

# Prediksi Harga Kopi Robusta Kabupaten Muara Enim Tahun 2025 Menggunakan Metode Random Forest, SVM & Tree

Yunita<sup>1\*</sup>, Nerma Yulisa<sup>2</sup>, Diajeng Putri Sekar Mayasari<sup>3</sup>, Pujiyanto<sup>4</sup>

<sup>1</sup>Informatika/Teknik dan Komputer  
Universitas Baturaja  
<sup>1\*</sup>oktaviayunita05@gmail.com

<sup>2</sup>Informatika/Teknik dan Komputer  
Universitas Baturaja  
<sup>2</sup>nermayulisa534@gmail.com

<sup>3</sup>Informatika/Teknik dan Komputer  
Universitas Baturaja  
<sup>3</sup>putrisekarmayasari@gmail.com

<sup>4</sup>Informatika/Teknik dan Komputer  
Universitas Baturaja  
<sup>4</sup>pujiyanto.mail@gmail.com

**Abstrak**— Penelitian ini bertujuan untuk memprediksi harga kopi robusta di Kabupaten Muara Enim pada tahun 2025 dengan menggunakan pendekatan data mining berbasis algoritma Random Forest, Support Vector Machine (SVM), dan Decision Tree. Harga kopi sebagai salah satu komoditas unggulan Indonesia mengalami fluktuasi yang signifikan sehingga dibutuhkan metode prediktif yang akurat guna membantu petani, pengusaha, dan pemangku kebijakan dalam menentukan strategi pasar. Penelitian dilakukan dengan mengumpulkan data harga kopi dari bulan Januari hingga April 2025 yang diperoleh langsung dari para pengumpul kopi di Muara Enim. Data tersebut diolah menggunakan perangkat lunak Orange Data Mining dan dibagi menjadi data training dan testing. Hasil analisis menunjukkan bahwa ketiga algoritma memiliki tingkat akurasi yang bervariasi, di mana algoritma Random Forest memberikan hasil prediksi paling stabil. Penelitian ini membuktikan bahwa penggunaan teknologi machine learning dapat membantu dalam meramalkan harga komoditas secara efektif.

**Kata Kunci**— Prediksi, Data Mining, Software Orange, Kopi Robusta, Random Forest, SVM, Decision Tree.

**Abstract**— This research aims to predict the price of robusta coffee in Muara Enim Regency for the year 2025 using a data mining approach based on the *Random Forest*, *Support Vector Machine (SVM)*, and *Decision Tree* algorithms. Coffee, as one of Indonesia's top export commodities, experiences significant price fluctuations, necessitating accurate predictive methods to assist farmers, entrepreneurs, and policymakers in making market strategy decisions. The study collected coffee price data from January to April 2025, obtained directly from local coffee collectors in Muara Enim. The dataset was processed using *Orange Data Mining* software and divided into training and testing sets. The analysis results show that each algorithm exhibits different accuracy levels, with the Random Forest algorithm producing the most stable predictions. This research demonstrates that machine learning technology can effectively support commodity price forecasting.

**Keywords**— Prediction, Mining Data, Software Orange, Robusta Coffee, Random Forest, SVM, Decision Tree.

## I. PENDAHULUAN

Kopi adalah minuman berkafein yang banyak disukai masyarakat. Tak jarang, orang-orang mengonsumsi kopi di pagi, siang, dan sore untuk meningkatkan energi mereka dan menghilangkan kantuk. Kopi menjadi salah satu komoditas ekspor unggulan yang dikembangkan di Indonesia karena mempunyai nilai ekonomis yang relative tinggi di pasaran dunia. Permintaan kopi Indonesia dari waktu ke waktu terus meningkat[5].

