Support Vector Regression (SVR) Analysis Report

1. Introduction

Support Vector Regression (SVR) is a machine learning model used for predicting continuous values (like prices, temperatures, etc.). Unlike simple linear regression, SVR can handle complex relationships in data using different **kernel functions**. This report explains how different SVR models performed on a given dataset and which one works best.

---

2. Methodology

2.1 How SVR Works

SVR tries to fit the best possible line (or curve) while keeping errors within a certain margin (called **epsilon**). The equation is:

$Y=wX+b$ $Y=wX+b$

- **Y** = Target variable (what we predict).
- **X** = Features (input data).
- **w** = Weights (how much each feature matters).
- **b** = Bias (starting point).

2.2 Steps Taken

Step 1: Data Preprocessing

- **Checked for missing data** (no empty values).
- **Exploratory Data Analysis (EDA):** Studied how features relate to the target.
- **Encoded categorical data** (converted text into numbers).
- **Scaled features** (to ensure all features contribute equally).
- **Split data:**
    - **70% Training set** (to train the model).
    - **30% Testing set** (to check performance).

Step 2: Model Training

- Tested **different kernel functions**:
    - **Linear** (straight line).
    - **Polynomial** (curved line).
    - **RBF (Radial Basis Function)** (flexible curve).
    - **Sigmoid** (S-shaped curve).
- Tuned **hyperparameters** (C, epsilon, gamma) to improve accuracy.

Step 3: Performance Evaluation

Measured using:

- **R^2 Score (0-1):** Closer to 1 = better fit.
- **Mean Absolute Error (MAE):** Average prediction error.
- **Mean Squared Error (MSE):** Larger errors penalized more.

3. Model Comparison & Results

Model Type	Kernel	Key Settings	R^2 Score	Performance
Linear SVR	linear	Default	**0.8950**	Best model (strong linear trend).
Polynomial SVR (Degree 3)	poly	C=1.0	**-0.0508**	Very poor (data not polynomial).
RBF SVR (Default)	rbf	C=10	**-0.0558**	Bad (default settings not good).
Sigmoid SVR	sigmoid	C=1.0	**-0.0574**	Worst (sigmoid doesn't fit).
Optimized RBF SVR	rbf	C=100, epsilon=0.1	**-0.0302**	Better than default but still bad.

Key Findings:

1. **Linear SVR works best (R^2 = 0.8950)** → Data has a strong linear trend.
2. **Polynomial & Sigmoid kernels fail** → Data doesn't fit curves well.
3. **RBF kernel can improve with tuning**, but still worse than Linear SVR.
4. **Higher C (C=100) helps slightly**, but not enough to beat Linear SVR.

4. Recommendations

Best Model Choice:

Use Linear SVR (it's simple and works best).

Avoid These Models:

- **Polynomial & Sigmoid SVR** (they perform terribly).
- **Default RBF SVR** (needs heavy tuning to be useful).

If You Want to Try Non-Linear Models:

- **Optimize RBF SVR further** (test different gamma, C, and epsilon).
- **Try feature engineering** (create new features to help non-linear kernels).

General Tips:

✓ **Always scale features** (SVR is sensitive to feature scales).
✓ **Start with Linear SVR** (if it works well, no need for complex kernels).
✓ **Use cross-validation** to find the best hyperparameters.

---

5. Conclusion

- **Best Model: Linear SVR ($R^2 = 0.8950$)** → Simple and effective.
- **Avoid:** Polynomial, Sigmoid, and default RBF kernels (they fail).
- **Tuning RBF helps but isn't worth it yet** (stick with Linear SVR unless data changes).

Final Decision: Use **Linear SVR** for reliable predictions.