

Klasterisasi Harga Kriptokurensi (BitCoin) Menggunakan Metode K-Means

Ardi Setiyansah dan Sriyanto

Fakultas Ilmu Komputer Institut Informatika dan Bisnis Darmajaya
Jl. ZA Pagar Alam No.93 Bandarlampung 35142
E-mail: ardi.2121210001@mail.darmajaya.ac.id, sriyanto@darmajaya.ac.id*

Abstrak

Kripto memiliki return atau tingkat keuntungan yang signifikan, namun disisi lain *kripto* juga memiliki potensi risiko yang tinggi dalam investasi, *Kripto* memiliki volatilitas yang ekstrem, lonjakan kenaikan dan penurunan harganya sangat cepat, volatilitas yang tinggi merupakan cerminan tingkat risiko yang dihadapi oleh para investor karena pergerakan nilai *Kripto* sangat tidak stabil, karena dapat naik turun sangat cepat. Penelitian ini bertujuan untuk mengklasterisasi harga tinggi dan rendah dari *kripto* menggunakan metode *K-Means*. Dari implementasi yang telah dilakukan diperoleh 2 kluster yaitu kluster 0 dan kluster 1. Dimana, kluster 0 yaitu harga tinggi dengan jumlah data 626 dan kluster 1 yaitu harga rendah dengan jumlah data 1.015 dari total 1.641 record data.

Kata kunci : Klustering, Kriptokurensi, K-Means, *loss*, *profit*, *unsupervised*, *bitcoin*, *Cryptocurrency price*, *K-Means Clustering*

Pendahuluan

Saat ini setidaknya ada 11.000 jenis mata uang kripto yang diperdagangkan. Namun demikian, untuk di Indonesia sendiri, Jenis aset kripto yang dapat diperdagangkan di Indonesia berdasarkan Perba Nomor 7 Tahun 2020 tentang Penetapan Daftar Aset Kripto Yang Dapat Diperdagangkan Di Pasar Fisik Aset Kripto ada 229 aset kripto. Bila dikerucutkan berdasarkan kapitalisasi pasar global, berdasarkan data CoinMarketCap. [1]

Salah satu jenis mata uang kripto yang memiliki nilai paling tinggi yaitu *bitcoin*, dimana pada Maret 2020 harga bitcoin yang sebelumnya berkisar 140 juta sempat jatuh ke harga 60 juta. Setelah itu harga bitcoin justru melonjak lebih tinggi hingga akhirnya pada 10 November 2021 harga *bitcoin* mencapai 970 juta. [2]

Dari pergerakan harga tersebut dapat dilihat bahwa kripto memang memiliki return / tingkat keuntungan yang signifikan, namun disisi lain *kripto* juga memiliki potensi risiko yang tinggi dalam investasi, *Kripto* memiliki volatilitas yang ekstrem, lonjakan kenaikan dan penurunan harganya sangat cepat, volatilitas yang tinggi merupakan cerminan tingkat risiko yang dihadapi oleh para investor karena pergerakan Nilai *Kripto* sangat tidak stabil, karena dapat naik turun sangat cepat.

Berdasarkan latar belakang masalah di atas dapat dilihat bahwa potensi risiko *loss* (kerugian)

pada investasi *kripto* sangat tinggi, hal ini disebabkan lonjakan kenaikan dan penurunan harga yang fluktuatif.

Tujuan dari penelitian ini yaitu melakukan klasterisasi harga tinggi dan rendah menggunakan metode berdasarkan harga *kripto* (*BitCoin*) dari rentan tahun 2018 sampai 2023, harapannya dengan penelitian ini dapat menjadi acuan investor kripto dalam berinvestasi dan menjadi referensi untuk penelitian berikutnya.