Sebagian besar wilayah Indonesia adalah wilayah lahan pertanian, sehingga Negara Indonesia terkenal dengan Negara agraris. Potensi pendapatan sumber daya terbesar didapatkan dari sektor pertanian. Perkebunan, sebagai salah satu subsector penting dalam sektor pertanian, mempunyai kontribusi yang signifikan terhadap perekonomian Indonesia. Kopi arabika tergolong salah satu komoditas unggulan didalam subsector perkebunan di Indonesia karena memiliki peluang pasar yang baik di dalam negeri maupun luar negeri. Produksi kopi Indonesia telah mencapai 600 ribu ton pertahunnya dan lebih dari 80 persen berasal dari perkebunan rakyat. Fluktuasi harga kopi yang terjadi tidak terlepas dari perkembangan produksi kopi dunia karena adanya faktor penyebab. Pesatnya perkembangan produksi kopi dunia telah menyebabkan terjadinya pasokan kopi dunia sehingga mengakibatkan persaingan antara negara produsen semakin ketat dan pada akhirnya harga cenderung tertekan. Metode time series sebagai metode peramalan, data time series pada umumnya untuk mengetahui bentuk pola variasi dimasa lalu dan mengetahui nilai di masa yang akan datang. Dengan melakukan proses determinasi metode time series memiliki kemampuan prediksi yang tinggi, dengan meninjau nilai di masa lalu maka nilai yang akan datang dapat diketahui. Signal yang kuat pada komponen determinan maka nilai yang akan datang dapat diketahui.

### 1. Data mining.

Teknik yang dibuat didalam data mining yaitu salah satunya bagaimana menelusuri

data yang ada untuk membentuk sebuah model, kemudian menggunakan model

tersebut bisa dapat mengenali pola data yang lain yang tidak berada dalam basis data yang sudah tersimpan. Teknik seperti ini juga dapat dimanfaatkan untuk kebutuhan prediksi. Dalam data mining, pengategorisasian data juga bisa dilakukan. Bertujuan agar kita dapat mengetahui pola universal data-data yang ada[4].

### 2. Prediksi

Prediksi adalah suatu proses sistematis untuk memperkirakan kejadian atau kondisi yang akan terjadi di masa depan dengan menggunakan data, informasi, atau pengetahuan yang tersedia saat ini. Prediksi dilakukan melalui analisis terhadap pola-pola atau tren masa lalu, baik secara kuantitatif maupun kualitatif, dengan tujuan untuk memberikan gambaran kemungkinan hasil yang belum terjadi. Prediksi bukan hanya sekadar "menebak", tetapi merupakan proses ilmiah yang melibatkan penggunaan metode logis, statistik, atau algoritmik untuk membuat proyeksi berdasarkan bukti yang ada.

### 3. Software Orange

Orange adalah teknologi pembelajaran mesin open source atau perangkat lunak penambangan data. Orange dapat digunakan untuk analisis dan visualisasi data eksploratif. Ini memberikan platform untuk pemilihan eksperimen, pemodelan prediktif, dan sistem rekomendasi dan dapat digunakan untuk penelitian genomik, biomedis, bioinformatika, dan pengajaran. Orange selalu disukai jika faktor inovasi, kualitas, atau keandalannya terlibat. Orange mempermudah pemakai bermain dengan data open source serta melaksanakan proses data analytics secara intuitif. Pada permasalahan riset ini Orange Data Mining menunjukkan sebagian widget untuk mencari data informasi kata yang dominan timbul dari konten status serta pendapat/komentar account twitter yang hendak menciptakan

tampilan word cloud dari widget Orange Data Mining[2]. Teknik datamining membantu dalam menemukan pengetahuan tersembunyi dalam tim data yang dapat digunakan untuk menganalisis dan memprediksi perilaku di masa depan. Klasifikasi adalah salah satu metode penambangan catatan yang menetapkan label kelas ke sekumpulan kasus yang tidak diklasifikasikan.

#### 4. Dataset

Dataset merupakan kumpulan dari objek dan sifat atau karakteristik dari suatu objek itu sendiri (atribut). Dataset terdiri dari sekumpulan file atau dokumen. Pada statistik, dataset biasanya berasal dari hasil pengamatan aktual yang diperoleh dengan cara mengambil sampel populasi. Dataset dapat menghasilkan algoritma yang bisa dipakai dalam pengujian pada perangkat lunak.

#### 5. Kopi Arabika

Kopi Arabika (nama ilmiah: *Coffea arabica*) adalah salah satu jenis kopi paling populer dan paling banyak dikonsumsi di dunia, dikenal karena cita rasanya yang halus, asam yang cerah, dan aroma yang kompleks. Kopi Arabika menyumbang sekitar 60–70% dari total produksi kopi dunia.

Jenis kopi ini berasal dari wilayah pegunungan di Etiopia dan dipercaya sebagai spesies kopi pertama yang dibudidayakan oleh manusia. Kopi Arabika dianggap memiliki kualitas yang lebih tinggi dibandingkan jenis lainnya seperti *Robusta* (*Coffea canephora*).

