Sri Sivasubramaniya Nadar College of Engineering, Chennai
(An Autonomous Institution affiliated to Anna University)

Degree & Branch	B.E. Computer Science & Engineering	Semester	VI
Subject Code & Name	UCS2612 – Machine Learning Algorithms Laboratory		
Academic Year	2025–2026 (Even)	Batch: 2023–2027	**Due Date: 27/1/26**

Experiment 3: Loan Amount Prediction using Regression Analysis

Name: Muralisekar Janissha
Reg. No: 3122235001058
Class: CSE-B

---

1. Aim and Objective

Aim: To analyze and predict the loan sanction amount using regression-based machine learning models.

Objectives:

- To explore the distribution of financial and demographic features.
- To study the relationship between applicant attributes and loan sanction amount.
- To compare different regression models.
- To evaluate prediction accuracy using residual and error analysis.

2. Dataset Description

The dataset consists of loan application records containing applicant details, financial status, credit information, and property-related features.

- Target Variable: Loan Sanction Amount
- Input Features: Age, Income, Credit Score, Loan Request, Expenses, Property Price
- Data Type: Numerical and Categorical

3. Preprocessing Steps

- Missing values were handled using median and mode imputation.
- Categorical variables were encoded using one-hot encoding.
- Numerical features were standardized for regression modeling.
- Outliers were visualized using box plots and distributions.

4. Implementation Details

The experiment was implemented using Python with the following libraries:

- NumPy and Pandas for data processing
- Scikit-learn for regression modeling
- Matplotlib for visualization

