kelompok dengan warna ungu merupakan kelompok saham murah dikarenakan kelompok ini berada pada rentang nilai PER dari 0 sampai 0,3... dan nilai PBV dari 0 hingga 0,3... lalu kelompok dengan warna toska merupakan kelompok saham sedang dengan rentang nilai PER dari 0 hingga 0,4... dan PBV dari 0,4 hingga 1. Kemudian kelompok dengan warna kuning merupakan kelompok saham mahal dengan rentang nilai PER 0,4 sampai 1 dan PBV 0 sampai 0,6.

Gambar 9. Plotting Histogram Cluster K-Means

Gambar 9. diatas merupakan *plotting histogram cluster K-Means* dimana terdapat 3 *cluster* yang terdiri dari *cluster 0* berisi 82 anggota, *cluster 1* berisi 52 anggota dan *cluster 2* berisi 16 anggota.

5. Hasil K-Means Clustering

```
[17] df2['Nama Perusahaan'] = df2['Nama Perusahaan']

[18] df2
```

index	PER	PBV	cluster	Nama Perusahaan
0	0.04365079365079365	0.012578616352201255	0	Astra Agro Lestari Tbk.
1	0.11451247165532881	0.047169811320754707	0	Ace Hardware Indonesia Tbk.
2	0.5028344671201814	0.0	2	Adhi Karya (Persero) Tbk.
3	0.0600907029478458	0.2987421383647798	0	Adaro Minerals Indonesia Tbk.
4	0.006802721088435373	0.03144654088050314	0	Adaro Energy Indonesia Tbk.
5	0.423469387755102	0.371069182389937	2	Aneka Gas Industri Tbk.
6	0.06179138321995464	0.0732704402515722	0	AKR Corporindo Tbk.
7	0.04591836734693877	0.022012578616352196	0	Asuransi Multi Artha Guna Tbk.
8	0.07199546485260772	0.04402515723270439	0	Bank Amar Indonesia Tbk.
9	0.13378684807256233	0.40251572327044016	1	Ateliers Mecaniques D'Indonesie Tbk.
10	0.2290249433106576	0.591194968553459	1	Sumber Alfaria Trijaya Tbk.
11	0.0634920634920635	0.053459119496855334	0	Aneka Tambang Tbk.
12	0.29648526077097503	0.11949685534591192	0	Cilacap Samudera Fishing Industry Tbk
13	0.036848072562358274	0.03144654088050314	0	Astra International Tbk.
14	0.13492063492063489	0.0732704402515722	0	Adi Sarana Armada Tbk.
15	0.01870748299319728	0.38050314465408797	1	Astra Otoparts Tbk.
16	0.18197278911564627	0.14779874213836475	0	Avia Avian Tbk.
17	0.177437641723356	0.408050314465408	1	Bank Capital Indonesia Tbk.
18	0.141156462585034	0.14150943396226412	0	Bank Central Asia Tbk.
19	0.9115646258503401	0.21383647798742134	2	Allo Bank Indonesia Tbk.
20	0.03968253968253968	0.028301086792452824	0	Bank Negara Indonesia (Persero) Tbk.
21	0.061224489795918366	0.05974842767295956	0	Bank Rakyat Indonesia (Persero) Tbk.
22	0.3713151927437641	0.028301086792452824	2	Pelayaran Nasional Bina Buana Ray Tbk
23	1.0	0.02515723270440251	2	Bumi Benowo Sukses Sejahtera Tbk.
24	0.01706802721088433	0.481132075471698	1	Bank Tabungan Negara (Persero) Tbk.

Gambar 10. Hasil K-Means Halaman 1

Gambar 10. diatas merupakan hasil yang didapatkan dari algoritma *K-Means* pada data saham. Terdapat 25 data pada halaman pertama hasil *K-Means*