#### 6. Random forest

Random Forest merupakan algoritma dalam machine learning yang digunakan untuk pengklasifikasian dataset dalam jumlah besar. Karena fungsinya bisa dipakai dengan banyak dimensi dan dengan berbagai skala dan performa yang tinggi. Random Forest diimplementasikan untuk berbagai sektor seperti perbankan, kesehatan, analisis keuangan, e-commerce dan sebagainya. Cara kerja

algoritma ini adalah dengan membuat beberapa pohon keputusan dan menyartukannya untuk memperoleh prediksi yang lebih stabil serta akurat. Hutan (Forest) yang dibangun oleh algoritma tersebut merupakan kumpulan pohon keputusan dimana biasanya dilatih dengan metode bagging, yang diharapkan menghasilkan model machine learning yang lebih baik.

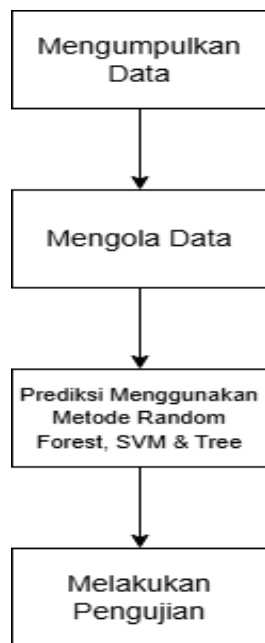
#### 7. SVM (Support Vector Machine)

Support vector machine merupakan salah satu metode regresi atau mengklasifikasi data berdasarkan data-data sebelumnya dan pemodelannya disupervisi terlebih dahulu. SVM termasuk ke dalam jenis klasifikator yang biner, linier dan non probabilistic. Support Vector Machine menggunakan decision boundary (batas keputusan) yang akan menentukan klasifikasi dari data-data pelatihan, sehingga dapat dibentuk sebuah model linier atau hyperplane yang paling optimal untuk mengklasifikasi data.

#### 8. Decision Tree

Decision tree adalah algoritma machine learning yang menggunakan seperangkat aturan untuk membuat keputusan dengan struktur seperti pohon yang memodelkan kemungkinan hasil, biaya sumber daya, utilitas dan kemungkinan konsekuensi atau resiko. Konsepnya adalah dengan cara menyajikan algoritma dengan pernyataan bersyarat, yang meliputi cabang untuk mewakili langkah-langkah pengambilan keputusan yang dapat mengarah pada hasil yang menguntungkan. Klasifikasi ini menggunakan observasi pada node untuk menemukan target pada leaves. Decision Tree merupakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia dengan kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih simple.

## II. METODOLOGI PENELITIAN



Gambar 1. Discovery in Database (KDD)

### 1. Mengolah Data

Beberapa instrumen penting digunakan pada penelitian ini. Pertama, data riset yang akan dimasukkan dan diolah terdiri dari riset pada toko di kabupaten muara enim tersebut. Kedua, data tersebut sudah tersedia dalam format Microsoft Excel. Dataset ini mencakup 320 harga kopi, baik grade A,B dan C yang diambil secara aktual dari tanggal 1 Januari 2025 sampai dengan 16 April 2025. Ketiga, perangkat lunak yang digunakan untuk pengolahan data adalah Microsoft Excel 2010. Pengolahan data sampel yang identik dengan populasi dilakukan dengan menggunakan perangkat Microsoft Excel. Langkah ini dimaksudkan untuk mendapatkan hasil pengujian data. Selain itu, aplikasi khusus digunakan untuk mengelola dan menguji data uji. Selanjutnya, data tersebut diolah dan

dianalisis sesuai dengan prosedur dan teknik pengolahan data yang telah ditentukan. Proses pengolahan data meliputi beberapa langkah. Pertama,

dilakukan pemilihan dan penyusunan klasifikasi data agar dapat dikategorikan dengan baik. Selanjutnya, setiap data diberikan kode-kode tertentu guna mempermudah analisis lebih lanjut. Terakhir, dilakukan analisis data sesuai dengan klasifikasinya yang telah ditetapkan sebelumnya.