Dilakukannya penelitian ini juga sebagai pembandingan dari penelitian-penelitian terdahulu dimana mayoritas penelitian terfokus dengan prediksi harga kriptokurensi. Beberapa penelitian terdahulu sebagai berikut:

1. Prediksi Harga *Cryptocurrency* Menggunakan Algoritma *Long Short Term Memory* (LSTM) [3]
2. Implementasi orange data mining untuk prediksi harga *bitcoin* [4]
3. Aplikasi prediksi jangka pendek harga *bitcoin* menggunakan metode arima
4. Prediksi Harga *Bitcoin* Menggunakan Metode *Random Forest* [5]
5. Prediksi harga *cryptocurrency* dengan metode *k-nearest neighbours* [6]

DOI : <http://dx.doi.org/10.32409/jikstik.22.4.3409>,

*)Penulis Korespondensi

- Prediksi harga *bitcoin* berdasarkan informasi blockchain menggunakan metode *long-short term memory* (LSTM) [7]

Metode Penelitian

Algoritma K-means

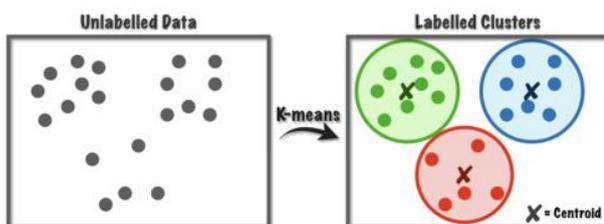
K-means adalah algoritma pengelompokan tanpa pengawasan yang dirancang untuk mempartisi data yang tidak berlabel menjadi sejumlah tertentu (itulah "K") dari pengelompokan yang berbeda. Dengan kata lain, *k-means* menemukan observasi yang memiliki karakteristik penting dan mengklasifikasikannya bersama ke dalam cluster. Solusi pengelompokan yang baik adalah solusi yang menemukan kluster sedemikian rupa sehingga pengamatan dalam setiap kluster lebih mirip daripada kluster itu sendiri.

K-Means clustering adalah algoritma deskriptif yang menskalakan dengan baik hingga besar data (Hartigan, 1975). Analisis kluster memiliki aplikasi yang luas, termasuk: segmentasi pelanggan, pengenalan pola, studi biologi, dan web klasifikasi dokumen. *k*-Means clustering mencoba untuk menemukan *k* partisi (McQueen, 1967) dalam data, di mana setiap pengamatan milik cluster dengan mean terdekat. [3]

Langkah-langkah dasar untuk *k*-means adalah :

- Pilih *k* observasi secara sewenang-wenang sebagai centroid cluster awal.
- Tugaskan setiap pengamatan ke cluster yang memiliki centroid terdekat.
- Setelah semua pengamatan ditetapkan, hitung ulang posisi dari *k* centroid.
- Ulangi langkah 2 dan 3 hingga centroid tidak lagi berubah.

Pengulangan ini membantu meminimalkan variabilitas dalam cluster dan memaksimalkan itu variabilitas di antara cluster, lihat Gambar 1.



Gambar 1: *k*-Means Clustering



Gambar 2: Tahapan penelitian

Tahapan Penelitian

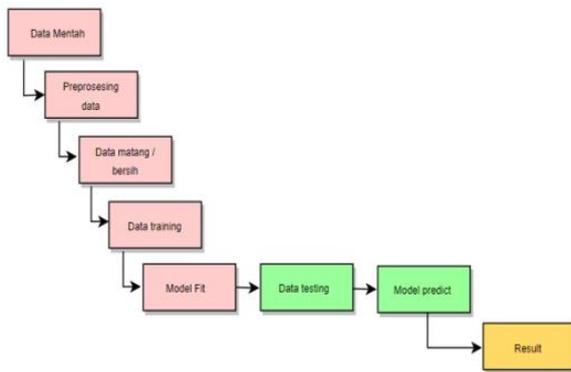
Dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset dalam memudahkan penelitian dan dapat berjalan dengan sistematis dan memenuhi tujuan yang diinginkan maka dibuat alur (lihat Gambar 2) dalam tahapan penelitian yang akan dilakukan berikut :

