



Komparasi Metode *Support Vector Machine* dan *Random Forest* untuk Prediksi Penjualan Solar Industri (HSD) pada PT Heva Petroleum Energi Palembang

Putri Octaria^{1*}, Shinta Puspasari², Evi Purnamasari³

¹⁻³ Teknik Informatika, Universitas Indo Global Mandiri Palembang, Indonesia

Email: 2021110127@students.uigm.ac.id^{1*}, shinta@uigm.ac.id², evips@uigm.ac.id³

Alamat: Jl. Jend. Sudirman Km.4 No. 62, 20 Ilir D. IV, Kec. Ilir Tim. I, Kota Palembang, Sumatera Selatan 30129

*Korespondensi penulis

Abstract. *The fluctuating nature of Industrial Solar or High Speed Diesel (HSD) sales poses a significant challenge for companies, particularly in developing appropriate distribution strategies and stock planning. This situation demands the application of data-driven analytical methods to support more effective decision-making. This study aims to predict Industrial Solar sales at PT Heva Petroleum Energi Palembang using two Machine Learning methods, namely Support Vector Machine (SVM) and Random Forest. The data used are monthly sales records for the period 2022–2024. The research process includes data collection, pre-processing with normalization and feature selection, model building, testing by dividing the data into training and test sets, and performance evaluation using the Mean Absolute Percentage Error (MAPE) metric. The results show that the Random Forest model produces a MAPE value of 12.48%, while the Support Vector Machine model obtains a MAPE value of 12.97%. This comparison shows that Random Forest is superior in predicting sales compared to SVM. Thus, it can be concluded that Random Forest is a more appropriate choice for application in modeling Industrial Solar sales. The implications of these findings are expected to provide a real contribution to companies in developing distribution policies and stock management that are more accurate, efficient, and sustainable, so as to be able to support the stability of company operations in the future.*

Keywords: *Industrial Diesel Fuel; Machine Learning; Prediction; Random Forest; Support Vector Machine.*

Abstrak. Penjualan Solar Industri atau High Speed Diesel (HSD) yang bersifat fluktuatif menjadi tantangan signifikan bagi perusahaan, khususnya dalam menyusun strategi distribusi dan perencanaan stok yang tepat. Kondisi ini menuntut penerapan metode analisis berbasis data untuk mendukung pengambilan keputusan yang lebih efektif. Penelitian ini bertujuan untuk memprediksi penjualan Solar Industri di PT Heva Petroleum Energi Palembang dengan memanfaatkan dua metode *Machine Learning*, yaitu Support Vector Machine (SVM) dan Random Forest. Data yang digunakan berupa catatan penjualan bulanan selama periode 2022–2024. Proses penelitian meliputi tahap pengumpulan data, pra-pemrosesan dengan normalisasi serta seleksi fitur, pembangunan model, pengujian dengan membagi data ke dalam set latih dan uji, serta evaluasi performa menggunakan metrik Mean Absolute Percentage Error (MAPE). Hasil penelitian menunjukkan bahwa model Random Forest menghasilkan nilai MAPE sebesar 12,48%, sedangkan model Support Vector Machine memperoleh nilai MAPE sebesar 12,97%. Perbandingan ini memperlihatkan bahwa Random Forest lebih unggul dalam memprediksi penjualan dibandingkan SVM. Dengan demikian, dapat disimpulkan bahwa Random Forest merupakan pilihan yang lebih tepat untuk diaplikasikan dalam memodelkan penjualan Solar Industri. Implikasi dari temuan ini diharapkan dapat memberikan kontribusi nyata bagi perusahaan dalam menyusun kebijakan distribusi dan pengelolaan stok yang lebih akurat, efisien, serta berkelanjutan, sehingga mampu mendukung stabilitas operasional perusahaan di masa mendatang.

Kata kunci: Machine Learning; Prediksi; Random Forest; Solar Industri; Support Vector Machine.

1. LATAR BELAKANG

High Speed Diesel (HSD) adalah bahan bakar hasil pengolahan minyak bumi melalui proses penyulingan, yang memiliki warna kuning kecokelatan jernih dan umum dikenal sebagai solar. Bahan bakar ini umumnya digunakan pada berbagai jenis mesin diesel berkecepatan tinggi (lebih dari 1000 rpm), serta dapat dimanfaatkan pada dapur berskala kecil untuk menghasilkan pembakaran yang lebih bersih. (Direktur Jendral Minyak Gas dan Bumi KESDM, 2023)

