**Lab2  
1. Dataset Selection & Preparation**

* Choose a real-world dataset (e.g., house prices, stocks, medical data).
  + https://www.kaggle.com/datasets/atharvasoundankar/big-4-financial-risk-insights-2020-2025
* Preprocess data (handle missing values, normalize if needed).
  + No missing values

**2. Test Set Selection (No Training)**

* Fix 10 test points (randomly or last 10 rows).

**3. Model Evaluation (Pretrained Models)**

* Use pretrained **Linear Regression** & **Ridge Regression (sag solver)**.
* Ridge settings:
  + Iterations: [100, 500, 1000]
  + Alpha: [0.1, 1.0, 10]

**4. Performance Metrics**

* Predict 10 test points using both models.
* Compute:
  + **MAE** (Mean Absolute Error)
  + **MSE** (Mean Squared Error)
  + **RMSE** (Root Mean Squared Error)
  + **MAPE** (Mean Absolute Percentage Error)

**Linear Regression Metrics:** MAE: 121.79, MSE: 17099.50, RMSE: 130.77, MAPE: 0.61%

**Ridge Regression Metrics**

Alpha: 0.1, Max Iter: 100: MAE: 145.64, MSE: 23445.84, RMSE: 153.12, MAPE: 0.76%

Alpha: 0.1, Max Iter: 500: MAE: 148.17, MSE: 24402.62, RMSE: 156.21, MAPE: 0.77%

Alpha: 0.1, Max Iter: 1000: MAE: 148.16, MSE: 24435.38, RMSE: 156.32, MAPE: 0.77%

Alpha: 1.0, Max Iter: 100: MAE: 145.71, MSE: 23469.18, RMSE: 153.20, MAPE: 0.76%

Alpha: 1.0, Max Iter: 500: MAE: 148.16, MSE: 24398.87, RMSE: 156.20, MAPE: 0.77%

Alpha: 1.0, Max Iter: 1000: MAE: 148.16, MSE: 24435.47, RMSE: 156.32, MAPE: 0.77%

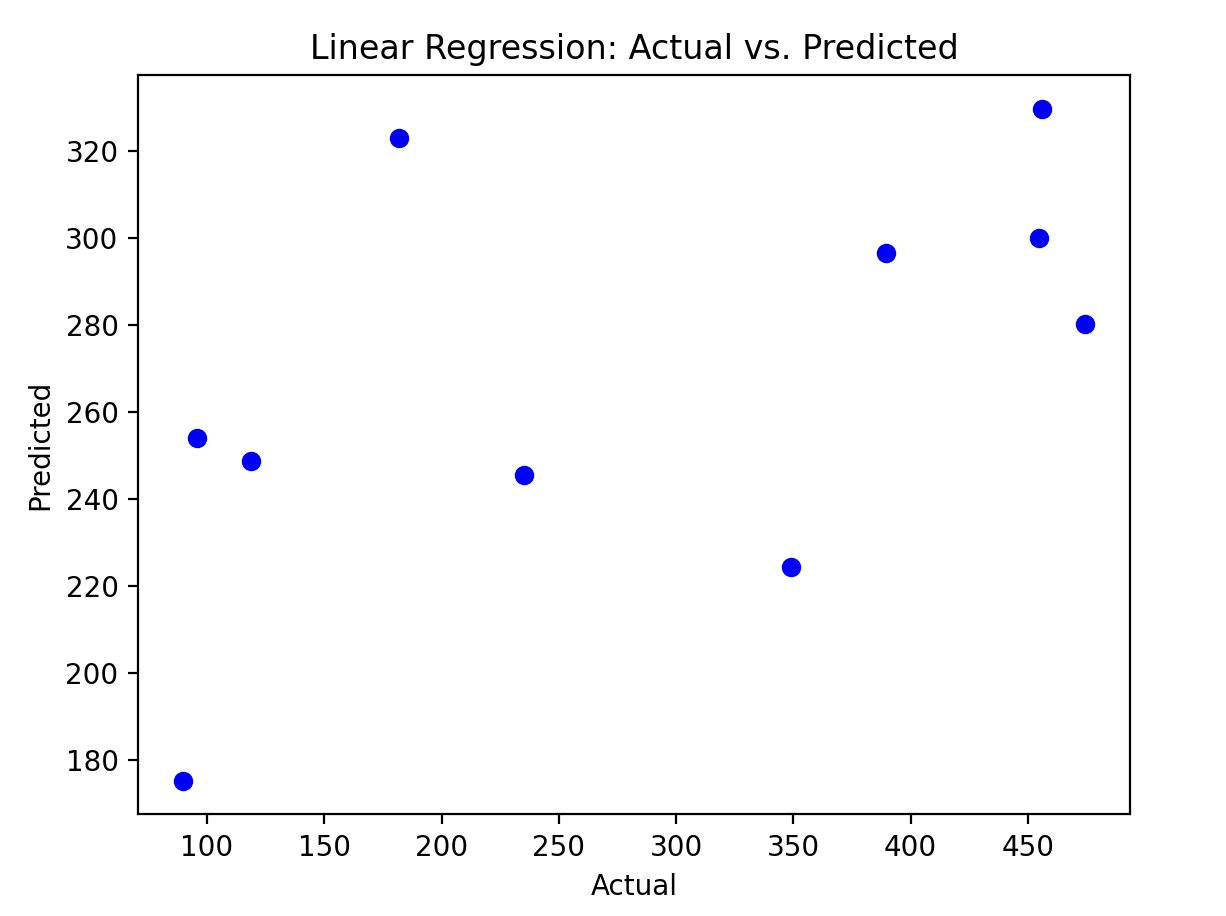
Alpha: 10, Max Iter: 100: MAE: 145.73, MSE: 23477.38, RMSE: 153.22, MAPE: 0.76%

Alpha: 10, Max Iter: 500: MAE: 148.16, MSE: 24397.37, RMSE: 156.20, MAPE: 0.77%

Alpha: 10, Max Iter: 1000: MAE: 148.16, MSE: 24435.34, RMSE: 156.32, MAPE: 0.77%

**5. Visualization & Interpretation**

* **Scatter Plot:** Show actual vs. predicted values.



* **Line Plot:** Compare Linear & Ridge predictions across different alpha values.  
  A graph with red and green lines

  AI-generated content may be incorrect.

A graph with lines and dots

Description automatically generated

**6. Optimization & Analysis**

* Adjust alpha & iterations if errors are high.
* **Final Discussion:**

กราฟที่แสดงการทำนายของโมเดล Linear Regression และ Ridge Regression มีลักษณะไม่เป็นเส้นตรง (non-linear) และมีการขึ้นลงที่ไม่สม่ำเสมอ ซึ่งบ่งบอกถึงการ overfitting และความแม่นยำในการทำนายที่ต่ำ และอาจจะเพราะข้อมูลที่มีน้อยด้วย