

## **ANALISIS DAN PREDIKSI PENUTUPAN HARGA SAHAM PADA PT ADARO ENERGY INDONESIA TBK MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR REGRESSION***

**Farikha Dwi Nur Qossina Januar<sup>1</sup>**  
**Ketut Tri Budi Artani<sup>2</sup>**  
**Nengah Widya Utami<sup>3</sup>**

<sup>1,2,3</sup>(STMIK Primakara)

<sup>1</sup>qossina.dwi@gmail.com

### **Abstract**

*The capital market is an important indicator of a country's economic development. The ability to predict stock prices accurately can provide many advantages for investors and companies. This study aims to analyze the comparison of the use of different attributes in predicting the closing price of the stock against the resulting accuracy using 11 attributes and 12 attributes with the K-Nearest Neighbor Regression algorithm. This study also aims to predict stock price for PT Adaro Energy Indonesia Tbk using the K-Nearest Neighbor Regression algorithm. The research methodology used in this research is Knowledge Discovery in Databases. The results showed that the use of 11 attributes in predicting the closing price of PT Adaro Energy Indonesia Tbk. gives better results than using 12 attributes. The model evaluation yielded outcomes with 11 attributes, showcasing a Root Mean Squared Error (RMSE) of 35,02. Additionally, the R-squared ( $R^2$ ) value stood at 0.99, accompanied by an Explained Variance Score (EVS) and Mean Absolute Error (MAE) of 0.99 and 24.54, respectively. This shows that selecting the right attribute is very important in predicting the closing price of a stock. Therefore, the selection of the right attributes must be considered in building an accurate predictive model.*

**Keywords:** Capital Market, KNN Regression, Knowledge Discovery in Databases, Stocks

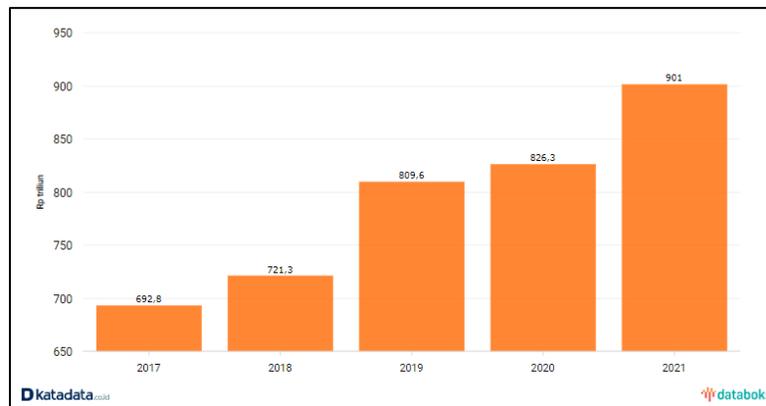
### **I. PENDAHULUAN**

Pasar modal merupakan sebuah wadah bertemunya antara penjual dan pembeli modal (Rianti, 2021). Pemilik modal dapat melakukan investasi modalnya dengan harapan memperoleh keuntungan di masa depan. Pasar modal memberikan kesempatan kepada pemiliki modal untuk bisa memperoleh imbalan berdasarkan karakteristik investasi yang dipilih (Wardiyah, 2017).

Berdasarkan pada Gambar 1, rata-rata kenaikan investasi pada periode 2017-2021 yaitu sebesar 6,9% (Dihni, 2022). Hal tersebut menunjukkan bahwa realisasi

investasi di Indonesia mengalami peningkatan. Kegiatan investasi yang berjalan secara terus menerus, hal tersebut tentunya dapat meningkatkan aktivitas ekonomi dan meningkatkan taraf kemakmuran masyarakat. Masyarakat yang melakukan investasi dikenal dengan istilah investor.

**Gambar 1**  
**Grafik Realisasi Investasi Periode 2017-2021**



Sumber: [databoks.katadata.co.id](http://databoks.katadata.co.id), 2022

Pertumbuhan investor pasar modal mengalami tren peningkatan setiap tahunnya. Berdasarkan pada laporan dari PT. Kustodian Sentral Efek Indonesia (KSEI) menyebutkan bahwa pada September 2022, total investor pasar modal di Indonesia sudah melampaui lebih dari 9 juta investor (KSEI, 2022a). Berdasarkan pada total investor yang ada, terdapat *Single Investor Identification* (SID) yang berinvestasi di saham sebanyak lebih dari 4 juta investor. Jumlah tersebut telah meningkat 22,45% dari akhir tahun 2021 yang jumlahnya sekitar 3,4 juta (KSEI, 2022b). Hal tersebut menunjukkan bahwa kinerja pasar modal di Indonesia terus memperlihatkan hasil yang positif yang dibuktikan dari adanya penambahan jumlah investor di pasar modal yang salah satunya merupakan investor yang berinvestasi pada saham.

Aktivitas dari investasi saham membuat para pemilik saham atau investor perlu memperhatikan perubahan harga sahamnya. Perubahan harga saham tersebut akan mencerminkan kinerja dari suatu perusahaan (Astutik et al., 2021). Harga saham meningkat maka perusahaan dianggap memiliki kinerja yang baik. Perubahan harga saham pada perusahaan dapat dilihat melalui Bursa Efek Indonesia (BEI).

