

Perbandingan Metode *Random Forest*, *Linier Regression*, *SVM* Untuk Memprediksi Harga Beras Premium

Doni Saputra^{1*}, Dio Raihan Trinadi², Desti Agustina³, Pujiyanto⁴

^{1*}Fakultas Teknik dan Komputer
Prodi Informatika
(Universitas Baturaja)

^{1*}dodon0853@gmail.com

²Fakultas Teknik dan Komputer Prodi
Informatika
(Universitas Baturaja)

²dioraihan742@gmail.com

³Fakultas Teknik dan Komputer Prodi
Informatika
(Universitas Baturaja)

³agustinadesti889@gmail.com

⁴Fakultas Teknik dan Komputer Prodi
Informatika
(Universitas Baturaja)

⁴pujiyanto.mail@gmail.com

Abstrak— Di Indonesia beras merupakan bahan pokok utama yang sangat mempengaruhi per ekonomian dan kesejahteraan masyarakat. Oleh karena itu, prediksi harga beras menjadi sangat penting dalam membantu pemerintah mengambil keputusan kedepan untuk menjaga stabilitas ekonomi negara dan menjaga kesejahteraan masyarakat. Penelitian ini bertujuan untuk memprediksi harga beras premium menggunakan algoritma *Random Forest*, *Linier Regression* dan *SVM* melakukan perbandingan metode manakah yang lebih akurat dalam memprediksi harga beras di hari yang akan datang. Data yang di gunakan dalam penelitian ini bersumber dari *compas.com* dari tahun 2024 – 2025 untuk memprediksi harga beras di tahun 2026. Model diuji menggunakan pendekatan supervised learning dengan validasi silang untuk memastikan keandalan hasil. Berdasarkan hasil pengujian yang di lakukan dengan algoritma *Random Forest*, *Linier Regression*, dan *SVM* memberikan hasil evaluasi bahwa metode *Random Forest* lah yang paling akurat hasilnya, karena berdasarkan evaluasi MAE (*mean absolute error*) nilai rata-rata terkecilah yang paling akurat untuk prediksi harga beras, maka dalam hal prediksi harga beras metode *Random Forest*lah yang sangat cocok untuk di gunakan para peneliti.

Kata kunci— Harga Beras, *Random Forest*, *Linier Regression*, *SVM*, Prediksi.

Abstract— In Indonesia, rice is the main staple food that greatly affects the economy and people's welfare. Therefore, rice price prediction is very important in helping the government make future decisions to maintain the country's economic stability and maintain people's welfare. This study aims to predict premium rice prices using the *Random Forest*, *Linear Regression* and *SVM* algorithms to measure more accurate methods in predicting rice prices in the future. The data used in this study came from *kompas.com* from 2024 - 2025 to predict rice prices in 2026. The model was tested using a supervised learning approach with cross-validation to ensure the results. Based on the results of testing carried out with the *Random Forest*, *Linear Regression*, and *SVM* algorithms, the evaluation results showed that the *Random Forest* method was the most accurate, because based on the MAE (*mean absolute error*) evaluation, the smallest average value was the most accurate for predicting rice prices, so in terms of predicting rice prices, the *Random Forest* method is the most suitable for use by researchers.

Keywords— Rice price, *Random Forest*, *Linier Regression*, *SVM*, Prediction

I. PENDAHULUAN

Di Indonesia beras merupakan bahan pokok utama yang sangat mempengaruhi per ekonomian dan kesejahteraan masyarakat. Oleh karena itu, prediksi harga beras menjadi sangat penting dalam membantu pemerintah mengambil keputusan kedepan untuk menjaga stabilitas ekonomi negara dan menjaga kesejahteraan masyarakat.