### 2. Prediksi Menggunakan Random Forest, SVM & Tree

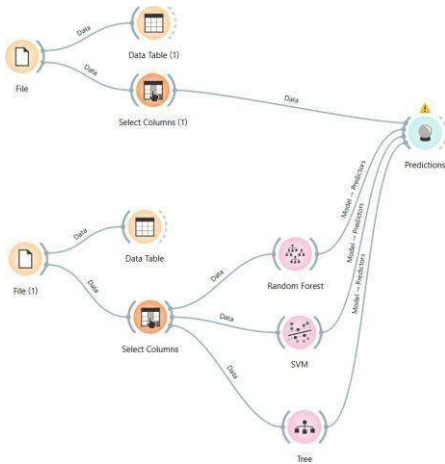
Tahapan ini merupakan inti dari metodologi, yaitu bagaimana data diolah dan dianalisis. Dalam penelitian prediksi harga kopi:

- Data dianalisis menggunakan perangkat lunak Orange Data Mining.
- Algoritma yang digunakan adalah:
  - Random Forest: Membangun banyak pohon keputusan dan menggabungkannya untuk hasil prediksi yang lebih akurat.
  - SVM (Support Vector Machine): Mengklasifikasikan data historis harga kopi dan membentuk hyperplane terbaik untuk prediksi.
  - Decision Tree: Membentuk model berbasis aturan-aturan logis seperti pohon untuk memetakan harga.

### 3. Pengujian

Hasil prediksi yang dilakukan oleh algoritma Random Forest, SVM & Decision Tree selanjutnya dilakukan pengujian untuk mengukur kinerja dari prediksi yang telah dihasilkan. Pengukuran kinerjanya melalui accuracy, precision, recall, dan F1 Score.

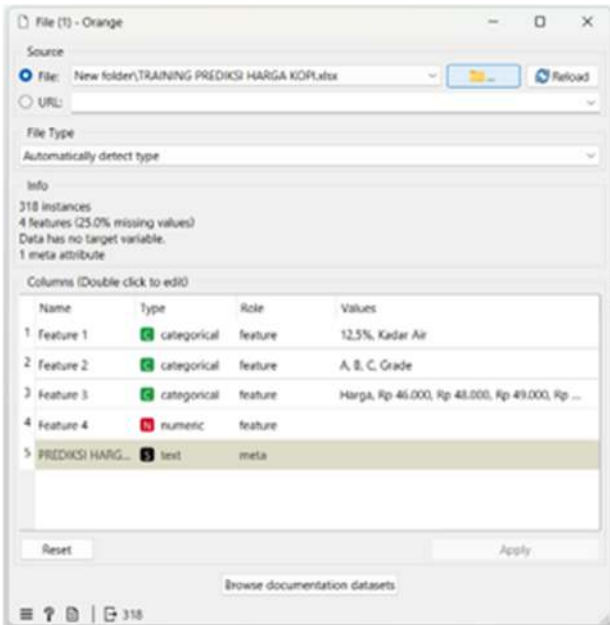
### III. HASIL DAN PEMBAHASAN



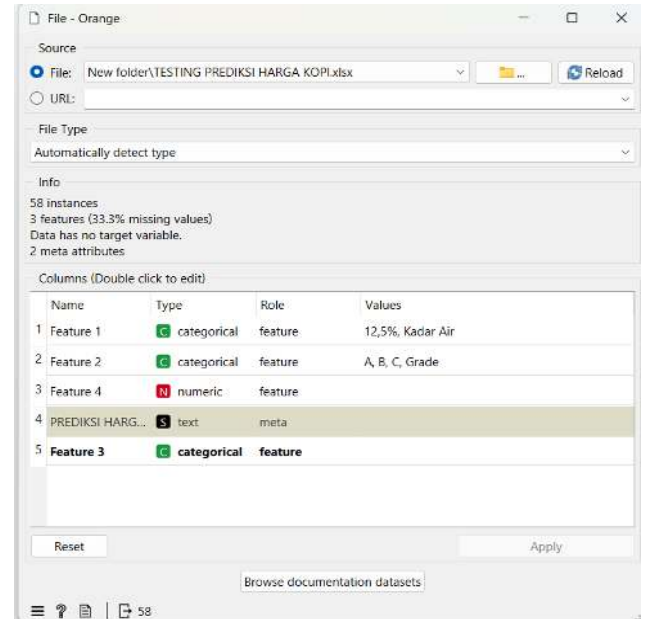
Gambar 2. Model Penelitian

Hasil penelitian dituliskan berdasarkan tahapan penelitian dari langkah awal penelitian hingga akhir seperti pada gambar 2.