**Gambar 2**  
**Grafik Harga Saham Beberapa Sektor**



Sumber: [www.investasimu.com](http://www.investasimu.com), 2022

Jumlah perusahaan tercatat dari berbagai sektor di Bursa Efek Indonesia (BEI) terus bertambah hingga tahun 2022. Hingga Oktober 2022, jumlah perusahaan yang tercatat telah mencapai 810 perusahaan. Pada Gambar 2 terlihat bahwa dari beberapa sektor perusahaan yang tercatat di BEI, sektor energi memiliki riwayat peningkatan tren yang sangat signifikan terhadap pergerakan harga saham dibandingkan dengan sektor-sektor lainnya. Perusahaan yang bergerak di sektor energi Indonesia salah satunya adalah perusahaan tambang batu bara.

Indonesia merupakan produsen batu bara terbesar ketiga di dunia. Pada tahun 2021, Indonesia telah memproduksi sebanyak 609,48 juta ton batu bara (Pahlevi, 2022). Selain itu, harga batu bara mengalami peningkatan hingga pada tanggal 22 Juni 2022, harga batu bara sudah mendekati US\$ 400/ton (CNBC, 2022). Secara historis, harga batu bara global berkorelasi positif dengan pergerakan harga sahamnya. Situasi tersebut tentunya menguntungkan untuk Indonesia terutama perusahaan-perusahaan tambang batu bara. Perusahaan tambang batu bara yang terlibat dalam kegiatan produksi batu bara salah satunya yaitu PT Adaro Energy Indonesia Tbk yang merupakan perusahaan tambang batu bara yang sudah tercatat di BEI sejak tahun 2008. Selain itu, Adaro menjadi salah satu perusahaan tambang batu bara yang masuk ke dalam daftar indeks IDX30 (R. Nugroho, 2022a). IDX30 sendiri merupakan indeks saham yang masuk ke dalam kategori indeks yang digunakan untuk mengukur

performa harga dari saham-saham yang memiliki kapitalisasi pasar yang besar dan likuiditas saham yang tinggi serta didukung oleh fundamental perusahaan yang baik.

**Tabel 1**  
**Saham Pertambangan Batu Bara IDX30 Tahun 2022**

No	Nama Perusahaan	Kapitalisasi Pasar	Frekuensi Transaksi (Kali)	Nilai Transaksi (Rp)
1	Adaro Energy Indonesia Tbk.	113,230,305	9.189	31.202.699.264
2	Bukit Asam Tbk.	48,962,802	8.268	11.487.200.256
3	Indo Tambangraya Megah Tbk.	44,406,053	4.527	1.129.920.000
4	Harum Energy Tbk.	24.568.389	3.223	13.316.200.448

Sumber: idxchannel.com, 2022

Berdasarkan pada Tabel 1, Adaro memiliki *market capital* yang paling besar diantara perusahaan pertambangan lainnya. *Market Capital* merupakan nilai dari sebuah perusahaan yang didasarkan pada harga pasar yang sedang berlaku saat ini (R. Nugroho, 2022). Semakin tinggi nilai *market capitalization*, maka semakin besar pula perusahaan tersebut menyerap uang investor. Hal ini menunjukkan, dengan menanamkan modal di perusahaan tersebut akan lebih jauh potensial untuk investasi di masa depan. Sejalan dengan *market capital*, frekuensi transaksi dan nilai transaksi saham Adaro juga paling tinggi di antara perusahaan tambang batu bara lainnya. Hal tersebut membuat Adaro masuk ke dalam daftar indeks IDX30 pada tahun 2022 (R. Nugroho, 2022a).

Dukungan fundamental perusahaan yang baik membuat harga saham Adaro terus menguat hingga perdagangan pada tanggal 30 Agustus 2022 menunjukkan peningkatan dan harga sahamnya mendekati rekor tertinggi di mana harga saham ditutup di level 3.540 naik 80 poin atau 2,31% dibandingkan sehari sebelumnya (Suryahadi, 2022a). Nilai laba bersih Adaro juga meningkat pesat sebesar 613% dari laba bersih periode tahun lalu yang hanya sekitar US\$ 169.964 (Suryahadi, 2022b). Berdasarkan pada idxchannel.com Adaro termasuk ke dalam 8 daftar saham pertambangan terbaik di tahun 2022 (R. Nugroho, 2022b).

**Gambar 3**  
**Grafik Harga Saham Adaro Tahun 2021-2022**



Sumber: finance.yahoo.com, 2022

Gambar 3 menunjukkan bahwa saham Adaro juga mengalami fluktuasi. Fluktuasi harga saham yang dialami oleh Adaro membuat investor atau calon investor ingin mengetahui bagaimana prediksi harga saham perusahaan tersebut ke depannya. Salah satu cara untuk mengetahui bagaimana prediksi harga saham tersebut, investor atau calon investor dapat memanfaatkan teknik data mining. Oleh karena itu, penelitian ini memberikan alternatif dengan menggunakan data mining untuk melakukan prediksi penutupan harga saham Adaro dengan menggunakan Algoritma *K-Nearest Neighbor Regression*.