Dalam era industri *modern* yang semakin bergantung pada energi, ketersediaan dan distribusi bahan bakar minyak (BBM) menjadi aspek vital dalam mendukung kelangsungan operasional berbagai sektor. Di tengah meningkatnya permintaan dan dinamika pasar energi yang *fluktuatif*, Salah satu tantangan terbesar yang dihadapi oleh perusahaan penyedia Solar Industri *High Speed Diesel (HSD)* adalah ketidakpastian dalam pola permintaan, yang dapat berubah drastis akibat faktor musiman, kondisi ekonomi, hingga kebijakan pemerintah (Lutfi, 2021)

PT Heva Petroleum Energi Palembang merupakan perusahaan yang bergerak di bidang transportasi dan distribusi Solar Industri *High Speed Diesel (HSD)* di wilayah Sumatera Selatan, tepatnya berlokasi di Komplek Bua Sakti, Jl. Sukabangun II Soak Simpung Blok C5. Perusahaan ini memiliki peran penting dalam mendistribusikan bahan bakar ke berbagai sektor industri dan komersial. Namun, dalam operasionalnya, PT Heva Petroleum menghadapi tantangan berupa fluktuasi permintaan solar industri yang dipengaruhi oleh berbagai faktor, seperti tren pasar, musim, harga, dan kondisi ekonomi.

Penelitian ini menggunakan metode *Support Vector Machine (SVM)* dan *Random Forest* untuk memprediksi penjualan solar industri. *SVM* dikenal efektif dalam menangani data berdimensi tinggi dan membangun *hyperplane optimal* untuk *klasifikasi* maupun *regresi*. Namun, metode ini sensitif terhadap data *outlier* dan memerlukan parameter yang tepat agar performanya maksimal. Sebaliknya, *Random Forest* adalah *metode ensemble* yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi. Metode ini dikenal *robust* terhadap *noise*, memiliki akurasi tinggi, dan cocok diterapkan pada data yang kompleks, meskipun memiliki kelemahan dalam konsumsi sumber daya komputasi.

Beberapa studi terdahulu telah menunjukkan efektivitas metode pembelajaran mesin dalam memprediksi penjualan BBM. Misalnya, penelitian oleh Arfan dan Paraga (2024) menyatakan bahwa belum ada penerapan metode yang lebih kompleks seperti *SVM* dan *Random Forest*, maka dari itu (Arfan & Paraga, 2024) memlihi membandingkan algoritma *K-Means*, *Naïve Bayes*, dan *Decision Tree* dalam prediksi penjualan BBM di SPBU Wadio

Nabire. Pada penulisannya (Arfan & Paraga, 2024) dengan hasil Decision Tree menunjukkan akurasi tertinggi (95%).

Selain itu, penelitian oleh (Darmawan et al., 2023) yang menggunakan metode *SVM* dan *Random Forest* untuk memprediksi kelulusan siswa juga menunjukkan bahwa *Random Forest* menghasilkan akurasi lebih tinggi (99,49%) dibandingkan *SVM* (98,98%). Hasil ini membuktikan bahwa kedua metode tersebut mampu memberikan hasil prediksi yang akurat meskipun diterapkan pada data yang kompleks dan berskala besar.

Berdasarkan hasil dari penelitian sebelumnya di atas, penelitian ini bertujuan untuk menerapkan metode *SVM* dan *Random Forest* dalam memprediksi penjualan solar industri di PT Heva Petroleum Energi Palembang. Diharapkan, model prediksi ini dapat menjadi alat bantu bagi perusahaan dalam meningkatkan efisiensi distribusi, perencanaan produksi, serta pengelolaan stok bahan bakar secara lebih akurat dan tepat sasaran.

2. KAJIAN TEORITIS

Machine Learning

Machine Learning merupakan salah satu cabang dari kecerdasan buatan yang bertujuan untuk mengembangkan algoritma yang memungkinkan komputer untuk belajar dari data tanpa harus diprogram secara *eksplisit*. Dengan kata lain, *Machine Learning* adalah sebuah sistem yang dirancang untuk bisa belajar dan melakukan tugas-tugas tanpa dibimbing langsung oleh pengguna. (Wijoyo A et al., 2024) Menurut (Gustriansyah et al., 2024), metode *Support Vector Machine (SVM)* dan *Random Forest* merupakan algoritma pembelajaran mesin kompleks nonlinear yang terbukti efektif dalam tugas klasifikasi maupun prediksi dengan tingkat akurasi yang tinggi. (Suhandi et al., 2024), tahap penting dalam penelitian *machine learning* meliputi pra-pemrosesan data, pembagian dataset, serta evaluasi model menggunakan *confusion matrix* dengan metrik akurasi, presisi, dan recall.