Nasi merupakan kebutuhan pokok bagi orang Asia Tenggara pada umumnya dan masyarakat

Indonesia pada khususnya. Istilah makan seringkali dikaitkan ketika kita mengonsumsi nasi sehingga nasi jarang sekali absen dalam setiap periode makan. Mengingat keberadaan beras dalam salah satu dari sembako serta merupakan makanan utama bagi masyarakat Indonesia membuat perhatian lebih bagi pemerintah dan para pemerhati khususnya untuk mengontrol harga dari beras itu sendiri di negara kita. [1]

Untuk menghindari permasalahan ketidakstabilan harga bahan pangan dimasa yang akan datang, dibutuhkan sebuah sistem yang dapat melakukan prediksi harga bahan pokok beras. Harapannya dengan adanya sistem tersebut dapat menjamin masyarakat secara luas terutama masyarakat menengah ke bawah bisa menjangkau harga beras dan juga dapat membuat hidup masyarakat lebih sejahtera, serta dapat menjadi tolak ukur bagi para pengusaha penggilingan padi, buruh tani dan juga para pedagang dalam membuat strategi pemasaran beras untuk meminimalkan resiko kerugian. [2]

Terdapat beberapa metode dalam melakukan pengklasifikasian salah satunya adalah metode Random Forest. Dimana *Random Forest* dapat meningkatkan akurasi karena adanya pemilihan secara acak dalam membangkitkan simpul anak untuk setiap node (simpul di atasnya) dan diakumulasikan hasil klasifikasi dari setiap pohon (*tree*), kemudian dipilih hasil klasifikasi yang paling banyak muncul. [3]

Random forest adalah algoritma machine learning yang mengikuti konsep supervised dalam pembentukan kelas pengklasifikasi. Algoritma ini menggabungkan hasil prediksi dari beberapa pohon keputusan[4]

Random Forest terbentuk dari sekumpulan pohon keputusan. Secara keseluruhan, random Forest menggabungkan hasil dari seluruh pohon Keputusan yang terbentuk selama proses pelatihan, menghasilkan keluaran kelas untuk tugas klasifikasi dan prediksi nilai rata-rata untuk model regresi.[5]

Linear regression dipilih sebagai metode analisis karena mampu menggambarkan hubungan linier antara variabel dependen. Serta Regresi linear memberikan landasan untuk memahami tren langsung secara linier.[6]

Linear regression merupakan salah satu algoritma yang memodelkan suatu persamaan untuk menghitung estimasi. Pada metode ini bertujuan untuk mencari pola pada nilai *Copyright* numerik, sehingga data yang dibutuhkan berupa data numerik agar dapat diolah dengan model algoritma ini. Linear regression memprediksi nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui. Secara

matematis memodelkan *variable* yang tidak diketahui atau tergantung dan *variable* yang dikenal atau independent sebagai persamaan linier.[7]

Support Vector Machine (SVM) adalah pengklasifikasi diskriminatif yang menghasilkan *hyperplane* pemisah. Toleransi kesalahan disertakan untuk membuat *hyperplane* pemisah menjadi kuat jika terjadi data kelas yang tidak dapat dipisahkan. Dalam Algoritma SVM ada trik kernel dimana ada SVM linear dan SVM non linear. SVM *hyperplane* linear bekerja hanya pada data yang dapat dipisahkan dengan cara linear.SVM Non Linear yaitu data yang berdistribusi pada kelas yang tidak linear sering digunakan pendekatan kernel pada fitur awal set.[8]

Untuk menentukan model algoritma mana yang akurat penulis menggunakan evaluasi *Mean Absolute Error* (MAE) adalah salah satu metode yang digunakan untuk mengukur tingkat keakuratan model peramalan. Nilai MAE menunjukkan rata – rata kesalahan (error) absolut antara hasil peramalan/prediksi dengan nilai riil [9].

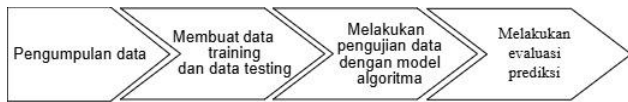
Orange merupakan perangkat lunak open source untuk *machine learning* dan data mining yang ditulis dengan bahasa Python. Orange dikembangkan oleh Laboratorium Bioinformatika, Fakultas Ilmu Komputer dan Informasi, Universitas Ljubljana.[10]

Orange merupakan salah satu aplikasi Data Mining yang dapat digunakan untuk melakukan analisis dan visualisasi data. Aplikasi bersifat *Open Source*, dan fitur-fitur yang disediakan cukup lengkap dan sangat memudahkan penggunaannya. Dimana Widget yang disediakan cukup mudah digunakan dan hasil dari analisis juga sudah tertera dengan jelas. Sehingga pengguna dapat lebih mudah memahami hasil dari data yang dianalisis [11].