Penelitian terdahulu yang dilakukan oleh Nurfauzan dan Fatimah (2022), menunjukkan hasil pemodelan menggunakan *K-Nearest Neighbor Regression* dengan empat variasi indikator yang berbeda, prediksi dengan menggunakan satu indikator yaitu Atribut *Close* memiliki hasil yang paling baik dengan nilai *root mean squared error* 169 dengan akurasi 98.8% dengan rasio perbandingan 70% data latih dan 30% data uji (Nurfauzan & Fatimah, 2022). Penelitian lain juga dilakukan oleh Tanuwijaya dan Hansun (2019), menunjukkan hasil bahwa aplikasi prediksi indeks saham LQ45 menggunakan *K-Nearest Neighbor Regression* berhasil dibangun dan akurasi tertinggi yang dicapai mencapai 91,81% dengan kode saham WSKT (Tanuwijaya & Hansun, 2019). Pada Penelitian ini akan dilakukan prediksi harga penutupan saham dengan membandingkan penggunaan 11 dan 12 atribut sehingga menjadi pembeda dari penelitian sebelumnya.

Berdasarkan pada latar belakang yang telah dijabarkan, penelitian ini bertujuan untuk membantu para investor dan calon investor saham Adaro untuk dapat memprediksi bagaimana pergerakan harga saham perusahaan di masa yang akan datang dan juga untuk mengetahui tingkat akurasi terhadap penggunaan algoritma K-Nearest Neighbor Regression.

## **II. KAJIAN PUSTAKA DAN PENGEMBANGAN HIPOTESIS**

### **2.1. Efisiensi Pasar Modal**

Efisiensi pasar modal adalah suatu keadaan dimana harga saham menyesuaikan secara cepat dengan adanya tambahan informasi, dan oleh karena itu harga saham tersebut sudah mencakupi seluruh informasi yang ada (Jones, 2013). Dengan kata lain, harga-harga aset di pasar dianggap selalu mencerminkan nilai sebenarnya atau fundamental dari aset tersebut, sehingga tidak mungkin untuk secara konsisten menghasilkan keuntungan yang berlebihan atau abnormal dengan menganalisis informasi publik.

Teori Efisiensi Pasar memiliki implikasi penting dalam investasi dan analisis keuangan. Dalam konteks saham, teori ini mengandaikan bahwa semua informasi publik dan pribadi yang relevan tentang suatu perusahaan atau aset sudah tercermin dalam harga sahamnya. Dengan kata lain, tidak ada informasi yang dapat memberi investor keuntungan yang tidak adil dalam jangka panjang.

### **2.2. Data Mining**

*Data mining* merupakan ekstraksi data dalam jumlah yang besar menjadi suatu informasi atau pengetahuan yang bermanfaat (Mustika et al., 2021). *Data mining* berisi proses pencarian pola yang diinginkan terhadap data yang digunakan untuk membantu proses pengambilan keputusan (Hermawati, 2013). Pola-pola yang terbentuk nantinya akan diproses dengan perangkat tertentu yang kemudian didapatkan suatu analisa data yang berguna dan kemudian dipelajari dengan lebih lanjut untuk memperoleh suatu informasi yang berguna.

Terdapat serangkaian aktivitas tertentu dalam data mining, semuanya melibatkan penggalian informasi baru yang bermakna dari data (Pratiwi et al., 2022). Adapun aktivitas tersebut diantaranya:

a. Uraian

Uraian dilakukan dengan menganalisis secara sederhana yang bertujuan untuk menemukan pola dan kecenderungan terhadap data yang ada.

b. Estimasi

Pertimbangan ini diwujudkan dengan record yang lengkap yang dimana telah tersedia nilai dari variabel nilai prediksi.

c. Prediksi

Prediksi merupakan sebuah teknik kombinasi dari beberapa teknik data mining lainnya yang menghasilkan nilai estimasi di masa yang akan datang.

d. Klasifikasi

Klasifikasi terdiri dari pemeriksaan atribut objek yang baru disajikan dan menetapkan kelas yang telah ditentukan sebelumnya.

e. Asosiasi

Asosiasi memiliki tugas untuk menganalisa data yang muncul, kemudian menemukan atribut yang muncul dalam satu waktu pada data mining.

f. Pengklasteran

Sebuah teknik yang digunakan melakukan pengelompokkan data yang memiliki kesamaan ke dalam sebuah wilayah yang sama.

### 2.3. **Prediksi**

Secara garis besar terdapat dua strategi untuk *data mining*, yaitu Strategi Prediktif dan Strategi Deskriptif (Han & Kamber, 2012). Strategi prediktif merupakan suatu proses yang menggunakan beberapa variabel dalam suatu model untuk melakukan prediksi. Penggunaan strategi ini memiliki tujuan untuk memprediksi hasil di masa depan. Strategi prediktif memiliki 2 metode yang dapat diterapkan yaitu klasifikasi dan regresi (Arhami & Nasir, 2020). Pada penelitian ini, strategi prediktif yang dilakukan akan menggunakan regresi dengan menggunakan algoritma K-Nearest Neighbor Regression. Prediksi dilakukan dengan menggunakan dataset Saham Adaro, dataset harga batu bara, dan dataset kurs mata uang rupiah.