Prediksi

Prediksi merupakan suatu kegiatan yang dilakukan untuk memperkirakan data, baik pada kondisi saat ini maupun untuk periode yang akan datang (Sari et al., 2023). Proses ini sangat erat kaitannya dengan perhitungan matematis dari kumpulan data dalam jumlah besar. Estimasi dapat dilakukan dengan berbagai metode, bahkan seringkali dikombinasikan agar menghasilkan prediksi yang lebih akurat. Manfaat utama dari prediksi adalah membantu pemilik usaha dalam mengambil keputusan terkait jumlah barang yang perlu disediakan perusahaan. (Duran et al., 2024) Selain itu, prediksi juga berperan penting dalam

perencanaan persediaan stok, karena hasil yang dihasilkan dapat memberikan gambaran yang optimal, sehingga risiko kesalahan akibat perencanaan yang kurang tepat dapat diminimalkan.(Miftahul Jannah et al., 2023)

Metode Support Vector Machine

Metode *Support Vector Machine (SVM)* dipilih dalam penelitian ini karena merupakan metode pembelajaran mesin berbasis kernel yang efektif untuk tugas klasifikasi maupun regresi, sehingga sesuai digunakan dalam memprediksi penjualan solar industri. *SVM* bekerja dengan memaksimalkan margin antar kelas untuk membentuk *hyperplane* optimal yang mampu mengidentifikasi pola penjualan, bahkan ketika data memiliki dimensi tinggi (Gustriansyah et al., 2024). Keunggulan utama *SVM* adalah kemampuannya dalam menghindari *overfitting*, terutama pada dataset berdimensi tinggi namun jumlah datanya terbatas. Meskipun demikian, *SVM* juga memiliki kelemahan, seperti sensitivitas terhadap pemilihan parameter (misalnya nilai C dan kernel), waktu pelatihan yang relatif lama pada dataset besar, serta rentan terhadap data *noise* dan *outlier* (Yunita & Kamayani, 2023) Oleh karena itu, diperlukan penanganan data yang baik serta pemilihan parameter yang tepat agar kinerja model dapat optimal dalam memprediksi penjualan solar industri di PT Heva Petroleum Energi.

Metode Random Forest

Metode *Random Forest* merupakan salah satu teknik *ensemble learning* yang menggunakan pohon keputusan sebagai base *classifier* yang dibangun secara acak kemudian dikombinasikan untuk menghasilkan prediksi yang lebih akurat (Fitri, 2023). Metode ini bekerja dengan membangun sejumlah pohon keputusan melalui proses sampling data dan pemilihan atribut secara acak, di mana setiap pohon terdiri dari node akar (root node), node internal sebagai simpul percabangan, serta node daun (leaf node) sebagai penentu hasil prediksi (Cahya et al., 2024). Random Forest memiliki berbagai kelebihan, di antaranya akurasi yang tinggi, robust terhadap *outliers* dan *noise*, lebih cepat dibandingkan beberapa metode ensemble lainnya, serta sederhana dan mudah diparalelkan sehingga *Random Forest* dipilih karena memiliki keunggulan dalam menangani variabel yang saling berinteraksi serta lebih tahan terhadap permasalahan *overfitting* dibandingkan dengan metode lainnya(Purnamasari et al., 2025).

Metrik Evaluasi

Metrik evaluasi digunakan untuk menilai seberapa baik kinerja prediksi suatu model. Dalam model deret waktu, evaluasi ini dilakukan dengan menghitung tingkat kesalahan yang dihasilkan oleh model (Duran et al., 2024) Penelitian ini menggunakan *MAPE* Berikut adalah penjelasan lebih lanjut mengenai *MAPE*

MAPE (*Mean Absolute Percentage Error*) merupakan salah satu metode statistik yang digunakan untuk menilai tingkat akurasi suatu model dalam melakukan prediksi. dilihat dari beberapa sumber, *MAPE* juga disebut sebagai *Mean Absolute Percentage Deviation* (MAPD)(Andika, 2024). Adapun rumus untuk menghitung *MAPE* adalah sebagai berikut:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \times 100\%$$

Keterangan:

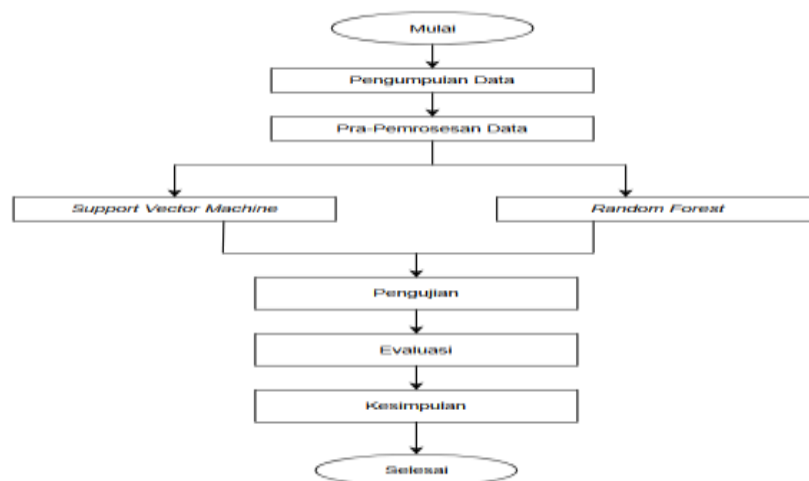
n : Adalah ukuran sample

A_i :Nilai data Aktual

F_i : Nilai data Prediksi

3. METODE PENELITIAN

Tahap pertama yang dilakukan dalam penelitian ini ialah pengambilan data, diikuti oleh tahap seleksi fitur, normalisasi data dan split data, dan lanjut pada tahap pemodelan algoritma *SVM* dan *Random Forest* serta pengujian akurasi dengan *MAPE*. Tahapan penelitian dapat dilihat pada Gambar 1



Gambar 1. Tahapan Penelitian.

Berdasarkan gambar diatas, alur atau tahap penelitian adalah sebagai berikut.

Tahap 1: Dataset dalam penelitian ini diperoleh dari data penjualan solar industri pada PT Heva Petroleum Energi Palembang dengan rentang waktu 2022 hingga 2024 dengan total 36 dataset.

Tahap 2: Tahap Pra-Pemrosesan Data, pada tahap ini terdapat 3 langkah yang dilakukan yaitu seleksi fitur, normalisasi data dan split data Tahap split data tahap ini split data dibagi menjadi 5 rasio yaitu 90:10, 80:20, 70:30, 60:40, 50:50.

Tahap 3: Penerapan Pemodelan Algoritma metode Support Vector Machine dan Random Forest.

Tahap 4 : pada tahap in model SVM dan Random Forest akan melakukan pengujian, dilanjutkan dengan evaluasi dan terakhir kesimpulan.

Pengumpulan data merupakan tahap penting dalam penelitian ini karena kualitas data yang diperoleh akan sangat mempengaruhi hasil analisis. Data yang terkumpul pada penelitian ini adalah data transaksi penjualan Solar Industri (*HSD*) dari PT Heva Petroleum Energi, yang diambil dalam rentang waktu tahun 2022 hingga 2024. Atribut data mencakup berbagai informasi penting terkait transaksi penjualan, Tanggal_penjualan, Jumlah_konsumen, Harga, Jumlah_liter, dan Total. data penjualan dapat dilihat pada **Tabel 1**.

Tabel 1. Dataset Penjualan Tahun 2022-2024.

No	No.Invoice	Tanggal	Konsume n	NPWP	No.Seri Faktur	.. .	Total Penjuala n
1	021/HSD/HPE/I/2022	12/01/2022		24*.2**.*	010.0**	..	10.000
		2	PT. Roda	1	*	.	
2	022/HSD/HPE/I/2022	17/01/2022		24*.2**.*	010.0**	..	10.000
		2	PT. Roda	1	*	.	
3	023/HSD/HPE/I/2022	20/01/2022	PT.	24*.2**.*	010.0**	..	10.000
		2	Sawitri	1	*	.	
4	024/HSD/HPE/I/2022	25/01/2022	PT.	24*.2**.*	010.0**	..	10.000
		2	Cakra	1	*	.	
5	025/HSD/HPE/I/2022	30/01/2022		24*.2**.*	010.0**	..	10.000
		2	PT. Roda	1	*	.	
		Total					50.000
6		15/02/2022	PT.	24*.2**.*	010.0**	..	10.000
	027/HSD/HPE/I/2022	2	Mandiri	1	*	.	
7		16/02/2022		24*.2**.*	010.0**	..	10.000
	027/HSD/HPE/I/2022	2	PT. Mitra	1	*	.	
8		25/02/2022	PT.	24*.2**.*	010.0**	..	10.000
	028/HSD/HPE/I/2022	2	Dania	1	*	.	
		Total					30.000
45	0455/HSD/HPE/XII/20	22/12/2022	PT. Rafa	24*.2**.*	010.0**	..	5000
5	24	4		1	*	.	

4. HASIL DAN PEMBAHASAN

Pra-Pemrosesan Data

Tahapan dalam pra-pemrosesan ini tetap mengikuti langkah yang serupa. Proses dimulai dengan pembersihan, di mana kolom yang tidak ada relevansinya atau tidak diperlukan untuk analisis lebih lanjut dihilangkan. Selanjutnya, data dinormalisasi agar nilainya berada dalam skala 1 dan 0. Akhirnya, data dipisahkan menjadi set pelatihan dan pengujian untuk mendukung analisis atau penerapan model berikutnya.