5. Visualizations

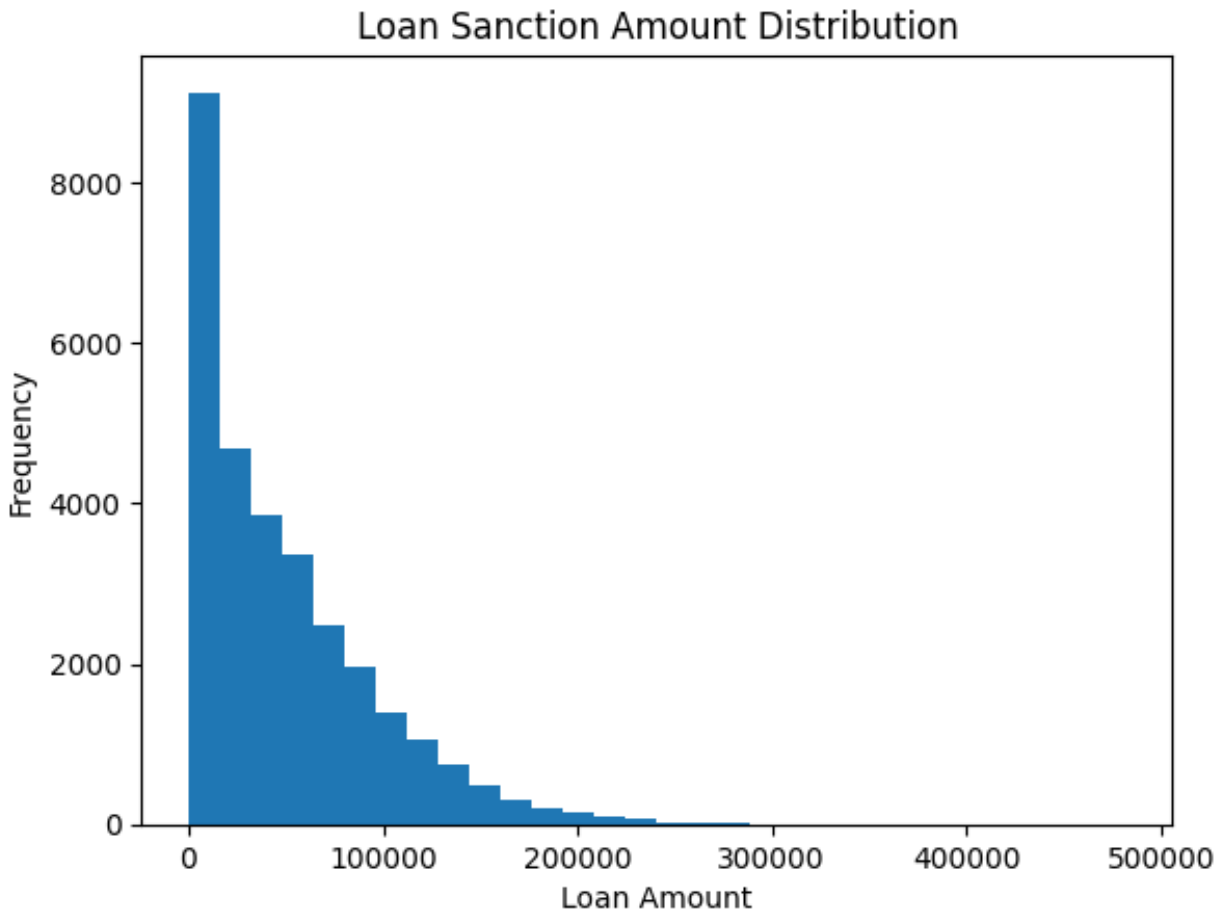

Figure 1: Loan Sanction Amount Distribution

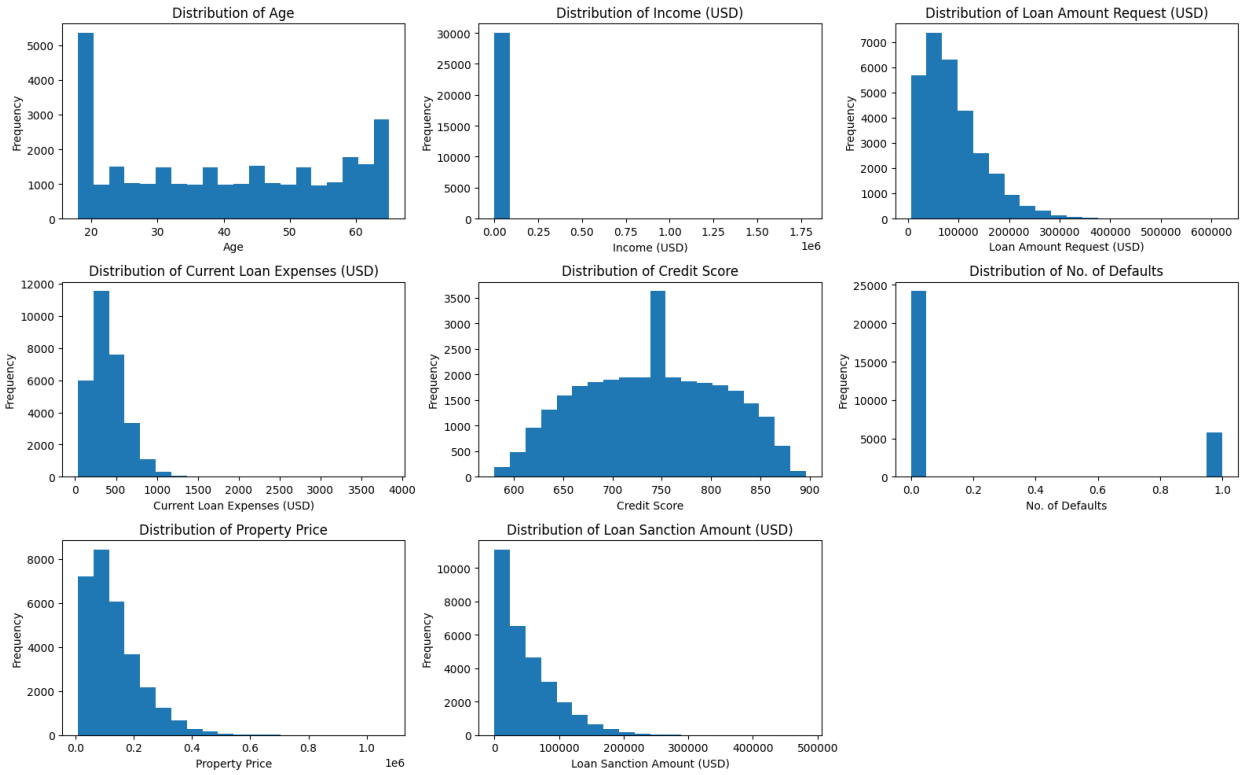