- Tahap Pemahaman bisnis juga dapat disebut fase pemahaman penelitian yang meliputi tujuan dan persyaratan proyek dengan jelas dalam hal bisnis atau unit penelitian secara keseluruhan, menterjemahkan tujuan dan batasan ke dalam perumusan definisi masalah data mining, menyiapkan strategi awal untuk mencapai tujuan tersebut.
- Tahap pemahaman data dengan melakukan pengumpulan data menggunakan analisis data eksplorasi untuk membiasakan diri dengan data dan menemukan wawasan awal, mengevaluasi kualitas data, apabila diinginkan pilih subset menarik yang mungkin berisi pola yang dapat ditindaklanjuti.
- Tahap persiapan data adalah mempersiapkan data mentah awal kumpulan data akhir yang akan digunakan untuk semua tahapan berikutnya. Pada tahapan ini memilih kasus dan variabel yang diinginkan, menganalisis yang sesuai untuk analisis. Melakukan transformasi pada variabel tertentu, jika diperlukan, membersihkan data mentah sehingga siap untuk alat pemodelan.
- Tahap pemodelan, Tahapan ini dilakukan dengan memilih dan menerapkan teknik pemodelan yang sesuai. Kalibrasi pengaturan model untuk mengoptimalkan hasil, seringkali, beberapa teknik berbeda dapat digunakan untuk masalah data mining yang

sama, apabila diperlukan mengulangi kembali ke tahapan persiapan data untuk membawa bentuk data sesuai dengan persyaratan khusus dari teknik data mining tertentu.

5. Tahap evaluasi adalah mengevaluasi satu atau lebih model yang disampaikan dalam tahap pemodelan untuk kualitas dan efektivitas sebelum menerapkannya untuk digunakan di lapangan. Pada tahap ini yang akan dilakukan adalah menentukan model sebenarnya mencapai tujuan yang ditetapkan.

Gambar 3 adalah ini rencana tahapan pemodelan yang dilakukan.



Gambar 3: Tahap pemodelan

Hasil dan Pembahasan

Dalam penelitian ini pengujian metode *K-Means* menggunakan bahasa Python melalui google colab, Sedangkan dataset yang digunakan berasal dari <https://id.investing.com/crypto/bitcoin/btc-idr-historical-data> dengan format .csv yang di convert ke .xlsx, dataset berisi histori harga BTC (*Bitcoin*) dari rentan januari 2019 sampai juni 2023.

Berikut adalah penerapan Algoritma *K-Means* untuk klusterisasi harga *cryptocurrency* (*Bitcoin*) menggunakan Python dan Google Colab :

1. Import data

```

+ Kode + Teks
[2] import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn.cluster import KMeans
  
```

Gambar 4: Import Library pada google colab

Gambar 4 adalah proses import library yang dibutuhkan dalam klusterisasi data menggunakan algoritma *K-Means* Terlihat pada line 2,

2. Membaca Dataset

```

[12] data = pd.read_excel('/content/Data-Historis-BTC_IDR-BTC-Indonesia-5_.xlsx')
data.head()
  
```

Gambar 5: Read Dataset

Pada line 12 terlihat pada Gambar 5 adalah perintah yang digunakan untuk membaca dataset sesuai directory penyimpanan pada google colab.

	Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Perubahan%
0	11/07/2023	462016000	459825984	462313984	459825984	0,02K	0,48%
1	10/07/2023	459825984	456628992	466859008	454000000	0,03K	0,70%
2	09/07/2023	456628992	457384992	458999008	455817984	0,01K	-0,17%
3	08/07/2023	457384992	459408992	459580000	456200000	0,01K	-0,44%
4	07/07/2023	459408992	455222016	460000000	453000000	0,01K	1,00%

Gambar 6: Table dataset Histori harga Bitcoin

Gambar 6 adalah *table dataset* hasil dari perintah *pd.read_excel* dapat dilihat pada Gambar 5 diatas. Dimana dataset tersebut terdiri dari tanggal, terakhir, pembukaan, tertinggi, Vol. Dan perubahan %.