II. METODOLOGI PENELITIAN

Penelitian ini bertujuan untuk menentukan perbandingan metode manakah yang lebih akurat untuk memprediksi harga beras antara metode *Random Forest*, *Regretion linier*, dan SVM. aplikasi yang di gunakan untuk simulasi adalah orange merupakan aplikasi *open source* yang telah

terbukti membantu peneliti menganalisa datanya dengan mudah[12].



Gambar 1. Tahap penelitian

Sesuai dengan tahapan penelitian, langkah pertama yang di lakukan adalah pengumpulan data, merupakan point yang sangat penting karena data adalah modal awal untuk melakukan penelitian ini, data yang kami kumpulkan adalah data harga beras perhari, dari bulan April 2024 sampai bulan April 2025. Kedua membuat data training dan data testing, data training di sini adalah data yang di butuhkan algoritma sebagai bahan belajar untuk membuat prediksi, sedangkan data testing adalah data yang akan di prediksi. Ketiga adalah proses pengujian data dengan model algoritma *Random Forest*, *Linier Regression*, dan *SVM*. Keempat melakukan evaluasi kinerja metode *random forest*, *linier regression*, dan *SVM* menganalisa hasil perbandingan dari metode tersebut.

2.1 Atribut data dan select columns

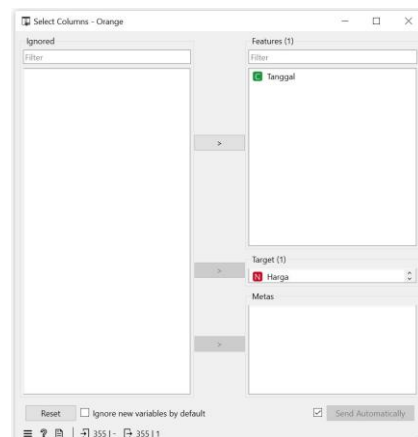
Kumpulan data harga beras premium berupa data sekunder yang di peroleh dari Internet tentang harga beras yang bersumber dari *website* *compas.com* artribut awal dari data ini adalah bulan/tahun, jika data aslinya untuk prediksi di orange kami menggunakan data perhari, kemudian kami rangkum dan yang kami cantumkan pada *table* di bawah ini adalah *table* data harga perbulan. Di atribut ke dua ada harga sebagai bahan acuan algoritma untuk mempelajari pola dan dapat menentukan hasil prediksi.

Tabel 1 data beras perbulan

No	Bulan/Tahun	Harga
1	01 April 2024	Rp. 16.230
2	01 Mei 2024	Rp. 16.230
3	01 Juni 2024	Rp. 15.810
4	01 Juli 2024	Rp. 15.830
5	01 Agustus 2024	Rp. 15.560
6	01 September 2024	Rp. 15.910
7	01 Oktober 2024	Rp. 15.610

8	01 November 2024	Rp. 15.630
9	01 Desember 2024	Rp. 15.690
10	01 Januari 2025	Rp. 15.460
11	01 Februari 2025	Rp. 15.407
12	01 Maret 2025	Rp. 15.900
13	01 April 2025	Rp. 15.556

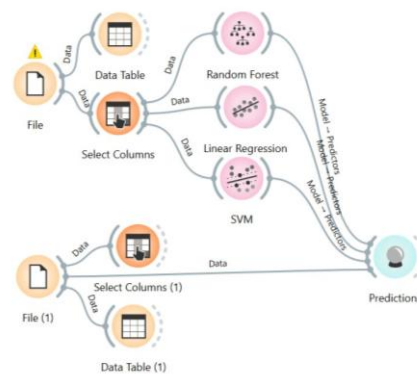
Pada proses di fitur *select columns* harga merupakan *table numeric* dan tanggal adalah data dengan *type categorical*, karena harga adalah target prediksi maka posisi *table* harus di letakan pada kolom target dan tanggal tetap berada di features karena tanggal merupakan variable input untuk memprediksi sesuatu.



Gambar 2. Filter data dengan fitur select columns

2.2 Proses Data Mining

Dalam menganalisa performa beberapa model klasifikasi pada orange tool, maka dilakukan perbandingan beberapa metode data mining untuk memilih metode yang terbaik dengan akurasi yang tinggi, dalam mengklasifikasi dataset harga beras premium.