index	PER	PBV	cluster	Nama Perusahaan
25	0.047619047619047616	0.053459119496855334	0	BFI Finance Indonesia Tbk.
26	0.22448979591836735	0.028301086792452824	0	Astrindo Nusantara Infrastruktur Tbk.
27	0.04365079365079365	0.5125786163522011	1	Blue Bird Tbk.
28	0.04875283446712018	0.053459119496855334	0	Bank Mandiri (Persero) Tbk.
29	0.38088080808080809	0.047169811320754707	2	Bank Bumi Arta Tbk.
30	0.061224489795918366	0.5031446540880502	1	Borneo Olah Sarana Sukses Tbk.
31	0.06802721088435373	0.05668377358490565	0	Bank Syariah Indonesia Tbk.
32	0.844671201814059	0.022012578616352196	2	Bumi Resources Minerals Tbk.
33	0.07653061224489796	0.6100628938017609	1	Bumi Serpong Damai Tbk.
34	0.06292517060802721	0.07547169811320763	0	Baramuti Sukses Sarana Tbk.
35	0.05555555555555555	0.8144654088050313	1	Bank BTPN Syariah Tbk.
36	0.020975066893424	0.5754716981132074	1	Bukalapak.com Tbk.
37	0.0034013605442176874	0.38050314465408797	1	Bumi Resources Tbk.
38	0.026543990929705215	0.1855345911949685	0	Bayan Resources Tbk.
39	0.05668934240362812	0.04402515723270439	0	Campina Ice Cream Industry Tbk.
40	0.06462585034013606	0.2044025157232704	0	Centralama Telekomunikasi Indonesia Tbk.
41	0.095230962309623	0.037735849056603765	0	Chemstar Indonesia Tbk.
42	0.1116780045351474	0.09748427672955973	0	Charoen Pokphand Indonesia Tbk.
43	0.03741496598639455	0.49999999999999999	1	Ciputra Development Tbk.
44	0.32312925170068024	0.10062893801761004	0	Dewi Shri Farnindo Tbk.
45	0.26870748299319724	0.022012578616352196	0	Nusa Konstruksi Engineering Tbk.
46	0.05895691609977324	0.02515723270440251	0	Central Omega Resources Tbk.
47	0.03401360544217687	0.03144654088050314	0	Puradelta Lestari Tbk.
48	0.031746031746031744	0.012578616352201255	0	Delta Dunia Makmur Tbk.
49	0.012471655328798183	0.009433962264150941	0	Dharma Satya Nusantara Tbk.

Gambar 11. Hasil K-Means Halaman 2

Gambar 11 diatas merupakan hasil yang didapatkan dari algoritma *K-Means* pada data saham. Terdapat 25 data pada halaman kedua hasil *K-Means*

index	PER	PBV	cluster	Nama Perusahaan
50	0.0209750566893424	0.0062893081761006275	0	Elmusu Tbk.
51	0.05952380952380952	0.0691823899371069	0	Elang Mahkota Teknologi Tbk.
52	0.33503401360544216	0.4622641509433961	1	Champ Resto Indonesia Tbk.
53	0.02834467120181406	0.009433962264150941	0	Energi Mega Persada Tbk.
54	0.02947845804988662	0.4779874213836477	1	Erajaya Swasembada Tbk.
55	0.02947845804988662	0.08805031446540879	0	Surya Esa Perkasa Tbk.
56	0.10260770975056689	0.028301886792452824	0	XL Axiata Tbk.
57	0.7471655328798187	0.5534591194968552	2	MD Picture Tbk.
58	0.012471655328798183	0.1823899371069182	0	Golden Energy Mines Tbk.
59	0.11337868480725626	0.012578616352201255	0	Gudang Garam Tbk.
60	0.04931972789115647	0.05345911949685534	0	GoTo Gojek Tokopedia Tbk.
61	0.5612244897959183	0.03459119496855345	2	Aman Agrindo Tbk.
62	0.4087301587301587	0.21698113207547165	2	Medikaloka Harmina Tbk.
63	0.017573696145124718	0.047169811320754707	0	Hexindo Adiperkasa Tbk.
64	0.07993197278911564	0.40251572327044016	1	H.M. Sampoerna Tbk.
65	0.016439909297052156	0.05660377358490565	0	Harum Energy Tbk.
66	0.7789115646258504	0.037735849056603765	2	Inti Bangun Sejahtera Tbk.
67	0.11791383219954647	0.07861635220125784	0	Indofood CBP Sukses Makmur Tbk.
68	0.0935374149659864	0.04402515723270439	0	Vale Indonesia Tbk.
69	0.036281179138321996	0.44968553459119487	1	Indofood Sukses Makmur Tbk.
70	0.01020408163265306	0.5503144654088049	1	Indo-Rama Synthetics Tbk.
71	0.0011337868480725596	0.01572327044025157	0	Indika Energy Tbk.
72	0.009637188208616776	0.0062893081761006275	0	Indah Kiat Pulp & Paper Tbk.
73	0.14512471655328796	0.047169811320754707	0	Indocement Tunggal Prakarsa Tbk.
74	0.48979591836734687	0.06603773584905659	2	Indo Pureco Pratama Tbk.