Seleksi Fitur

Tahap Pada tahap seleksi fitur Dataset awal penelitian ini semula memiliki 11 atribut, namun setelah melalui tahap seleksi fitur, dipilih 3 atribut yang dinilai paling relevan dan mampu memprediksi dengan baik, yaitu bulan, jumlah konsumen, dan jumlah liter. Proses seleksi ini bertujuan untuk mengurangi kompleksitas model sekaligus meningkatkan akurasi prediksi. Data yang telah disederhanakan ini kemudian digunakan sebagai input pada tahap pemodelan menggunakan algoritma yang telah ditentukan, sehingga proses analisis menjadi lebih fokus dan efisien. Hasil dari seleksi Fitur dapat dilihat pada **Tabel 2**.

Tabel 2. Hasil Seleksi Fitur.

	Bulan	Jumlah Konsumen	Penjualan Liter
2022	1	5	40000
	2	4	30000
	3	3	20000
	4	4	25000
	5	4	40000
	6	5	50000
	7	3	30000
	8	5	35000
	9	6	45000
	10	6	40000
	11	15	105000
	12	25	180000
2023	1	26	130000
	2	20	110000
	3	16	85000
	4	10	60000
	5	9	60000
	6	10	75000
	7	12	85000
	8	11	60000
	9	15	75000
	10	16	80000

2024	11	15	75000
	12	25	125000
	1	26	130000
	2	20	110000
	3	16	85000
	4	11	55000
	5	9	45000
	6	10	50000
	7	17	90000
	8	18	95000
	9	15	80000
	10	19	95000
	11	48	24000
	12	60	30000

Normalisasi Data

Tahap pemodelan prediksi diawali dengan proses normalisasi data untuk menyamakan skala nilai pada setiap fitur, sehingga algoritma *Machine Learning* dapat bekerja secara optimal. Pada penelitian ini digunakan metode Min-Max Scaler, yang mentransformasi nilai setiap atribut ke dalam rentang 0 hingga 1. Proses ini bertujuan meminimalkan perbedaan skala antar variabel dan meningkatkan kinerja model. Implementasi kode normalisasi ditunjukkan pada Gambar 2. sedangkan hasil normalisasi data disajikan pada Tabel 3.

```
# Normalisasi data
scaler_X = MinMaxScaler()
X_scaled = scaler_X.fit_transform(X_SVM)
scaler_y = MinMaxScaler()
y_scaled = scaler_y.fit_transform(y_SVM.values.reshape(-1,
1)).ravel()
```

Gambar 2. Kode Program Normalisasi Data.

Tabel 3. Hasil Normalisasi Data.

Bulan	Jumlah Konsumen	Jumlah Liter
1	0.035088	0.035088
2	0.017544	0.0625
3	0.000000	0.000
4	0.017544	0.03125
5	0.017544	0.071429

6	0.035088	0.107143
7	0.000000	0.035714
8	0.035088	0.053571
9	0.052632	0.089286

Split Data

Setelah data dinormalisasi, tahap selanjutnya adalah membagi data menjadi dua bagian, yaitu data latih dan data uji. Pembagian ini bertujuan agar model dapat dilatih menggunakan sebagian data dan kemudian diuji menggunakan data yang belum pernah dilihat sebelumnya untuk mengetahui performa prediksinya. Pada penelitian ini, pembagian data dilakukan dengan rasio rasio 90% untuk data latih dan 10% untuk data uji, sebagaimana terlihat dalam Gambar 3 kode program split data.

```
#Split Data
X_train_SVM, X_test_SVM, y_train_SVM, y_test_SVM =
train_test_split(
X_scaled, y_scaled, test_size=0.2, random_state=42)
```

Gambar 3. Kode Program Split Data.

Penerapan Metode *Support Vector Machine*

Pada tahap pelatihan menggunakan metode Support Vector Machine (SVM), data dibagi menjadi 80% sebagai data latih. Model kemudian dilatih untuk mengenali pola hubungan antara variabel input (bulan, jumlah konsumen, dan jumlah liter) dengan output berupa penjualan Solar Industri. Hasil dari proses pelatihan ditunjukkan pada Tabel 4.

Tabel 4 . Hasil Pelatihan SVM.

