RIDGE REGRESSION MODEL – TRAINING & EVALUATION DOCUMENTATION

1. INTRODUCTION

The purpose of this project is to develop a Ridge Regression model to handle multicollinearity in input features and improve model generalization. Ridge Regression introduces L2 regularization to reduce model variance and prevent overfitting.

2. OBJECTIVE

- Train a Ridge Regression model
- Tune the alpha parameter to find the optimal regularization strength
- Evaluate performance using MAE, MSE, RMSE, and R^2
- Save the final trained model
- Analyze bias–variance behaviour
- Visualize performance across different alpha values

3. DATASET

The dataset was split into:

- **Training set (X_train, y_train)**
- **Testing set (X_test, y_test)**

Standardization was applied before model training.

4. METHODOLOGY

4.1 RIDGE REGRESSION TRAINING

Ridge Regression formula:

$$\boxed{\beta = (X^T X + \alpha I)^{-1} X^T y}$$

- Alpha values tested:
  [0.01, 0.1, 1, 10, 100, 200, 500]
- Model trained using scikit-learn's Ridge().

4.2 METRICS USED

Metric	Explanation
MSE	Measures squared error (penalizes large errors)
RMSE	Square root of MSE – more interpretable

Metric	Explanation
MAE	Average absolute difference
R^2 Score	Proportion of variance explained

5. HYPERPARAMETER TUNING

For each alpha, the following were recorded:

- Train MSE
- Train RMSE
- Train MAE

Plots were generated:

1. **Train MSE vs Alpha (log scale)**
2. **Train RMSE vs Alpha**
3. **Train MAE vs Alpha**

These graphs show how regularization affects model complexity.

6. UNDERFITTING & OVERFITTING ANALYSIS

Condition	Meaning
High train + high test error	Underfitting
Low train + high test error	Overfitting
Balanced train & test error	Good model

Using these metrics, the best alpha was selected.

7. MODEL SAVING

The final optimized Ridge model was saved as:

ridge.pkl

Using pickle:

```
import pickle

pickle.dump(model, open("ridge.pkl", "wb"))
```

8. RESULTS

- Ridge Regression improved stability in presence of multicollinearity
- Optimal alpha selected based on evaluation metrics
- Performance visualizations validated bias–variance tradeoff
- Final model can be reused anytime using pickle

9. CONCLUSION

Ridge Regression successfully addressed overfitting and improved prediction reliability. Hyperparameter tuning using alpha, combined with evaluation metrics and visualization, ensured an optimized and well-generalized model.