Gambar 12. Hasil *K-Means* Halaman 3

Gambar 12. diatas merupakan hasil yang didapatkan dari algoritma *K-Means* pada data saham. Terdapat 25 data pada halaman ketiga hasil *K-Means*.

index	PER	PBV	cluster	Nama Perusahaan
75	0.3735827664399093	0.6635220125786162	1	MNC Vision Networks Tbk.
76	0.18990929705215417	0.5660377358490565	1	Itama Ranoraya Tbk.
77	0.05555555555555555	0.053459119496855334	0	Indosat Tbk.
78	0.002834467120181405	0.037735849056603765	0	Indo Tambangraya Megah Tbk.
79	0.04081632653061225	0.03144654088050314	0	Japfa Comfeed Indonesia Tbk.
80	0.0873015873015873	0.022012578616352196	0	Jasa Marga (Persero) Tbk.
81	0.017006802721088433	0.6320754716981131	1	Kedawung Setia Industrial Tbk.
82	0.0663265306122449	0.795597482767294	1	Kawasan Industri Jababeka Tbk.
83	0.8027210884353742	0.0911949685534591	2	Kioson Komersial Indonesia Tbk.
84	0.13492063492063489	0.12578616352201255	0	Kalbe Farma Tbk.
85	0.10770975056689343	0.12264150943396224	0	Bangun Karya Perkasa Jaya Tbk.
86	0.17800453514739228	0.811320754716981	1	Link Net Tbk.
87	0.03458049886621315	0.04402515723270439	0	Lippo General Insurance Tbk.
88	0.0663265306122449	0.691823899371069	1	Lippo Karawaci Tbk.
89	0.02494331065759637	0.6540880503144653	1	Matahari Department Store Tbk.
90	0.028911564625850344	0.3867924528301886	1	PP London Sumatra Indonesia Tbk.
91	0.03798105941043084	0.06289308176100628	0	Mitra Adiperkasa Tbk.
92	0.08503401360544219	0.16981132075471694	0	Mahaka Radio Integra Tbk.
93	0.006235827664399092	0.40251572327044016	1	Mitrabara Adiperdana Tbk.
94	0.01530612244897959	0.08490566037735847	0	Prima Andalan Mandiri Tbk.
95	0.37074829931972786	0.17942582031886788	2	Merdeka Copper Gold Tbk.
96	0.003968253968253966	0.5251572327044024	1	Medco Energi Internasional Tbk.
97	0.090702947845805	0.5974842767295596	1	Midi Utama Indonesia Tbk.
98	0.22392290249433106	0.67295959748427671	1	Mitra Keluarga Karyasehat Tbk.
99	0.009637188208616776	0.42452830188679236	1	Mulia Industrindo Tbk.

Gambar 13. Hasil *K-Means* Halaman 4

Gambar 13 diatas merupakan hasil yang didapatkan dari algoritma *K-Means* pada data saham. Terdapat 25 data pada halaman keempat hasil *K-Means*

index	PER	PBV	cluster	Nama Perusahaan
100	0.09410430839002266	0.38050314465408797	1	Multipolar Tbk.
101	0.03798105941043084	0.20125786163522008	0	Multipolar Technology Tbk.
102	0.21031746031746032	0.022012578616352196	0	Mega Manunggal Property Tbk.
103	0.020408163265306124	0.481132075471698	1	Media Nusantara Citra Tbk.
104	0.028911564625850344	0.5125786163522011	1	Mitra Pinasthika Mustika Tbk.
105	0.07029478458049887	0.06289308176100628	0	Metrodata Electronics Tbk.
106	0.1978458049886621	0.04402515723270439	0	Dayamitra Telekomunikasi Tbk.
107	0.41213151927437636	0.09748427672955973	2	Murni Sadar Tbk.
108	0.03231292517006803	0.6540880503144653	1	PAM Mineral Tbk.
109	0.005668934240362807	0.4779874213836477	1	Ancora Indonesia Resources Tbk.
110	0.005668934240362812	0.594339226415093	1	Pan brothers Tbk.
111	0.026077097505668938	0.022012578616352196	0	Perusahaan Gas Negara Tbk.
112	0.01870748299319728	0.6069182389937106	1	Panca Mitra Multiperdana Tbk.
113	0.02324263038548753	0.03144654088050314	0	Pinago Utama Tbk.
114	0.036281179138321996	0.496853459119496	1	Pinin Financial Tbk.
115	0.041383219954648526	0.46855345911949675	1	Cikarang Listrindo Tbk.
116	0.07142857142857142	0.9999999999999999	1	Prima Globalindo Logistik Tbk.
117	0.061224489795918366	0.05660377358490565	0	Prodia Widayahusada Tbk.
118	0.008503401360544218	0.04402515723270439	0	Bukit Asam Tbk.
119	0.15362811791383219	0.540880503144654	1	PP (Persero) Tbk.
120	0.06575963718820861	0.028301886792452824	0	Pakuwon Jati Tbk.
121	0.25510204081632654	0.07547169811320753	0	Rukun Raharja Tbk.
122	0.0663265306122449	0.371069182389937	1	Surya Citra Media Tbk.
123	0.06802721088435373	0.6823899371069181	1	Organon Pharma Indonesia Tbk.
124	0.016439909297052156	0.44654088050314465	1	Millennium Pharmacon International Tbk.

