New Data Prediction and Accuracy

Try to predict data on WEC machine on another location tasmania. Which is totally new. so predict how accurate our model to predict total new data

In [86]:
```
from joblib import import load
new_data = pd.read_csv('Tasmania_Data.csv',header=None)
new_data
```

Out[86]:

	0	1	2	3	4	5	6	7	
0	546.1931	194.5337	160.0840	435.5412	220.8238	449.5166	396.1027	231.7794	69.
1	281.0669	390.3761	561.0742	295.9217	565.5344	236.3035	393.0648	340.0667	213.
2	566.0000	566.0000	346.5334	202.2120	389.6777	277.7876	421.9610	260.3339	464.
3	2.1045	352.3969	285.2965	566.0000	532.4306	176.8103	566.0000	0.0000	234.
4	168.5854	550.6155	0.0000	566.0000	450.8427	0.0000	566.0000	566.0000	458.
...	...	...	...	...	...	...	...	...	
71995	447.4090	80.6710	74.8115	495.8420	520.5531	205.9689	20.4930	541.6594	368.
71996	565.5693	522.9972	437.3214	88.5190	8.5200	429.1224	564.0161	38.1109	339.
71997	49.7652	90.4154	110.6000	532.7479	428.7800	101.1560	13.6231	506.2370	450.
71998	539.4314	154.6262	493.8225	2.3327	217.3275	480.6632	565.6854	165.0820	83.
71999	565.6854	98.0576	437.2540	58.9012	160.7589	424.0946	509.1169	221.6505	139.

72000 rows × 49 columns

In [87]:
```
# Separate features and target
X_new = new_data.iloc[:, :-1]
y_true = new_data.iloc[:, -1]
```

In [88]:
```
# Load the saved scaler
scaler = load("scaler.joblib")
X_new_scaled = scaler.transform(X_new)
```

In [89]:
```
# Load the saved best model
best_saved_model= load("Best_Model_SVM_Linear_Kernel.joblib")
best_saved_model
```

Out[89]:

SVR

```
SVR(kernel='linear')
```

In [91]:
```
# Predict using the loaded model
y_pred = best_saved_model.predict(X_new_scaled)
```

In [92]:
```
# Evaluate the performance
rmse = mean_squared_error(y_true, y_pred)
r2 = r2_score(y_true, y_pred)

# Print evaluation results
print(f"RMSE on New Data: {rmse:.4f}")
print(f"R^2 Score on New Data: {r2:.4f}")

# Save predictions alongside true values for further analysis
results = pd.DataFrame({
    "Actual": y_true,
    "Predicted": y_pred
})
```

```
RMSE on New Data: 19177.6745
R^2 Score on New Data: 1.0000
```

In [93]:
```
results.to_csv("new_data_predictions.csv", index=False)
print("Predictions saved to 'new_data_predictions.csv'.")
```

```
Predictions saved to 'new_data_predictions.csv'.
```

In [ ]:

Observation on Best Model's Performance on New Data (Tasmania) Performance Metrics: RMSE: 19,177.67, R^2: 1.0000

1. **Interpretation of R^2 Score:** The $R2$ score of 1.0000 indicates that the model perfectly explains the variance in the target variable (Powerall) for the new dataset. This is an exceptional result, suggesting the model is highly effective in capturing the underlying patterns and relationships in the data.

2. **RMSE Interpretation:** The RMSE of 19,177.67 represents the average prediction error in the same units as the target variable (Powerall). While the RMSE value might appear high, it must be compared to the scale of Powerall in the new dataset. If Powerall values are in a similar range, this RMSE could indicate a very accurate prediction.

3. **Consistency Across Datasets:** The performance on the new dataset (Tasmania) aligns with the model's exceptional results on the previous datasets. This consistency demonstrates that the model generalizes well to unseen data, reinforcing its reliability and robustness.

4. **Insights on Data and Model:** The high $R2$ score and acceptable RMSE suggest that the placement of WECs and their features in Tasmania align with patterns learned from the Adelaide dataset. The SVM Linear Kernel model effectively captures these relationships, making it highly suitable for predicting Powerall in similar contexts.

5. **Model Robustness:** Achieving such a high $R2$ score on a completely new dataset (Tasmania) indicates the model's robustness. It shows the model's ability to generalize its learning across different wave energy converter configurations and power outputs.

The End