### 2.4. **K-Nearest Neighbor Regression (KNN Regression)**

*KNN Regression* merupakan salah satu algoritma data mining yang digunakan untuk melakukan regresi atau peramalan yang dimana hasil regresi didapatkan dari

rata-rata “k” tetangga terdekat (Kurniawan et al., 2022). Nilai rata-rata tersebut akan menjadi nilai output untuk data baru yang dimasukkan ke dalam algoritma KNN. Jauh dekatnya data tetangga dihitung berdasarkan rumus *Euclidean* berikut ini:

$$d(x_{test,i}, x_{train,j}) = \sqrt{\sum_{i,j}^n (x_{test,i} - x_{train,j})^2} \quad (1)$$

Keterangan:

$d(x_{test,i}, x_{train,j})$	= jarak euclidean
$x_{test,i}$	= data testing ke- $i$
$x_{train,j}$	= data testing ke- $j$
$n$	= banyaknya data
$i, j$	= urutan data ke- 1, 2, 3, ..., $n$

Cara kerja algoritma KNN *Regression* perlu dilakukan penentuan data latih, data uji, dan nilai k. Data latih akan diurutkan berdasarkan hitungan jarak terdekat antara data uji dan data latih. Pada Akhir tahapan, akan diambil rata-rata data latih terkecil sesuai dengan jumlah k yang digunakan untuk menentukan kelas regresi. Berikut ini merupakan alur perhitungan menggunakan algoritma *KNN Regression* (Cahya, 2018).

- Menentukan parameter k (nilai tetangga terdekat)
- Menghitung jarak *Euclidean* (mengukur jarak antara data latih dan data uji)
- Melakukan perangkingan berdasarkan nilai k
- Menentukan jarak yang terdekat sampai nilai k
- Memasangkan kelas yang sesuai
- Mencari jumlah kelas yang terdekat dan menetapkan kelas tersebut untuk dievaluasi

### III. METODE PENELITIAN

Penelitian ini menggunakan jenis data kuantitatif. Sumber data yang digunakan yaitu data sekunder yang datanya tidak didapatkan secara langsung. Data yang digunakan dalam penelitian ini berupa data historis harga saham PT Adaro Energy Indonesia Tbk, data historis harga batu bara, dan data historis kurs mata uang rupiah yang diperoleh dari situs *finance.yahoo.com* dan *investing.com*.

Metode yang digunakan dalam penelitian ini adalah metode *Knowledge Discovery in Databases* (KDD). KDD merupakan salah satu penerapan metode dalam data mining untuk mencari informasi penting dan pola dalam data dengan melibatkan algoritma

untuk mengidentifikasi pola pada data (Syahril et al., 2020). Berikut di bawah ini merupakan tahapan dari metode KDD yang perlu dilakukan secara berurutan (Pratiwi et al., 2022):

a. *Selection*

*Selection* merupakan proses pertama dalam KDD di mana perlu dilakukan seleksi dari kumpulan data harian yang ada (data saham Adaro) sebelum menuju ke tahap penggalan informasi.

b. *Pre-processing/ Cleaning*

*Pre-processing/ Cleaning* dilakukan pada data yang menjadi fokus KDD. Proses ini meliputi pembuangan duplikasi data, mengisi data yang kosong (*data imputation*), pemeriksaan data yang tidak konsisten, dan perbaikan kesalahan pada data. Berikut pada Tabel 2 akan disajikan perintah *code* program untuk mengisi data yang kosong.

**Tabel 2**  
**Code Program Untuk Mengisi Data Kosong**

<b>Code Program:</b>	<code>adro_ml.fillna(adro_ml.mean(), inplace=True)</code>
----------------------	---

Kode program pada Tabel 2 digunakan untuk mengisi nilai yang hilang dalam data dengan nilai rata-rata dari setiap kolom. Fungsi `fillna` digunakan untuk menggantikan nilai yang hilang dalam data dengan nilai rata-rata yang sesuai dari kolom tersebut. Parameter `inplace=True` memastikan bahwa perubahan diterapkan langsung pada data asli tanpa perlu meng-assign hasil kembali ke variabel.

**Gambar 4**  
**Pengecekan Data Kosong**

```

0s ▶ adro_ml.isnull().sum()
adro_high_price    0
adro_low_price     0
adro_open_price    0
coal_price         6
idr_currency_rate  2
is_payday          0
is_dm1_holiday    0
is_dm2_holiday    0
is_dp1_holiday    0
is_dp2_holiday    0
adro_close_price   0
dtype: int64

```

Sumber: Data diolah, 2023

**Gambar 5**  
**Kondisi Data setelah pengisian data**

```

0s ▶ adro_ml.isnull().sum()
adro_high_price    0
adro_low_price     0
adro_open_price    0
coal_price         0
idr_currency_rate  0
is_payday          0
is_dm1_holiday    0
is_dm2_holiday    0
is_dp1_holiday    0
is_dp2_holiday    0
adro_close_price   0
dtype: int64

```

Sumber: Data diolah, 2023

Pada saat dilakukan pengecekan data terlihat bahwa terdapat total 8 (6+2) baris data yang kosong (lihat pada Gambar 4), kemudian setelah dilakukannya pengisian data dengan perintah code program pada Tabel 2 maka semua baris yang kosong akan otomatis terisi sehingga saat dilakukan pengecekan keterangan data kosong sudah tidak ada (lihat pada Gambar 5).