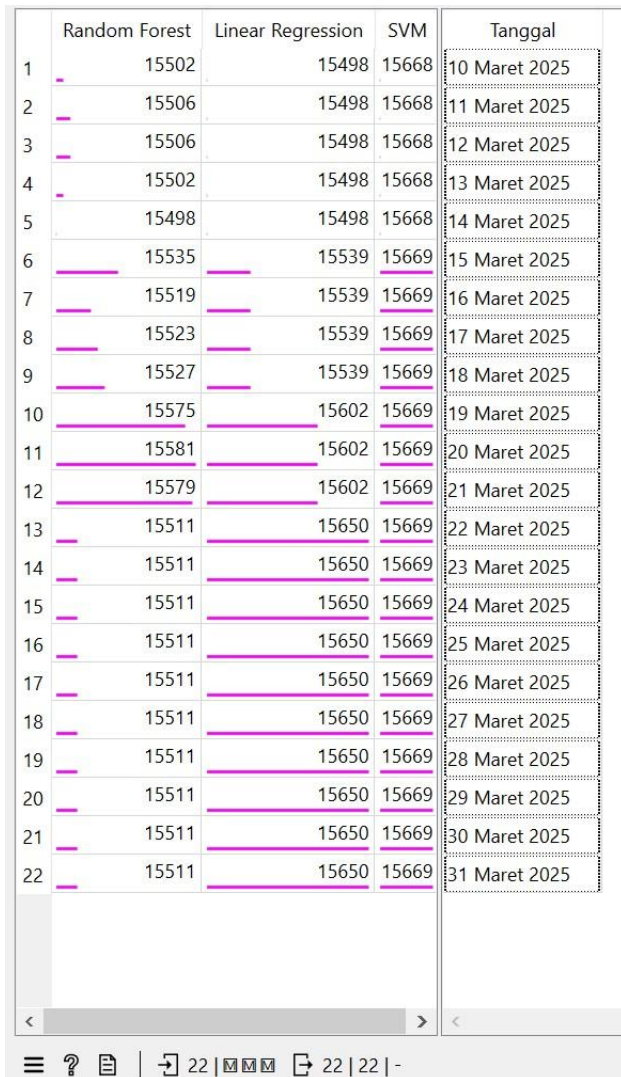
Gambar 3. Desain klasifikasi dataset untuk prediksi

III. HASIL DAN PEMBAHASAN

3.1 Hasil simulasi prediksi

Gambar 4 di bawah ini adalah tabel prediksi yang menunjukkan hasil percobaan prediksi harga beras di pertengahan bulan maret sampai akhir bulan maret guna mencari keakuratan model algoritma mana yang paling akurat untuk di gunakan menghitung prediksi di tahun yang akan datang.

Untuk menentukan model mana yang paling akurat, gunakan MAE (*mean absolute error*) Mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual, semakin kecil mae, semakin akurat model prediksi.



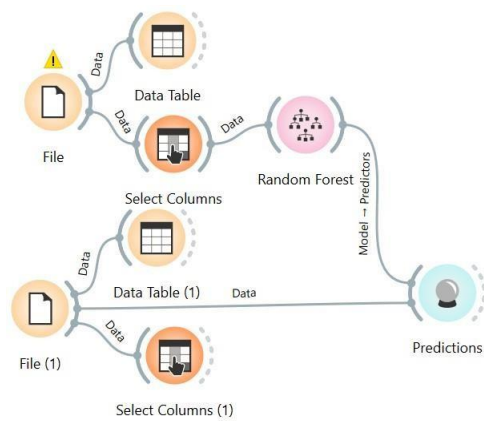
Gambar 4. Hasil simulasi prediksi

Random Forest, Linear Regression, dan SVM: Menampilkan angka prediksi harga dan juga *visual bar* (garis warna magenta) yang menunjukkan nilai relatif. Ini memudahkan kita melihat perbedaan antar model secara visual. Pada tanggal 10–14 Maret 2025, semua algoritma menunjukkan harga konstan, Mulai 15 Maret 2025, harga dari semua algoritma mulai berubah naik.