Data	Aktual (Liter)	Prediksi SVM (Liter)	Persentase Error (%)
1	60000	60836	1.39%
2	95000	101702	7.05%
3	80000	102521	28.15%
4	130000	102137	21.43%
5	45000	65778	46.17%
6	75000	67852	9.53%
7	40000	69491	73.73%
8	240000	211931	11.70%
9	40000	32773	18.07%
10	40000	44321	10.80%
11	50000	67852	35.70%
12	60000	60426	0.71%
13	60000	78581	30.97%
14	50000	51337	2.67%

15	180000	139674	22.40%
16	30000	33183	10.61%
17	130000	102137	21.43%
18	20000	33593	67.97%
19	95000	112430	18.35%
20	25000	40608	62.44%
21	80000	95505	19.38%
22	125000	139674	11.74%
23	55000	63730	15.87%
24	105000	102931	1.97%
25	75000	102931	37.24%
26	85000	78171	8.03%
27	109999	86031	21.79%
28	30000	48444	61.48%
29	75000	95505	27.34%
30	35000	58762	67.89%
31	85000	76532	9.96%
32	45000	60836	35.19%

Secara umum, model SVM mampu memberikan hasil prediksi yang cukup mendekati nilai aktual, meskipun pada beberapa data masih terdapat selisih yang cukup besar. Hal ini terlihat dari persentase error yang bervariasi, mulai dari sangat kecil (0,71%) hingga cukup tinggi (73,73%). Variasi ini menunjukkan bahwa performa SVM sangat bergantung pada distribusi data latih serta sensitivitas parameter kernel yang digunakan.

Penerapan Metode *Random Forest*

Proses pelatihan pada metode *Random Forest* juga menggunakan 80% data latih dengan atribut input yang sama. Random Forest membangun banyak pohon keputusan secara acak, kemudian menggabungkan hasil prediksi dari tiap pohon untuk menghasilkan nilai akhir. Hasil pelatihan dapat dilihat pada Tabel 5.

Tabel 5. Hasil Pelatihan Random Forest.

Data	Aktual (Liter)	Prediksi SVM (Liter)	Persentase Error (%)
1	60000	60836	1.39%
2	95000	101702	7.05%
3	80000	102521	28.15%
4	130000	102137	21.43%
5	45000	65778	46.17%
6	75000	67852	9.53%
7	40000	69491	73.73%
8	240000	211931	11.70%
9	40000	32773	18.07%
10	40000	44321	10.80%
11	50000	67852	35.70%
12	60000	60426	0.71%
13	60000	78581	30.97%

14	50000	51337	2.67%
15	180000	139674	22.40%
16	30000	33183	10.61%
17	130000	102137	21.43%
18	20000	33593	67.97%
19	95000	112430	18.35%
20	25000	40608	62.44%
21	80000	95505	19.38%
22	125000	139674	11.74%
23	55000	63730	15.87%
24	105000	102931	1.97%
25	75000	102931	37.24%
26	85000	78171	8.03%
27	109999	86031	21.79%
28	30000	48444	61.48%
29	75000	95505	27.34%
30	35000	58762	67.89%
31	85000	76532	9.96%
32	45000	60836	35.19%

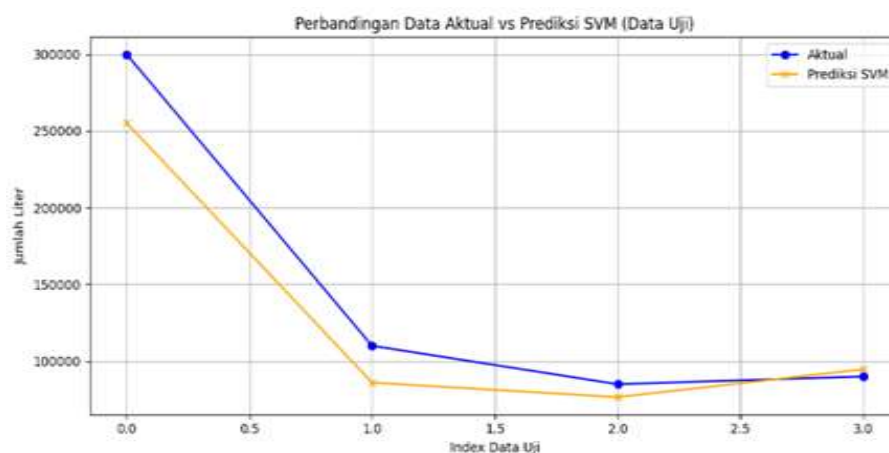
Hasil yang diperoleh pada Random Forest menunjukkan pola error yang mirip dengan SVM, karena memang kedua metode diuji menggunakan data yang sama. Namun, Random Forest cenderung lebih stabil dalam memprediksi data dengan variasi yang besar, walaupun pada beberapa titik error tetap tinggi. Secara keseluruhan, metode ini mampu memberikan hasil prediksi yang konsisten karena sifat ensemble learning yang robust terhadap outlier.