c. *Transformation*

Data yang terpilih harus diubah sesuai dengan algoritma yang akan digunakan dalam data mining. Hal ini dikarenakan terdapat banyak algoritma yang ada dalam data mining di mana setiap algoritmanya memerlukan format data yang berbeda-beda.

d. *Data Mining*

Pada tahap ini, terdapat sebuah proses pencarian pola atau informasi menarik dalam data terpilih dengan menggunakan metode atau algoritma tertentu.

e. *Interpretation/ Evaluation*

Pada tahap ini, pengetahuan atau informasi yang diperoleh perlu disampaikan dan diubah ke dalam bentuk yang lebih mudah dimengerti oleh user seperti ringkasan atau dalam bentuk visualisasi.

f. *Knowledge*

Pada tahap ini pengetahuan dan informasi yang dihasilkan akan diimplementasikan sesuai dengan kegunaan pengetahuan dan informasi tersebut.

#### IV. HASIL DAN PEMBAHASAN

Penelitian ini dimulai dengan melakukan pengumpulan data yaitu data historis harian dari harga saham Adaro sebanyak 1.252 data, data historis harga batu bara sebanyak 1.290 data, dan data historis kurs mata uang rupiah sebanyak 1.303 data. Data tersebut selanjutnya akan diproses dengan menggunakan metode *Knowledge Discovery in Databases* (KDD).

##### 4.1. *Selection*

Pada tahap *selection*, data yang digunakan untuk tujuan analisis ini adalah data historis harian saham PT. Adaro Energy Indonesia Tbk berupa atribut *open* (*adro\_open\_price*), *high* (*adro\_high\_price*), *low* (*adro\_low\_price*), *close* (*adro\_close\_price*),

dan volume (*adro\_volume*). Data historis tersebut diseleksi dari bulan September 2017 hingga bulan Agustus 2022 sebanyak 1252 data. Seleksi data pada historis saham Adaro hanya diseleksi selama 5 tahun terakhir yang digunakan untuk memprediksi harga penutupan di hari berikutnya. Dengan menggunakan data yang mencakup berbagai kondisi, termasuk normal dan tidak normal (Pandemi Covid-19), dapat membantu meningkatkan keandalan model prediksi.

#### 4.2. *Pre-processing/ Cleaning*

Pada tahap *pre-processing/ cleaning*, dilakukan proses penghapusan data yang bernilai *null* yang dimana data tersebut dianggap sebagai data *error* atau tidak sah dikarenakan data tersebut bernilai kosong. Data yang bernilai *null* tersebut terjadi pada tanggal transaksi 19 Juni 2019. Penyebab data tersebut kosong belum diketahui, namun hal tersebut tidak berpengaruh signifikan terhadap penelitian ini sehingga dapat dilakukan penghapusan data pada baris tersebut. Berikut pada Tabel 3 akan ditampilkan bukti data yang bernilai *null* pada tanggal 19 Juni 2019.

**Tabel 3**  
**Data Null Saham Adaro**

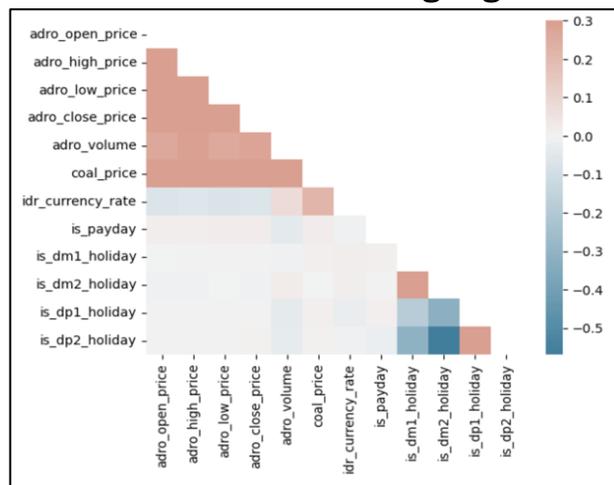
No	Date	Open	High	Low	Close	Volume
1	6/19/2019	null	null	null	null	null

Sumber: [finance.yahoo.com](https://finance.yahoo.com), 2022

Proses penghapusan data yang telah dilakukan, dataset saham Adaro berjumlah 1251 data, kemudian akan dilanjutkan pada proses penambahan atribut harga batu bara (*coal\_price*) dan kurs mata uang rupiah (*idr\_currency\_rate*) dengan mengintegrasikan data historis saham Adaro dengan data historis harga batu bara dan juga data historis kurs mata uang rupiah. Selain atribut tersebut, dilakukan penambahan beberapa atribut lagi seperti *atribut is\_payday*, *is\_dm1\_holiday*, *is\_dm2\_holiday*, *is\_dp1\_holiday*, dan *is\_dp2\_holiday*. Penambahan atribut tersebut akan diisi dengan data bernilai 1 (*true*) dan 0 (*false*) yang digunakan untuk melihat kondisi dari suatu hari ketika akan melakukan suatu prediksi harga saham pada hari tertentu. Kondisi hari tersebut diperlukan untuk menentukan apakah hari tersebut hari gajian (*is\_payday*), 1 hari sebelum hari libur (*is\_dm1\_holiday*), 2 hari sebelum hari libur (*is\_dm2\_holiday*), 1 hari setelah hari libur (*is\_dp1\_holiday*), dan 2 hari setelah hari

libur (*is\_dp2\_holiday*). Penambahan atribut tersebut juga didasarkan atas hubungan korelasi antara atribut-atribut lainnya. Pada Gambar 6 di bawah ini akan ditampilkan grafik *heatmap* yang menyatakan korelasi antara atribut-atribut yang digunakan.