#### 1. Data File



Gambar 3. File Data Training



Gambar 4. File Data Testing

Pada Gambar 3 & Gambar 4 melakukan input terhadap data hipertensi ke dalam widget file yang dimana datanya terdiri dari 318 instance dan 4 feature untuk gambar 3 dan 59 instance dan 3 feature untuk gambar 4.

#### 2. Data Tabel

##### 2.1 Tabel Training

Tanggal	Kadar Air	Grade	Harga
2025-01-01 00:00:00	12,5%	A	Rp 68.000
2025-01-01 00:00:00	12,5%	B	Rp 62.000
2025-01-01 00:00:00	12,5%	C	Rp 55.000
2025-01-02 00:00:00	12,5%	A	Rp 68.000
2025-01-02 00:00:00	12,5%	B	Rp 62.000
2025-01-02 00:00:00	12,5%	C	Rp 55.000
2025-01-03 00:00:00	12,5%	A	Rp 68.000
2025-01-03 00:00:00	12,5%	B	Rp 62.000
2025-01-03 00:00:00	12,5%	C	Rp 55.000
2025-01-04 00:00:00	12,5%	A	Rp 68.000
2025-01-04 00:00:00	12,5%	B	Rp 62.000
2025-01-04 00:00:00	12,5%	C	Rp 55.000
.....	.....	.....	.....
2025-04-13 00:00:00	12,5%	C	Rp 48.000
2025-04-14 00:00:00	12,5%	A	Rp 62.000
2025-04-14 00:00:00	12,5%	B	Rp 56.000
2025-04-14 00:00:00	12,5%	C	Rp 49.000
2025-04-15 00:00:00	12,5%	A	Rp 64.000
2025-04-15 00:00:00	12,5%	B	Rp 58.000
2025-04-15 00:00:00	12,5%	C	Rp 51.000
2025-04-16 00:00:00	12,5%	A	Rp 65.000
2025-04-16 00:00:00	12,5%	B	Rp 59.000

Gambar 5. Tabel Data Training

### 2.2 Tabel Testing

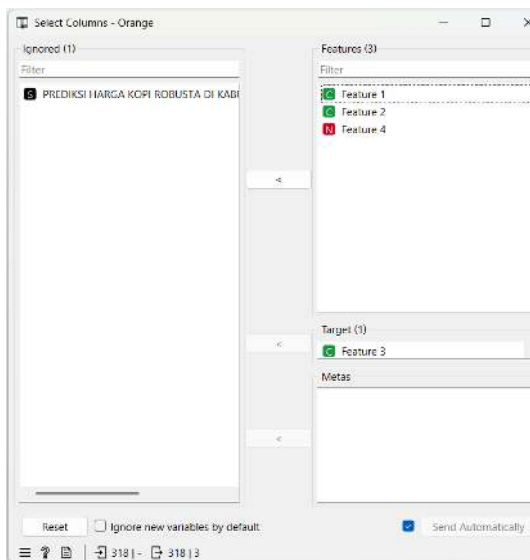
2025-04-14 00:00:00	12,5%	A	?	Rp. 62.000
2025-04-14 00:00:00	12,5%	B	?	Rp. 56.000
2025-04-14 00:00:00	12,5%	C	?	Rp. 49.000
2025-04-14 00:00:00	12,5%	A	?	Rp. 64.000
2025-04-15 00:00:00	12,5%	B	?	Rp. 58.000
2025-04-15 00:00:00	12,5%	C	?	Rp. 51.000
2025-04-16 00:00:00	12,5%	A	?	Rp. 65.000
2025-04-16 00:00:00	12,5%	B	?	Rp. 59.000
2025-04-16 00:00:00	12,5%	C	?	?
2025-04-17 00:00:00	12,5%	A	?	?
2025-04-17 00:00:00	12,5%	B	?	?
2025-04-17 00:00:00	12,5%	C	?	?

Gambar 6. Tabel Data Testing

Datatable training tempat untuk melihat data yang sudah di input kedalam data file, Gambar 5 adalah bentuk data hipertensi yang digunakan dalam penelitian. Sedangkan Datatable testing tempat untuk melihat data yang sudah di input kedalam data file, Gambar 6 adalah bentuk data hipertensi yang digunakan dalam penelitian.