Gambar 14. Hasil *K-Means* Halaman 5

Gambar 14. diatas merupakan hasil yang didapatkan dari algoritma *K-Means* pada data saham. Terdapat 25 data pada halaman kelima hasil *K-Means*

index	PER	PBV	cluster	Nama Perusahaan
125	0.008503401360544218	0.6100628930817609	1	Sampoerna Agro Tbk.
126	0.11394557823129252	0.46855345911949675	1	Industri Jamu dan Farmasi Sido Muncul Tbk.
127	0.12755102040816327	0.10062893081761004	0	Gaya Abadi Sempurna Tbk.
128	0.0	0.02515723270440251	0	Samudera Indonesia Tbk.
129	0.10204081632653061	0.02515723270440251	0	Semen Indonesia (Persero) Tbk.
130	0.024376417233560092	0.0691823899371069	0	Golden Eagle Energy Tbk.
131	0.12301587301587301	0.022012578616352196	0	Summarecon Agung Tbk.
132	0.030045351473922906	0.6666666666666665	1	Soechi Lines Tbk.
133	0.009637188208616776	0.4779874213836477	1	Saratoga Investama Sedaya Tbk.
134	0.08900226757369614	0.531446540880503	1	Surya Semesta Internusa Tbk.
135	0.022108843537414963	0.04402515723270439	0	Sawit Sumbermas Sarana Tbk.
136	0.049886621315192746	0.08490566037735847	0	Sumber Tani Agung Resources Tbk.
137	0.011904761904761906	0.03144654088050314	0	Triputra Agro Persada Tbk.
138	0.21088435374149656	0.14779874213836475	0	Tower Bersama Infrastructure Tbk.
139	0.019274376417233556	0.7295597484276728	1	Tunas Baru Lampung Tbk.
140	0.02494331065759637	0.03144654088050314	0	Timah Tbk.
141	0.006802721088435373	0.009433962264150941	0	Pabrik Kertas Tjiwi Kimia Tbk.
142	0.10147392290249434	0.10062893081761004	0	Telkom Indonesia (Persero) Tbk.
143	0.018140589569160995	0.022012578616352196	0	TBS Energi Utama Tbk.
144	0.09297052154195011	0.12893081761006286	0	Sarana Menara Nusantara Tbk.
145	0.0209750566893424	0.03459119496855345	0	United Tractors Tbk.
146	0.15816326530612243	0.9999999999999999	1	Unilever Indonesia Tbk.
147	0.6190476190476191	0.2641509433962236	2	Wir Asia Tbk.
148	0.0566893420362812	0.022012578616352196	0	Widodo Makmur Unggas Tbk.
149	0.11507936507936507	0.4779874213836477	1	Yeloo Integra Datamart Tbk.

Gambar 15. Hasil *K-Means* Halaman 6

Gambar 15. diatas merupakan hasil yang didapatkan dari algoritma *K-Means* pada data saham. Terdapat 25 data pada halaman keenam hasil *K-Means*

3. Pembahasan

Berdasarkan hasil yang telah didapatkan melalui implementasi bahasa pemrograman *Python* didapatkan hasil sebagai berikut. Metode *elbow* pada Penelitian ini digunakan untuk mengoptimalkan jumlah *cluster* yang akan digunakan pada algoritma *K-Means*. Hasil yang didapatkan berupa 3 *cluster* setelah melalui pengujian dari k uji 2, k uji 3, k uji 4 sampai dengan k uji 8. Pada *line chart* terbentuk siku yang cukup tajam pada k uji 3 sehingga jumlah k yang akan digunakan pada algoritma *K-Means* adalah 3.

Penelitian ini juga menggunakan algoritma *K-Means* yang menggunakan jumlah k yang telah diperoleh dari metode *elbow* diatas yaitu k=3. Hasil yang diperoleh dari proses perhitungan menggunakan *Python* yaitu 3 *cluster*. Terdiri dari *cluster* pertama atau 0 dengan jumlah 82 anggota. Lalu *cluster* kedua atau 1 dengan jumlah 52 anggota. Kemudian *cluster* ketiga atau 2 dengan jumlah 16 anggota.