**Gambar 6**  
**Hasil Korelasi Atribut Yang Digunakan**



Sumber: Data diolah, 2023

Berdasarkan pada Gambar 6 berupa gambar *heatmap*, grafik tersebut menjelaskan bahwa semakin gelap atau pekat warna dari kolom yang menghubungkan dua atribut maka kedua atribut tersebut berkorelasi semakin kuat. Terlihat bahwa semua atribut yang digunakan untuk penelitian ini saling berkorelasi. Korelasi yang terjadi antara atribut berpotensi mempengaruhi harga saham dengan begitu atribut-atribut tersebut digunakan pada penelitian ini.

Berikut ini pada Tabel 4 akan disajikan hasil dari *pre-processing/cleaning* data yang telah dilakukan.

**Tabel 4**  
**Data Historis Setelah *Pre-processing/Cleaning***

<b>Data Historis</b>	<b>Sebelum <i>Preprocessing/Cleaning</i></b>	<b>Setelah <i>Preprocessing/Cleaning</i></b>
Saham Adaro	1252	1251
Harga Batu Bara	1290	1251
Kurs Mata Uang Rupiah	1303	1251

Sumber: Data diolah, 2023

Proses pengurangan data pada saat *pre-processing/ cleaning* dilakukan dengan menselaraskan jumlah data harga batu bara dan kurs mata uang rupiah dengan jumlah dari data Saham Adaro.

#### **4.3. Transformation**

Proses ini akan dilakukan perubahan terhadap bentuk data yang belum memiliki format yang sesuai dengan algoritma yang akan digunakan. Namun, pada penelitian ini transformasi data tidak dilakukan karena bentuk data yang digunakan sudah memiliki format yang sesuai dengan algoritma yang akan digunakan.

#### **4.4. Data Mining**

*Data mining* dalam penelitian ini menggunakan algoritma *K-Nearest Neighbor Regression* dan *tools* yang digunakan adalah *Google Colaboratory*. Dalam proses ini juga akan dilakukan perbandingan penggunaan atribut yang berbeda yaitu penggunaan 11 atribut dan 12 atribut terhadap tingkat akurasi dan prediksi harga penutupan saham yang dihasilkan. Adapun perbandingan *data training* dan *data testing* yang digunakan adalah sebesar 90% dan 10%. Dari proses *data mining* ini menghasilkan nilai  $k$  optimal pada 11 atribut yaitu nilai  $k = 2$  sedangkan pada 12 atribut nilai  $k = 33$ .

#### **4.5. Interpretation & Evaluation**

Pada tahap *Interpretation & Evaluation*, evaluasi performa model yang telah dilatih sebelumnya dengan menggunakan data testing. Dalam hal ini, terdapat beberapa metrik evaluasi seperti *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), *R-squared* ( $R^2$ ), dan *Explained Variance Score* (EVS) untuk mengevaluasi performa model.  $R^2$  dan EVS digunakan untuk mengukur sejauh mana variasi harga saham dapat dijelaskan oleh model prediksi. Tabel 5 merupakan hasil evaluasi model dari pengujian data menggunakan 11 atribut yang menghasilkan nilai  $k$  optimal bernilai 2 dan 12 atribut yang menghasilkan nilai  $k$  optimal bernilai 33.

Berdasarkan pada penggunaan metrik evaluasi model yang terdapat pada Tabel 5, terdapat hasil evaluasi model dari penggunaan 11 atribut dan 12 atribut yang telah dilakukan pengujian data dengan melakukan percobaan menggunakan 3 macam komposisi *data training* dan *data testing* yaitu 90:10, 80:20, dan 70:30. Semakin kecil nilai RMSE, MAE maka semakin baik pula performa model yang dihasilkan. Pada EVS dan  $R^2$ , semakin mendekati 1 nilainya maka performa model semakin baik. Terlihat bahwa evaluasi model terbaik terdapat pada komposisi *data training* dan *data testing*

dengan perbandingan (90:10). Penggunaan 4 metrik evaluasi model mencerminkan pengaruh dari atribut-atribut yang digunakan untuk memprediksi harga penutupan saham Adaro. Hasil evaluasi dari penggunaan atribut-atribut tersebut menunjukkan bahwa evaluasi prediksi harga penutupan saham Adaro dengan menggunakan 11 atribut memiliki hasil evaluasi yang lebih baik dibandingkan dengan menggunakan 12 atribut yang di dalamnya terdapat Atribut Volume.