3. Info Dataset

```

[14] data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1641 entries, 0 to 1640
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   Tanggal    1641 non-null   object
 1   Terakhir   1641 non-null   int64
 2   Pembukaan  1641 non-null   int64
 3   Tertinggi  1641 non-null   int64
 4   Terendah   1641 non-null   int64
 5   Vol.       1640 non-null   object
 6   Perubahan% 1641 non-null   object
dtypes: int64(4), object(3)
memory usage: 89.9+ KB
  
```

Gambar 7: Info Dataset

Pada Gambar 7 adalah hasil dari perintah *data.info()* terlihat pada line [14], dimana perintah tersebut berfungsi untuk menampilkan informasi dataset.

4. Describe dataset

```
[15] data.describe()
```

	Terakhir	Pembukaan	Tertinggi	Terendah
count	1.641000e+03	1.641000e+03	1.641000e+03	1.641000e+03
mean	3.445537e+08	3.442850e+08	3.505779e+08	3.380175e+08
std	2.383803e+08	2.384103e+08	2.425206e+08	2.339207e+08
min	4.805900e+07	4.805900e+07	4.869600e+07	4.750800e+07
25%	1.341500e+08	1.340810e+08	1.358180e+08	1.329050e+08
50%	2.973140e+08	2.973000e+08	3.020000e+08	2.937010e+08
75%	5.195830e+08	5.195830e+08	5.279130e+08	5.091200e+08
max	9.584970e+08	9.580190e+08	9.780600e+08	9.462910e+08

Gambar 8: Describe dataset

Pada Gambar 8 describe dataset adalah informasi berupa Count, mean, std, min, 25%, 50%, 75%, max dari dataset yang digunakan.

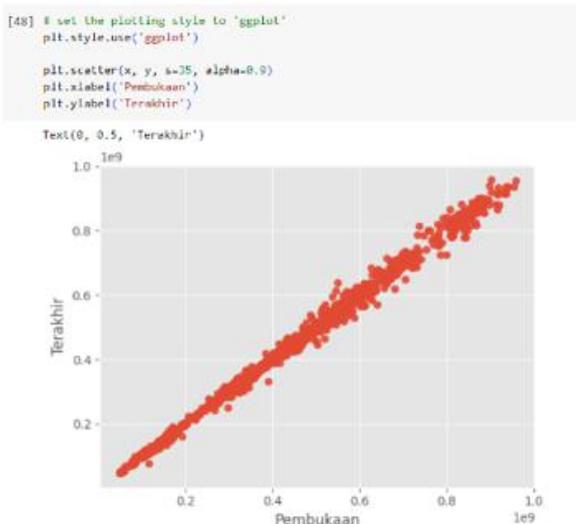
5. Menentukan Variable

```
[47] df = data_encode.copy()
      x = df['Pembukaan']
      y = df['Terakhir']
```

Gambar 9: Describe dataset

Pada Gambar 9 adalah perintah yang digunakan untuk menentukan variable pada dataset yang akan menjadi paramater dalam klastering, pada dataset variable yang di gunakan adalah pembukaan dan Terakhir.

6. Melihat Sebaran data menggunakan GG-PLOT



Gambar 10: Visualisasi Sebaran Data

Pada Gambar 10 adalah perintah dan hasil visualisasi sebaran data menggunakan ggplot

7. Perintah algoritma K-means

```
[49] X = df[['Pembukaan', 'Terakhir']].values

      kmeans = KMeans(n_clusters=2, init='k-means++', n_init='auto', random_state=0).fit(X)

[50] cluster_labels = kmeans.predict(X)
      print(cluster_labels)

[0 0 0 ... 1 1 1]
```

```
[51] cluster_labels = kmeans.labels_
      print(cluster_labels)

[0 0 0 ... 1 1 1]
```

```
[52] centers = kmeans.cluster_centers_
      print(centers)

[[6.05925899e+08 6.06132364e+08]
 [1.82501163e+08 1.82808298e+08]]
```