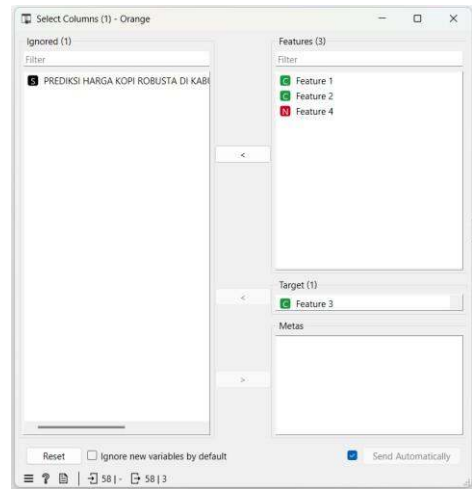
### 3. Data column/target

#### 3.1 Data column Training



Gambar 7. Tabel Column Training

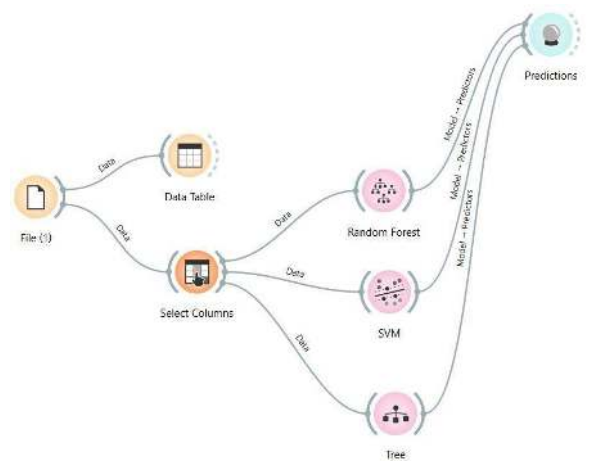
### 3.2 Data column Testing



Gambar 7. Tabel Column Testing

Columns akan dilakukan pengambilan satu features yang akan dijadikan sebagai target variable, selanjutnya akan dilakukannya pengujian terhadap data tersebut.

### 4. Algoritma Random Forest, SVM & Tree



Gambar 8. Tabel Column Testing

Pada gambar 8 dilakukan penghubungan data dari select columns ke Random Forest, Support Vector Machine (SVM) & Tree. Selanjutnya akan dihubungkan ke dalam proses predictions. Dalam proses predictions membutuhkan dua data yang sudah dilakukan perhitungan algoritma dan datayang belum dilakukan perhitungan algoritma oleh karena itu

peneliti menghubungkan data dari select columns ke proses predictions untuk melihat hasilnya.

menyusun strategi pasar yang lebih adaptif terhadap fluktuasi harga komoditas.

### 5. Predictions

Gambar 9. Predictions

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	0.666	0.042	0.006	0.003	0.042	0.007
SVM	0.570	0.042	0.023	0.022	0.042	0.008
Tree	0.637	0.042	0.024	0.023	0.042	-0.006

Gambar 10. Hasil Perhitungan Predictions

Gambar 9 dan Gambar 10 merupakan hasil prediction yang sudah di hitung melalui aplikasi Orange.

### KESIMPULAN

Penelitian ini menunjukkan bahwa pendekatan *data mining* menggunakan algoritma Random Forest, Support Vector Machine (SVM), dan Decision Tree efektif dalam memprediksi harga kopi robusta di Kabupaten Muara Enim tahun 2025. Dari ketiga algoritma yang diuji, Random Forest terbukti memberikan hasil prediksi yang paling stabil dan akurat dibandingkan dengan SVM dan Decision Tree. Hasil ini diperoleh melalui analisis data harga kopi yang dikumpulkan dari bulan Januari hingga April 2025, dan diproses menggunakan perangkat lunak Orange Data Mining. Teknologi ini memberikan kemudahan dalam analisis visual serta mendukung proses klasifikasi dan prediksi secara efisien. Dengan demikian, implementasi metode ini diharapkan dapat membantu petani, pelaku usaha, dan pengambil kebijakan dalam

### PENUTUP

Berdasarkan hasil pengolahan dan analisis data terhadap harga kopi robusta di Kabupaten Muara Enim menggunakan algoritma Random Forest, SVM, dan Decision Tree, dapat disimpulkan bahwa pendekatan data mining memberikan hasil prediksi yang cukup akurat dan bermanfaat bagi pemangku kepentingan dalam rantai pasok kopi. Di antara ketiga algoritma, Random Forest menunjukkan performa yang paling stabil dan akurat dalam meramalkan harga kopi. Penelitian ini juga menunjukkan bahwa penggunaan teknologi seperti Orange Data Mining sangat membantu dalam proses analisis data secara visual dan efisien.