Tabel 2 perhitungan MAE Random Forest

Tgl	Akt ual	Pre_ RF	Pre_ LR	Pre_ S VM	Err_ RF	Err_ LR	Err_ S VM
10 mar ch 202 5	154 98	1550 6	1549 8	15668	8	0	170
11 mar ch 202 5	154 98	1550 2	1549 8	15668	4	0	170
12 mar ch 202 5	154 98	1550 2	1549 8	15668	4	0	170
13 mar ch 202 5	154 98	1550 6	1549 8	15668	8	0	170
14 mar ch 202 5	154 98	1551 1	1549 8	15668	13	0	170
15 mar ch 202 5	155 39	1552 7	1553 9	15669	12	0	130
16 mar ch 202 5	155 39	1552 3	1553 9	15669	16	0	130
17 mar ch 202 5	155 39	1551 9	1553 9	15669	20	0	130
18 mar ch 202 5	155 39	1553 1	1553 9	15669	8	0	130

5								mar ch 202 5	56	5	0						
19 mar ch 202 5	156 02	1555 8	1560 2	15669	44	0	67	31 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113		
20 mar ch 202 5	156 02	1555 6	1560 2	15669	46	0	67										
												MAE	28,7 2	42,7 2	122,7 7		
21 mar ch 202 5	156 02	1556 3	1560 2	15669	39	0	67	<p>Pada table 2 di atas terdapat atribut tanggal hari yang akan di prediksi, kemudian kolom kedua adalah harga aktual yang di kumpulkan oleh penulis dari beberapa sumber, dan yang ketiga adalah kolom prediksi menggunakan model algoritma <i>random forest</i>, ke empat kolom prediksi harga dari model <i>Linier Regression</i>, kelima ada kolom yang berisi nilai prediksi dari model SVM, kolom delapan menunjukkan <i>error</i> absolut yang di hasilkan dari model <i>random forest</i>, kolom ke tujuh dan kedelapan juga sama menampilkan error absolut dari model linier regression dan SVM, setelah di dapatkan nilai error absolut, di lanjutkan dengan menghitung nilai rata-rata MAE dengan rumus excel “=AVERAGE(F2:F23)” maka menghasilkan nilai MAE <i>Random Forest</i> 28,72. <i>Linier Regression</i> 42,72. SVM 122,77. Berdasarkan hasil MAE (<i>mean absolute error</i>) semakin kecil nilai mae semakin akurat hasil sebuah prediksi, dan di dapati model algoritma <i>Random Forest</i> yang paling cocok di gunakan untuk prediksi harga beras.</p> <p>3.2 Prediksi harga beras dengan metode <i>random forest</i>.</p> <p>Berdasarkan simulasi di atas, metode <i>Random Forest</i> lah yang hasil evaluasinya mendapat nilai terkecil dengan nilai 28,72 yang artinya memiliki keakuratan dalam memprediksi harga beras di tahun 2026, dengan menggunakan data harga beras dari tanggal 1 april 2024 sampai 30 april 2025, data testing atau data yang akan di prediksi mulai dari tanggal 1 januari 2026 sampai 10 januari 2026.</p>									
22 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113										
23 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113										
24 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113										
25 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113										
26 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113										
27 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113										
28 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113										
29 mar ch 202 5	155 56	1551 5	1565 0	15669	41	94	113										
30	155	1551	1565	15669	41	94	113										



Gambar 5. Workflow prediksi harga beras tahun 2026

Pada gambar 5 di atas peneliti hanya menggunakan model algoritma *Random Forest* untuk memprediksi harga beras di tahun yang akan datang. Dengan adanya hasil prediksi ini di harapkan masyarakat dapat mengambil keputusan yang lebih tepat dan bijak berdasarkan informasi yang tersedia.

	Random Forest	tanggal
1	15556	1 januari 2026
2	15556	2 januari 2026
3	15556	3 januari 2026
4	15556	4 januari 2026
5	15556	5 januari 2026
6	15556	6 januari 2026
7	15556	7 januari 2026
8	15556	8 januari 2026
9	15556	9 januari 2026
10	15556	10 januari 2026

Gambar 6. Hasil prediksi *Random Forest*

Gambar table di atas adalah hasil prediksi harga beras di awal bulan januari dengan metode *Random Forest*.