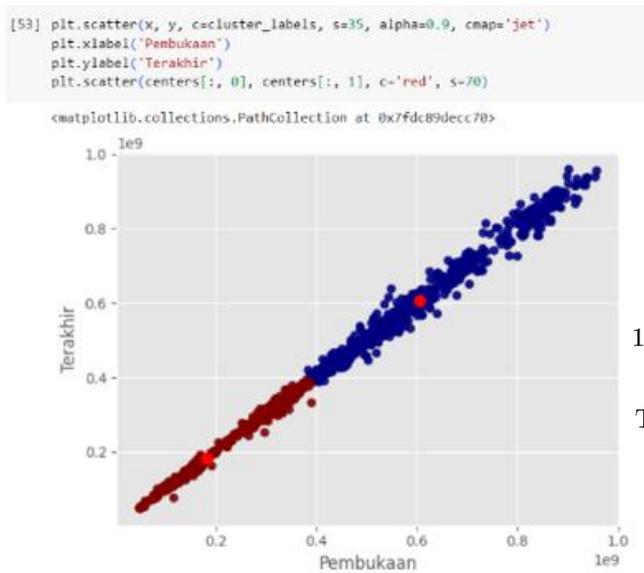
Gambar 11: Perintah Algoritma K-Means

Pada gambar 11 adalah perintah yang digunakan untuk memproses klasterisasi data menggunakan algoritma k-means, berikut adalah deskripsi dari perintah diatas:

- Line 49 : pada line 49 pada gambar diatas adalah perintah yang digunakan untuk menentukan nilai K sesuai dengan parameter dataset yang di gunakan yaitu pembukaan dan terakhir, dimana nilai K = 2 artinya perintah ini akan mencari 2 klaster dari dataset
- Line 50 : Pada line 50 pada gambar diatas adalah perintah yang digunakan untuk memprediksi label sesuai sebaran data dan parameter yang digunakan.
- Line 51 : Pada line 51 pada gambar diatas adalah perintah yang digunakan untuk menampilkan lable dari kedua klaster dalam hal ini kalster 1 = 0 dan klaster 2 =1

Line 52 : pada line 52 pada gambar diatas adalah perintah yang digunakan untuk mendapatkan titik pusat klaster dari dataset.

8. Visualisasi ggplot Klaster K-Means



Gambar 12: Visualisasi Kluster K-Means

Pada Gambar 12 adalah perintah dan hasil vi-

sualisasi untuk klusterisasi, terlihat terdapat 2 kluster yaitu kluster 1 berwarna merah dan kluster 2 berwarna biru.

9. Hasil klusterisasi K-Means

Pada Gambar 13 adalah table hasil klusterisasi algoritma K-Means, dimana pada kolom lable telah terdapat hasil klustering yaitu lable 1 dan 0

10. Hasil Klusterisasi K-Means

Tabel 1: Jumlah record data berdasarkan lable

Lable	Count Record Data
0	626
1	1.015

Dari hasil klustering menggunakan algoritma K-Means dari 1.641 record data 626 data masuk kedalam kluster 0 (Harga Tinggi) dan 1.015 masuk kedalam kluster 1 (Harga rendah), lihat Tabel 1.

Index	Tanggal	Terakhir	Pembukaan	Tertinggi	Terendah	Vol.	Perubahan%	Lable
25	16/06/2023	394244992	381128992	395700000	378000000	0.03K	662	1
26	15/06/2023	381128992	376015008	382508992	373100000	0.03K	507	1
27	14/06/2023	376015008	385200000	388203008	372000000	0.02K	210	1
28	13/06/2023	385200000	385926016	390368992	382100000	0.02K	18	1
29	12/06/2023	385926016	387310016	389291008	382672992	0.02K	34	1
30	11/06/2023	387310016	387199008	391705984	384596000	0.01K	383	1
31	10/06/2023	387199008	393900000	394278016	380000000	0.03K	154	1
32	09/06/2023	393900000	394211008	397244000	392000000	0.02K	7	1
33	08/06/2023	394211008	392159008	397000000	390000000	0.03K	431	1
34	07/06/2023	392159008	404990016	405500000	390000000	0.02K	256	0
35	06/06/2023	404990016	388009992	406000000	381999008	0.03K	703	0
36	05/06/2023	388009992	405508000	405620000	382742016	0.04K	301	0
37	04/06/2023	405508000	404448000	408664992	403999008	0.01K	405	0
38	03/06/2023	404448000	406910016	408044992	400648000	0.01K	57	0
39	02/06/2023	406910016	400500000	406920000	398000000	0.02K	528	0
40	01/06/2023	400500000	407458008	408739008	398318016	0.02K	155	0
41	31/05/2023	407458008	413963008	413964992	403000000	0.02K	145	0
42	30/05/2023	413963008	415401984	417497984	413004000	0.03K	33	0
43	29/05/2023	415401984	418895008	422293008	413000000	0.03K	79	0
44	28/05/2023	418895008	400748000	419020000	400748000	0.04K	709	0
45	27/05/2023	400748000	398596016	402444000	398001984	0.01K	434	0
46	26/05/2023	398596016	395024000	400000000	393539992	0.02K	444	0
47	25/05/2023	395024000	393500000	396225984	388000000	0.01K	440	0
48	24/05/2023	393500000	394937008	397000000	390000000	0.03K	293	0