Saran dari penelitian ini adalah agar metode ini dapat diintegrasikan dalam sistem informasi pertanian atau komoditas daerah, sehingga petani dan pelaku industri dapat merancang strategi produksi dan pemasaran secara lebih terarah dan adaptif terhadap perubahan pasar.

### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah berkontribusi dalam pelaksanaan penelitian ini, khususnya kepada para pengumpul kopi di Kabupaten Muara Enim yang telah bersedia memberikan data yang dibutuhkan.

Ucapan terima kasih juga disampaikan kepada institusi dan pihak pendukung lainnya atas fasilitas dan dukungan teknis yang diberikan. Penelitian ini tidak akan berjalan lancar tanpa bantuan dan kerja sama dari semua pihak terkait.

## REFERENSI

- [1] Yallah, H., Darwis, M., & Henrowati, R. (2024). Penerapan Algoritma Id3 Melalui Aplikasi Orange Untuk Prediksi Akurasi Akreditasi Sekolah Dasar Di Depok. *Jurnal Manajemen Informatika dan Sistem Informasi*, 7(2), 131-141.
- [2] Hikmah, I. R., & Yasa, R. N. (2021). Perbandingan Hasil Prediksi Diagnosis pada Indian Liver Patient Dataset (ILPD) dengan Teknik Supervised Learning Menggunakan Software Orange. *Jurnal Telematika*, 16(2), 69-76.
- [3] Mardiani, E., Rahmansyah, N., Ningsih, S., Lantana, D. A., Wirawan, A. S. P., Wijaya, S. A., & Putri, D. N. (2023). Komparasi Metode Knn, Naive Bayes, Decision Tree, Ensemble, Linear Regression Terhadap Analisis Performa Pelajar Sma. *Innovative: Journal Of Social Science Research*, 3(2), 13880-13892.
- [4] Putri, A. J., Syafira, A. S., Purbaya, M. E., & Purnomo, D. (2022). Analisis sentimen e-commerce lazada pada jejaring sosial twitter menggunakan algoritma support vector machine. *Jurnal TRINISTIK: Jurnal Teknik Industri, Bisnis Digital, dan Teknik Logistik*, 1(1), 16-21.
- [5] Nurhafida, S. I., & Sembiring, F. (2021, September). Analisis Text Clustering Masyarakat Di Twiter Mengenai Mcdonald'Sxbts Menggunakan Orange Data Mining. In *Prosiding Seminar Nasional Sistem Informasi Dan Manajemen Informatika Universitas Nusa Putra* (Vol. 1, pp. 28-35).
- [6] Indriyanti, I., Ichsan, N., Fatah, H., Wahyuni, T., & Ermawati, E. (2022). Implementasi Orange Data Mining Untuk Prediksi Harga Bitcoin. *Jurnal Responsif: Riset Sains dan Informatika*, 4(2), 118-125.
- [7] Punkastyo, D. A., Septian, F., & Syaripudin, A. (2024). Implementasi data mining menggunakan algoritma Naïve Bayes untuk prediksi kelulusan siswa. *Journal of System and Computer Engineering*, 5(1), 24-35.
- [8] Zulfikri, M. R., Wiriasto, G. W., & Al Sasongko, S. M. PERANCANGAN SISTEM KLASIFIKASI JENIS CACAT PADA BIJI KOPI ARABIKA MENGGUNAKAN ARSITEKTUR MOBILENETV2 BERBASIS ANDROID.
- [9] Lubis, R. A., Hasibuan, K. S., Sari, N., Lubis, S. H., & Ramadani, P. (2022). Pengaruh Nilai Tukar Rupiah dan Harga Kopi Internasional Terhadap Nilai Ekspor Kopi
- [10] Indonesia Tahun 2004-2021. *Jurnal Penelitian Ekonomi Manajemen dan Bisnis*, 1(4), 226-240.
- [11] Kurniawan, D., Apriani, S., & Pujianto. (2024). *Prediksi angka kemiskinan di Kabupaten Ogan Komering Ulu menggunakan metode data mining berbasis SVM, Neural Network & SGD*. *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, 13(1)