**Tabel 5**  
**Hasil Evaluasi Model Menggunakan 11 Atribut dan 12 Atribut**

Evaluasi Model	11 Atribut			12 Atribut		
	(90:10)	(80:20)	(70:30)	(90:10)	(80:20)	(70:30)
<b>RMSE</b>	35,02	36,27	36,72	517,02	537,37	535,64
<b>R<sup>2</sup></b>	0,99	0,99	0,99	0,02	0,12	0,16
<b>EVS</b>	0,99	0,99	0,99	0,04	0,13	0,16
<b>MAE</b>	24,54	24,15	25,52	411,89	425,77	424,52

Sumber: Data diolah, 2023

#### 4.6. Prediksi

Model yang telah dilatih sebelumnya lalu digunakan untuk memprediksi nilai pada data baru. Prediksi harga penutupan saham "adro\_close\_price" menggunakan method "*predict*" pada model yang telah dilatih untuk melakukan prediksi. Berikut ini pada Tabel 6 disajikan lima data perbandingan harga penutupan saham aktual dan prediksi menggunakan 11 atribut dan 12 atribut yang telah melalui proses data mining menggunakan algoritma *K-Nearest Neighbor Regression*.

**Tabel 6**  
**Data Historis Setelah Pre-processing/Cleaning**

Tanggal	Aktual	Prediksi	
		11 Atribut	12 Atribut
29 Agustus 2022	Rp 3.460	Rp 3.410	Rp 1.570,45
30 Agustus 2022	Rp 3.540	Rp 3.435	Rp 1.491,82
31 Agustus 2022	Rp 3.540	Rp 3.485	Rp 1.811,67
1 September 2022	Rp 3.700	Rp 3.410	Rp 1.769,24
2 September 2022	Rp 3.780	Rp 3.485	Rp 1.869,24

Sumber: Data diolah, 2023

Berdasarkan pada Tabel 6, data antara harga aktual dibandingkan dengan harga prediksi untuk mengetahui apakah hasil prediksi dengan harga aktual tidak terdapat perbedaan yang signifikan untuk menunjukkan validitas pendekatan ini dalam memprediksi harga saham.

#### **4.7. Knowledge**

Berdasarkan pada penelitian yang telah dilakukan pada tahapan sebelumnya, ditemukan bahwa penggunaan atribut-atribut dalam melakukan prediksi akan mempengaruhi model prediktif yang dibentuk dan tentunya hal tersebut juga akan mempengaruhi hasil prediksi dari penutupan harga saham yang dihasilkan. Hal ini terlihat jelas pada hasil evaluasi penggunaan 11 atribut dan 12 atribut untuk melakukan prediksi penutupan harga saham pada Adaro yang telah ditampilkan pada Tabel 6.

## **V. SIMPULAN, KETERBATASAN DAN SARAN**

Berdasarkan penelitian dan pengujian dataset dalam prediksi penutupan harga saham Adaro, dengan pengujian data sebanyak 1251 data dari tahun 2017 sampai dengan tahun 2022 menggunakan 11 atribut dan 12 atribut yang diuji. Sebanyak 1251 data dalam penelitian ini dibagi menjadi data training dan data testing dengan perbandingan masing-masing yaitu 90% (1.125 data training) dan 10% (126 data testing). Kesimpulan pada penelitian ini adalah sebagai berikut:

- a. Analisa dan prediksi pada penutupan harga saham PT. Adaro Energy Indonesia Tbk. menggunakan algoritma K-Nearest Neighbor Regression dengan melakukan pengujian 11 atribut dan 12 atribut memiliki hasil prediksi harga penutupan saham yang berbeda yang dimana hal tersebut dapat dilihat dari nilai RMSE, R<sup>2</sup>, EVS dan MAE. Hasil prediksi menggunakan 11 atribut dinilai lebih baik dibandingkan dengan menggunakan 12 atribut. Hal ini dikarenakan, di dalam 12 atribut tersebut terdapat atribut volume yang mempengaruhi hasil akurasi yang dihasilkan. Atribut volume mempengaruhi hasil akurasi karena volume cenderung memiliki variasi yang besar dan berkisar di dalam rentang yang lebar. Variasi yang besar dalam atribut tersebut menyebabkan prediksi menjadi kurang akurat.

b. Hasil penelitian menunjukkan bahwa algoritma KNN Regression dengan menggunakan 11 atribut mampu menghasilkan nilai RMSE sebesar 35,02,  $R^2 = 0,99$ , EVS = 0,99 dan MAE = 24,54. Nilai-nilai tersebut menunjukkan bahwa tingkat keakuratan untuk melakukan prediksi penutupan harga saham lebih baik dibandingkan dengan penggunaan 12 atribut yang dimana terdapat atribut volume di dalamnya.

Terdapat beberapa saran yang dapat diberikan setelah dilakukannya penelitian ini terhadap pengembangan penelitian berikutnya dalam memprediksi penutupan harga saham seperti di bawah ini:

- a. Pentingnya memilih atribut yang paling relevan dan signifikan dalam memprediksi target variabel, dan melakukan evaluasi performa model untuk memastikan hasil yang didapat sudah optimal.
- b. Melakukan pembuatan atribut baru dari kombinasi atribut yang ada dan juga relevan untuk memperbaiki kinerja model yang dapat memberikan hasil yang lebih akurat dan performa yang lebih baik dalam memprediksi target variabel.
- c. Pada penelitian berikutnya dapat menambahkan beberapa algoritma lainnya seperti LSTM (Long Short-Term Memory), Neural Networks, Extreme Gradient Boost, dan lainnya sebagai perbandingan dalam melakukan prediksi.

