Regression Workflow Report (LinearSVR)

1. Data Preparation

- **Dataset:** Loaded from `/content/drive/MyDrive/AI/selected_features_dataset.csv` with `actual_time` as the target variable
- **Preprocessing Steps:**
  - **Categorical Encoding:** Used One-Hot Encoding with `drop_first=True` to prevent the dummy variable trap.
  - **Splitting:** The data split into training (80%) and test (20%) sets with a fixed random seed for reproducibility.
  - **Scaling:** Features standardized using `StandardScaler` to ensure all predictors are on the same scale before model training.

2. Models Evaluated

Model	R^2 Score	MSE
Linear Regression	0.979157	2.072741
Ridge Regression	0.979145	2.073961
Linear SVR	0.977161	2.271192

a. Linear Regression

- Standard linear regression applied to scaled features.
- Achieved R^2 score of **0.9792** and MSE of **2.0727** on the test set.
- Indicates good model fit and reasonable prediction accuracy.

b. Ridge Regression

- Applied L2 regularization with `alpha=1.0` to reduce overfitting and manage multicollinearity.
- R^2 score of **0.9791** and MSE of **2.0740**, nearly identical to plain linear regression, confirming limited benefit from regularization in this context.

c. Linear SVR

- Employed linear Support Vector Regression (`LinearSVR`) with `epsilon=0.1` and `max_iter=10000`.
- Model raised a **convergence warning**, suggesting it did not reach the optimal solution within the iteration limit. Further tuning or increasing `max_iter` may be required.
- R^2 score of **0.9772** and MSE of **2.2712**. Performance slightly lower than linear regression models.

3. Notes

- **Feature Scaling:** Essential for both linear and SVR models to ensure effective model training.
- **Convergence Warning:** For LinearSVR, the model's warning indicates that results might improve with more iterations or different hyperparameters.
- **Model Interpretability:** All three models maintain interpretability, with linear regression offering the most straightforward insight into feature effects.

4. Summary

- All three models deliver strong predictive performance ($R^2 \approx 0.98$).
- Linear regression and Ridge regression perform almost identically.
- Linear SVR slightly trails but is still effective; convergence warning should be addressed for optimal results