Cluster pertama merupakan kumpulan saham dengan nilai murah, kemudian *cluster* kedua merupakan kumpulan saham dengan nilai sedang, lalu *cluster* ke tiga merupakan kumpulan saham dengan nilai mahal. *Cluster* pertama merupakan saham dengan nilai murah karena *cluster* pertama memiliki pusat *cluster* PER (0.0764131) dan PBV (0.06350667). Lalu *cluster* kedua merupakan saham dengan nilai sedang karena *cluster* kedua memiliki pusat *cluster* PER (0.0739033) dan PBV (0.5481374). Kemudian *cluster* ketiga merupakan saham dengan nilai murah karena *cluster* ketiga memiliki pusat *cluster* PER (0.6020763) dan PBV (0.14052673).

Saham dengan nilai murah cenderung memiliki risiko yang rendah, saham dengan nilai sedang cenderung memiliki risiko yang sedang, dan saham dengan nilai mahal cenderung memiliki risiko yang tinggi. Sehingga investor cenderung disarankan untuk memilih saham dengan risiko yang rendah untuk meminimalisir kerugian dalam membeli saham dibandingkan ketika investor memilih saham dengan nilai mahal maka risiko kerugian akan besar.

Tabel 2 dibawah merupakan pusat *cluster* terakhir dari perhitungan algoritma *K-Means* yang telah dilakukan dengan menggunakan bahasa pemrograman *Python*

Tabel 2 : Pusat *Cluster* Algoritma *K-Means*

Cluster	PER	PBV
0	0,0764131	0,06350667
1	0,0739033	0,5481374
2	0,6020763	0,14052673

5. KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil pengolahan data saham dengan menggunakan algoritma *K-Means*, dapat disimpulkan bahwa Perhitungan yang telah dilakukan menghasilkan 3 *cluster* saham yang terdiri dari *cluster* pertama sebagai saham dengan nilai murah berisi 82 anggota, *cluster* kedua sebagai saham dengan nilai sedang berisi 52 anggota, dan *cluster* ketiga sebagai saham dengan nilai mahal berisi 16 anggota. Saham dengan nilai murah cenderung memiliki risiko yang rendah, saham dengan nilai sedang cenderung memiliki risiko yang sedang, dan saham dengan nilai mahal cenderung memiliki risiko yang tinggi. Sehingga investor cenderung disarankan untuk memilih saham dengan risiko yang rendah untuk meminimalisir kerugian dalam membeli saham.

Saran

Berdasarkan penelitian yang telah dilakukan terdapat banyak kekurangan yang dapat diperbaiki dan dikembangkan. Oleh karena itu saran untuk penelitian selanjutnya yaitu penelitian ini dapat dikembangkan lebih jauh dengan membuat aplikasi berbasis *web* ataupun *android* agar semua orang dapat menggunakan algoritma *K-Means* untuk mengelompokkan saham sebagai referensi. Kemudian Penelitian ini dapat menggunakan periode data yang lebih panjang, seperti data dalam periode 1 tahun.

DAFTAR PUSTAKA

- [1] R. Suganthi and P. Kamalakannan, "Analyzing Stock Market Data Using Clustering Algorithm," *Int. J. Futur. Comput. Commun.*, vol. 4, no. 108-111. 10.7763/IJFCC.2015.V4.366., 2015.
- [2] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma *K-Means* Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [3] E. P. Jaya and R. Kuswanto, "Pengaruh Return on Assets, Debt To Equity Ratio Dan *Price to Book Value* Terhadap Return Saham Perusahaan Lq45 Terdaftar Di Bursa Efek Indonesia Periode 2016 – 2018," *J. Bina Akunt.*, vol. 8, no. 1, pp. 51–67, 2021, doi: 10.52859/jba.v8i1.136.
- [4] S. Sukamulja, *Pengantar Pemodelan Keuangan dan Analisis Pasar Modal*. Yogyakarta: CV Andi Offset, 2017.
- [5] E. Tandelilin, *Pasar Modal Manajemen Portofolio & Investasi*. Yogyakarta: PT Kanisius, 2017.
- [6] E. Irwansyah, "Clustering," 2017. <https://socs.binus.ac.id/2017/03/09/clustering/>.
- [7] Suyanto, *Data Mining untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika, 2017.
- [8] B. Orleans and E. P. Putra, "Clustering Algoritma (*K-Means*)," 2022. <https://sis.binus.ac.id/2022/01/31/clustering-algoritma-K-Means/>.
- [9] H. Wimmer, "Effects of Normalization Techniques on Logistic Regression in Data Science," *Proc. Conf. Inf. Syst. Appl. Res.*, pp. 1–9, 2018.