Gambar 13: Hasil klusterisasi K-Means

Penutup

1. Pada penelitian ini telah berhasil dilakukan klusterisasi harga bitcoin menggunakan algo-

ritma K-Means dengan 2 kluster.

2. Lable yang dihasilkan dari algoritma K-Means dalam penelitian ini yaitu lable 0 = Harga tinggi dan lable 1=Harga rendah

3. Hasil klustering menggunakan algoritma K-Means dari 1.641 record data, 626 data masuk kedalam klaster 0 (Harga Tinggi) dan 1.015 masuk kedalam klaster 1 (Harga rendah)
 4. Algoritma K-Means dapat di implementasikan dengan memanfaatkan fitur google colab dan bahasa pemrograman python.
 5. Pada penelitian ini hasil klasterisasi dapat divisualisasikan menggunakan *k-Means Scatter Plot*, sehingga memudahkan dalam menganalisa data
- [4] Indriyanti, Nurul Ichsan, Haerul Fatah, Tri Wahyuni dan Erni Ermawati, "Implementasi Orange Data Mining untuk Prediksi Harga Bitcoin", Jurnal Responsif Riset Sains dan Inf, vol. 4, No.2 , no. E-ISSN: 2685-6964, pp. 118-1, Agustus 2022.
 - [5] H. S. Siti Saadah, "Prediksi Harga Bitcoin Menggunakan Metode Random Forest," Jurnal Politeknik Caltex Riau, Vols. Vol. 7, No. 1, Mei 2021, no. e- ISSN : 2460-5255, pp. 24-32, 2021.
 - [6] A. S. Haerul Fatah, "Prediksi Harga Cryptocurrency dengan Metode K-Nearest Neighbours ", Jurnal PILAR Nusa Mandiri, Vols. Vol. 14, No. 2 September 2018,

Daftar Pustaka

- [1] Anonim, "Komoditas Berjangka," Badan Pengawas Perdagangan Berjangka Komoditi, 2021. diakses daring pada : https://bappebti.go.id/resources/docs/brosur_leaftlet_2001_01_10_7zwvgs5w.pdf. [Accessed Monday June 2023].
- [2] Anonim, "Data Historis BTC/IDR BTC Indonesia", Id Investing, diakses daring pada <https://-id.investing.com/crypto/bitcoin/btc-idr-historical-data>. [Accessed Monday June 2023].
- [3] S. W. Moch Farryz Rizkilloh, "Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term," Jurnal Resti, Vols. Vol. 6 No. 1 (2022) 25 - 31, no. 31, pp. 25 - 31, 2022.
- [4] J. A. Hartigan, "Clustering Algorithms," in Clustering Algorithms, the University of Michigan, Wiley, 1975, 2010.
- [5] S. L. Nur Fitriani Bintang Pradana, "Aplikasi Prediksi Jangka Pendek Harga Bitcoin Menggunakan Metode Arima", Jurnal Ilmiah Informatika Komputer, Vol. 25 No. 3, , no. <https://doi.org/10.35760/ik.2020.v25i3.3128>, Desember 2020.
- [6] K. D. Larasati, "Prediksi Harga Bitcoin Berdasarkan Informasi " Tugas Akhir, Prodi Statistika, Fak. Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta, 2020.