#### DAFTAR PUSTAKA

- Arhami, M., & Nasir, M. (2020). *DATA MINING Algoritma dan Implementasi*. ANDI.
- Astutik, E. P., Retnosari, & Nilasari, A. P. (2021). Analisis Determinan Harga Saham Perusahaan Food and Beverage Yang Terdaftar di Bursa Efek Indonesia. *Edunomika*, 5(2), 1099–1107.
- Cahya. (2018). *Contoh Implementasi Data Mining Algoritma K-Nearest Neighbors (k-NN) Menggunakan PHP dan MySQL Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu*. <https://extra.cahyadsn.com/knn>
- CNBC. (2022). *Pergerakan Harga Batu Bara Menyulut Harga Saham Emitennya*. <https://www.cnbcindonesia.com/market/20220621235153-17-349147/pergerakan-harga-batu-bara-menyulut-harga-saham-emitennya>
- Dihni, V. (2022). *Realisasi Investasi Indonesia Naik Terus dalam 5 Tahun Terakhir*.
- Han, J., & Kamber, M. (2012). *Data Mining Concepts and Techniques* (3rd ed.). Morgan Kauffan.

- Hermawati, F. A. (2013). Data mining. In *Data Mining*. CV ANDI OFFSET.
- Jones, C. P. (2013). *Investments Analysis and Management (12th ed.)*. Wiley.
- KSEI. (2022a). *Statistik Pasar Modal Indonesia*.
- KSEI. (2022b). Statistik Pasar Modal Indonesia\_September 2022. In *Publikasi PT Kustodian Sentral Efek Indonesia* (Issue September).
- Kurniawan, C. C., Rostianingsih, S., & Santoso, L. W. (2022). Penerapan Metode KNN-Regresi dan Multiplicative Decomposition untuk Prediksi Data Penjualan pada Supermarket X. *Jurnal INFRA*, 10(2).
- Mustika, Ardilla, Y., Manuhutu, A., Ahmad, N., Hasbi, I., Guntoro, Manuhutu, A. M., Ridwan, M., & Hozairi. (2021). *Data Mining dan Aplikasinya*.
- Nugroho, R. (2022a). *Simak Daftar Saham IDX30 Periode Agustus 2022-Januari 2023, Ada 3 Saham Baru*. <https://www.idxchannel.com/market-news/simak-daftar-saham-idx30-periode-agustus-2022-januari-2023-ada-3-saham-baru>
- Nugroho, R. (2022b). *Simak Delapan Daftar Saham Pertambangan Terbaik 2022 yang Bisa Anda Beli*. <https://www.idxchannel.com/market-news/simak-delapan-daftar-saham-pertambangan-terbaik-2022-yang-bisa-anda-beli>
- Nugroho, R. (2022). *Apa Itu Market Cap? Istilah yang Tak Bisa Dilewatkan dalam Investasi Saham*. <https://www.idxchannel.com/market-news/apa-itu-market-cap-istilah-yang-tak-bisa-dilewatkan-dalam-investasi-saham>
- Nurfauzan, D., & Fatimah, T. (2022). Implementasi Algoritma K-Nearest Neighbors Regression Dalam Memprediksi Harga Saham Implementasi Algoritma K-Nearest Neighbors Regression Dalam Memprediksi Harga Saham. *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI), September*, 164–173.
- Pratiwi, W., Utami, N., & Putra, G. (2022). Klasifikasi Penentuan Penerima Bantuan Sosial Tunai (BST) Menggunakan Algoritma C4.5 Di Desa Keramas, Gianyar, Bali. *Jurnal Informatika, Teknologi Dan Sains*, 4(3), 101–107. <https://doi.org/10.51401/jinteks.v4i3.1667>
- Pahlevi, R. (2022). *Ini Emiten Batu Bara dengan Produksi Terbesar di Indonesia*. <https://databoks.katadata.co.id/datapublish/2022/05/10/ini-emiten-batu-bara-dengan-produksi-terbesar-di-indonesia>
- Rianti, A. (2021). *Analisis Dampak Covid-19 Pada Volume Transaksi Saham dan Harga Saham (Studi Kasus Pada Kelompok Perusahaan Transportasi yang Terdaftar di Bei Tahun 2020)*. 1(3), 171–178.
- Suryahadi, A. (2022a). *Harga Saham ADRO Naik Dekati Rekor Tertinggi*.
- Suryahadi, A. (2022b). *Kinerja Adaro Energy (ADRO) di Atas Ekspektasi, Simak Rekomendasi Analis*.

- Syahril, M., Erwansyah, K., & Yetri, M. (2020). Penerapan Data Mining Untuk Menentukan Pola Penjualan Peralatan Sekolah Pada Brand Wigglo Dengan Menggunakan Algoritma Apriori. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 3(1), 118. <https://doi.org/10.53513/jsk.v3i1.202>
- Tanuwijaya, J., & Hansun, S. (2019). LQ45 Stock Prediction Using K-Nearest Neighbor Regression. *International Journal of Recent Technology and Engineering*, 8(3), 2388–2391. <https://doi.org/10.35940/ijrte.C4663.098319>
- Wardiyah, M. (2017). *Manajemen Pasar Uang dan Pasar Modal*. CV Pustaka Setia.