Pengujian Metode *Support Vector Machine*

Setelah proses pelatihan, model SVM diuji menggunakan 20% data uji yang belum pernah dilihat sebelumnya oleh model. Hasil pengujian disajikan pada Tabel 6 dan divisualisasikan pada Gambar 4. Dari hasil pengujian diperoleh nilai Mean Absolute Percentage Error (MAPE) sebesar **12,48%**. Nilai ini menunjukkan bahwa model SVM cukup baik dalam melakukan prediksi penjualan Solar Industri, dengan tingkat kesalahan rata-rata sekitar 12% dari data aktual.

Tabel 6. Hasil pengujian SVM.

Data	Aktual (Liter)	Prediksi <i>SVM</i> (Liter)	Persentase Error (%)
1	300000	255280	14.91%
2	110000	86031	21.79%
3	85000	76532	9.96%
4	90000	94686	5.21%



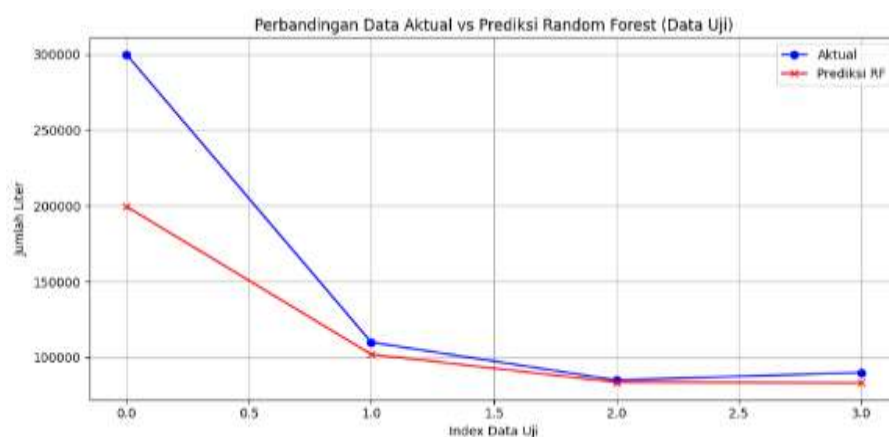
Gambar 4. Grafik pengujian SVM.

Pengujian Metode *Random Forest*

Model Random Forest juga diuji menggunakan data uji yang sama agar hasil perbandingan lebih objektif. Hasil pengujian ditunjukkan pada Tabel 7 dan divisualisasikan pada Gambar 5. Berdasarkan hasil uji, diperoleh nilai MAPE sebesar **12,97%**. Nilai ini sedikit lebih tinggi dibandingkan dengan SVM, yang berarti tingkat kesalahan prediksi Random Forest sedikit lebih besar.

Tabel 7. Hasil pengujian Random Forest.

Data	Aktual (Liter)	Prediksi RF (Liter)	Persentase Error (%)
1	300000	199601	33.46%
2	110000	102050	7.227%
3	85000	83620	1.623%
4	90000	83157	7.603%



Gambar 5. Grafik Pengujian Random Forest.

Namun demikian, meskipun selisih nilai MAPE keduanya relatif kecil, Random Forest tetap menunjukkan keunggulan pada aspek kestabilan model dan kemampuan menghindari

overfitting, sedangkan SVM lebih sensitif terhadap pemilihan parameter dan data yang memiliki variasi ekstrem.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode Support Vector Machine (SVM) dan Random Forest sama-sama mampu memprediksi penjualan Solar Industri (HSD) di PT Heva Petroleum Energi Palembang dengan tingkat akurasi yang cukup baik. Hasil pengujian menunjukkan bahwa model SVM memperoleh nilai MAPE sebesar 12,48%, sedangkan Random Forest menghasilkan nilai MAPE sebesar 12,97%. Hal ini menunjukkan bahwa SVM memiliki tingkat akurasi prediksi yang sedikit lebih baik dibandingkan Random Forest, meskipun perbedaannya relatif kecil. Namun demikian, Random Forest tetap unggul dalam hal kestabilan model dan kemampuannya menghadapi variasi data yang kompleks. Dengan demikian, tujuan penelitian ini tercapai, yaitu membandingkan performa kedua metode dalam memprediksi penjualan Solar Industri, dan hasilnya dapat dijadikan dasar pertimbangan perusahaan dalam menentukan strategi distribusi dan perencanaan stok yang lebih efektif.

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis sampaikan kepada Ibu Dr.Shinta Puspasari, S.Kom.,M.Kom. selaku pembimbing pembimbing I (penulis 2) dan Ibu Evi Purnamasari, S.SI., M.Kom. selaku pembimbing II (penulis 3) yang telah memberikan arahan, bimbingan, serta masukan yang sangat berharga dalam penyusunan skripsi ini hingga dapat terselesaikan dengan baik. Segala perhatian, ilmu, dan dukungan yang diberikan menjadi landasan penting bagi penulis dalam menyelesaikan penelitian ini.