IV. KESIMPULAN

Berdasarkan hasil penelitian menentukan perbandingan metode yang paling akurat antara metode *Random Forest*(28,72), *Linier Regression*(42,72), *SVM*(122,77) kemudian di evaluasi dengan perhitungan MAE (mean absolute error) dalam pengukuran ini metode yang memiliki nilai mae terendahlah yang paling akurat untuk melakukan prediksi dan di dapati hasil metode *Random Forest* mendapatkan *Absolute_error* paling rendah, yang artinya metode ini lah yang paling cocok di gunakan untuk prediksi harga beras di tahun 2026. Algoritma ini mampu menghasilkan prediksi yang stabil dan akurat dengan menggabungkan banyak *decision tree* dan melakukan voting (untuk klasifikasi) atau rata-rata (untuk regresi). Pengembangan selanjutnya yang dapat dilakukan yaitu mengintegrasikan algoritma *Random Forest* dengan model lain seperti *Gradient Boosting* atau *kNN* untuk membandingkan performa. Menerapkan metode ini dalam konteks data *real-time* untuk prediksi langsung (misalnya dalam *dashboard monitoring*).

UCAPAN TERIMA KASIH

Terimakasih di sampaikan kepada dosen pengampu yang telah memberikan pelatihan dalam penelitian ini dan ucapan terimakasih juga untuk teman-teman seperjuangan yang sudah berkontribusi dalam penelitian ini.

REFERENSI

- [1] R. E. Putra dan A. S. Sinaga, "Perkiraan Harga Beras Premium DKI Jakarta Menggunakan Regresi Linier."
- [2] B. Ghulam, A. Shidiq, M. T. Furqon, dan L. Muflikhah, "Prediksi Harga Beras menggunakan Metode Least Square," 2022. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [3] V. Wanika Siburian, J. Sistem Komputer Universitas Sriwijaya Palembang, dan I. Elvina Mulyana, *Prediksi Harga Ponsel Menggunakan Metode Random Forest*. 2018.
- [4] J. Hutahaean dan D. Yusup, "Perbandingan Metode Linear Regression, Random Forest & K-Nearest Neighbor Untuk Prediksi Produksi Hasil Panen Padi Di Provinsi Jawa Barat," 2024.
- [5] A. Kholiq Fajar, M. Zaenal Mutaqin, M. Malik Mutoffar, dan D. Setiyadi, "Sekolah Tinggi Teknologi Bandung, 5 Program Studi Sistem Informasi," *STMIK Sinar Nusantara Jln. Sersan Aswan*, vol. 7, no. 1, 2024, doi: 10.36595/misi.v5i2.
- [6] R. N. Silalahi dan M. Muljono, "Perbandingan Kinerja Metode Linear Regression, LSTM dan GRU Untuk Prediksi Harga Penutupan Saham Coca-Cola," *Komputika : Jurnal Sistem Komputer*, vol. 13, no. 2, hlm. 201–211, Okt 2024, doi: 10.34010/komputika.v13i2.12265.

- [7] E. Mardiani *dkk.*, “Copyright @ Analisis Prediksi Pendapatan Penduduk dengan Metode K- Nearest Neighbor, Decision Tree, Naive Bayes, Ensemble Methods, dan Linear Regression”.
- [8] S. K. Parinduri, R. Rosnelly, A. Purnama, A. Sihotang, dan M. C. Adelina, “Perbandingan Algoritma Support Vector Machine Dan K-Nearest Neighbors Pada Kematangan Buah Sawit,” *JURNAL DEVICE*, vol. 13, hlm. 147–155, 2023, [Daring].
Tersedia pada:
<https://indonesiabaik.id/infografis/indonesia->
- [9] A. A. Suryanto, A. Muqtadir, dan S. Artikel, “Penerapan Metode Mean Absolute Error (Mea) Dalam Algoritma Regresi Linear Untuk Prediksi Produksi Padi Info Artikel : Abstrak,” no. 1, hlm. 11, 2019.
- [10] I. Rosianal *dkk.*, “Perbandingan Hasil Prediksi Diagnosis pada Indian Liver Patient Dataset (ILPD) dengan Teknik Supervised Learning Menggunakan Software Orange,” *Jurnal Telematika*, vol. 16, no. 2.
- [11] A. C. Handoko dan H. Hendry, “Perbandingan Metode Supervised Learning Untuk Prediksi Diabetes Gestasional Dengan Software Orange,” *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 4, hlm. 1238–1247, Nov 2023, doi: 10.29100/jipi.v8i4.4166.
S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [12] S. Alim, “Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes Orange Data Mining Implementation For Student Graduation Classification Using K-Nearest Neighbor, Decision Tree And Naive Bayes Models.”