DAFTAR REFERENSI

- Andika, R. (2024). Penerapan model eksponensial dan logistik dalam prediksi populasi: Studi kasus Kota Palembang. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(2), 853–861. <https://doi.org/10.23960/jitet.v12i2.4005>
- Arfan, U., & Paraga, N. (2024). Perbandingan algoritma K-Means, Naïve Bayes dan Decision Tree dalam memprediksi penjualan bahan bakar minyak. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(4), 1379–1389. <https://doi.org/10.57152/malcom.v4i4.1566>
- Cahya, D., Buani, P., Rahmawati, A., Informatika, P. S., Mandiri, U. N., Studi, P., Informasi, S., Mandiri, U. N., & Forest, R. (2024). Klasifikasi mental disorder dengan menggunakan. *[Nama jurnal tidak tersedia]*, 9(2), 101–109.

- Darmawan, A., Yudhisari, I., Anwar, A., & Makruf, M. (2023). Pola prediksi kelulusan siswa Madrasah Aliyah Swasta dengan Support Vector Machine dan Random Forest. *Jurnal Minfo Polgan*, 12(1), 387–400. <https://doi.org/10.33395/jmp.v12i1.12388>
- Direktur Jenderal Minyak Gas dan Bumi KESDM. (2023). *Standar dan mutu (spesifikasi) bahan bakar minyak jenis minyak solar dengan campuran biodiesel (B100) sebesar 35% (B35) dengan angka setana (CN) 48* (pp. 5–6). Situs Ditjen Migas. <https://migas.esdm.go.id/cms/uploads/regulasi/regulasi-kkkl/2023/170.K.HK.02.DJM.2023.pdf>
- Duran, P. A., Vitianingsih, A. V., Riza, M. S., Maukar, A. L., & Wati, S. F. A. (2024). Data mining untuk prediksi penjualan menggunakan metode simple linear regression. *Teknika*, 13(1), 27–34. <https://doi.org/10.34148/teknika.v13i1.712>
- Fitri, E. (2023). Analisis perbandingan metode regresi linier, Random Forest Regression dan Gradient Boosted Trees Regression method untuk prediksi harga rumah. *Journal of Applied Computer Science and Technology*, 4(1), 58–64. <https://doi.org/10.52158/jacost.v4i1.491>
- Gustriansyah, R., Suhandi, N., Puspasari, S., & Sanmorino, A. (2024). Machine learning method to predict the toddlers' nutritional status. *Jurnal Infotel*, 16(1), 32–43. <https://doi.org/10.20895/infotel.v15i4.988>
- Lutfi, M. (2021). Pemanfaatan limbah oli bekas menjadi bahan bakar high speed diesel (HSD). *JST (Jurnal Sains Terapan)*, 7(1), 57–62. <https://doi.org/10.32487/jst.v7i1.1121>
- Miftahul Jannah, Haviz, M. H., Dewi Sartika, & Evi Purnamasari. (2023). Prediksi penjualan produk pada PT Bintang Sriwijaya Palembang menggunakan K-Nearest Neighbour. *Jurnal Software Engineering and Computational Intelligence*, 1(2), 80–89. <https://doi.org/10.36982/jseci.v1i2.3542>
- Purnamasari, E., Verano, D. A., Informatika, T., Indo, U., & Mandiri, G. (2025). Model data-driven untuk prediksi digitalisasi UMKM menggunakan GMM dan XGBoost. *[Nama jurnal tidak tersedia]*, 5(2), 204–214.
- Sari, L., Romadloni, A., & Listyaningrum, R. (2023). Penerapan data mining dalam analisis prediksi kanker paru menggunakan algoritma Random Forest. *Infotekmesin*, 14(1), 155–162. <https://doi.org/10.35970/infotekmesin.v14i1.1751>
- Suhandi, N., Gustriansyah, R., Destria, A., Amalia, M., & Kris, V. (2024). Prediksi kualitas susu menggunakan metode K-Nearest Neighbors. *[Nama jurnal tidak tersedia]*, 14(2), 197–208.
- Wijoyo, A. A. S., Ristanti, S., Sya'ban, S., Amalia, M., & Febriansyah, R. (2024). Pembelajaran machine learning. *OKTAL (Jurnal Ilmu Komputer dan Science)*, 3(2), 375–380. <https://journal.mediapublikasi.id/index.php/oktal/article/view/2305>
- Yunita, R., & Kamayani, M. (2023). Perbandingan algoritma SVM dan Naïve Bayes pada analisis sentimen penghapusan kewajiban skripsi. *Indonesian Journal of Computer Science*, 12(5), 2879–2890. <https://doi.org/10.33022/ijcs.v12i5.3415>