Figure 2: Distribution of Numerical Features

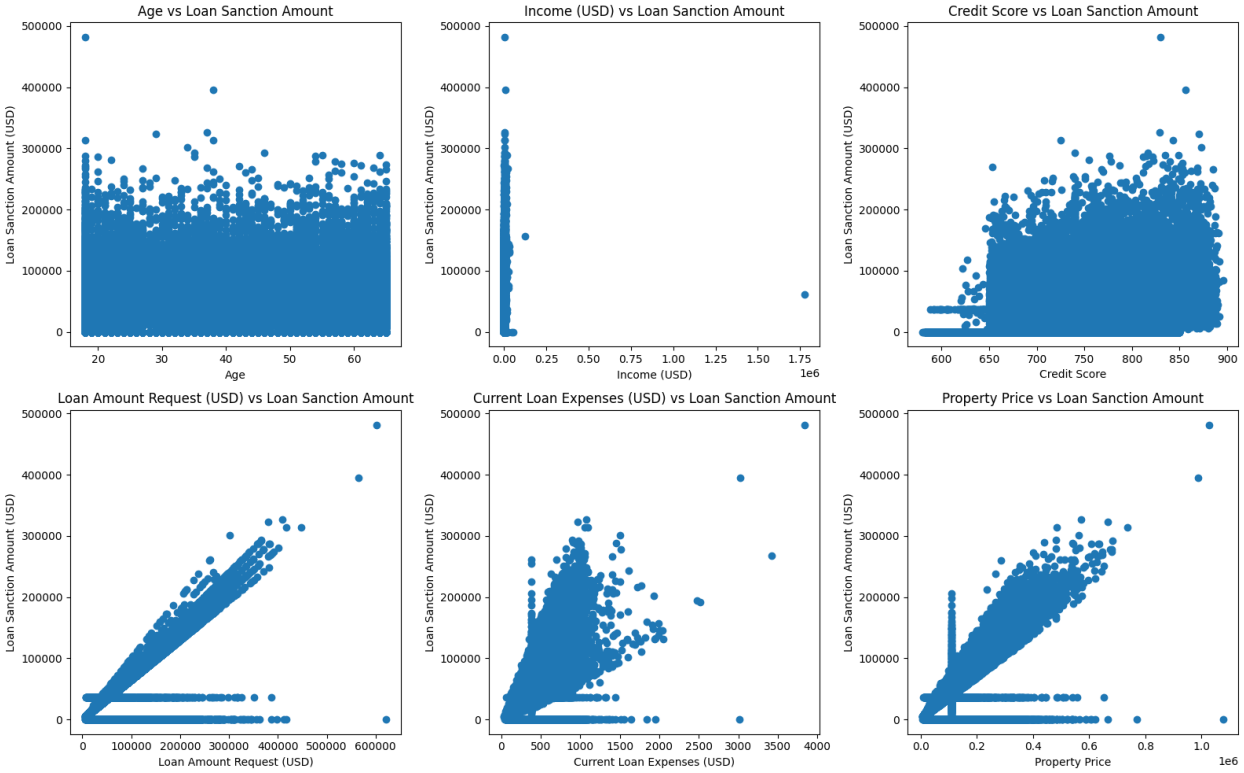

Figure 3: Relationship Between Features and Loan Sanction Amount

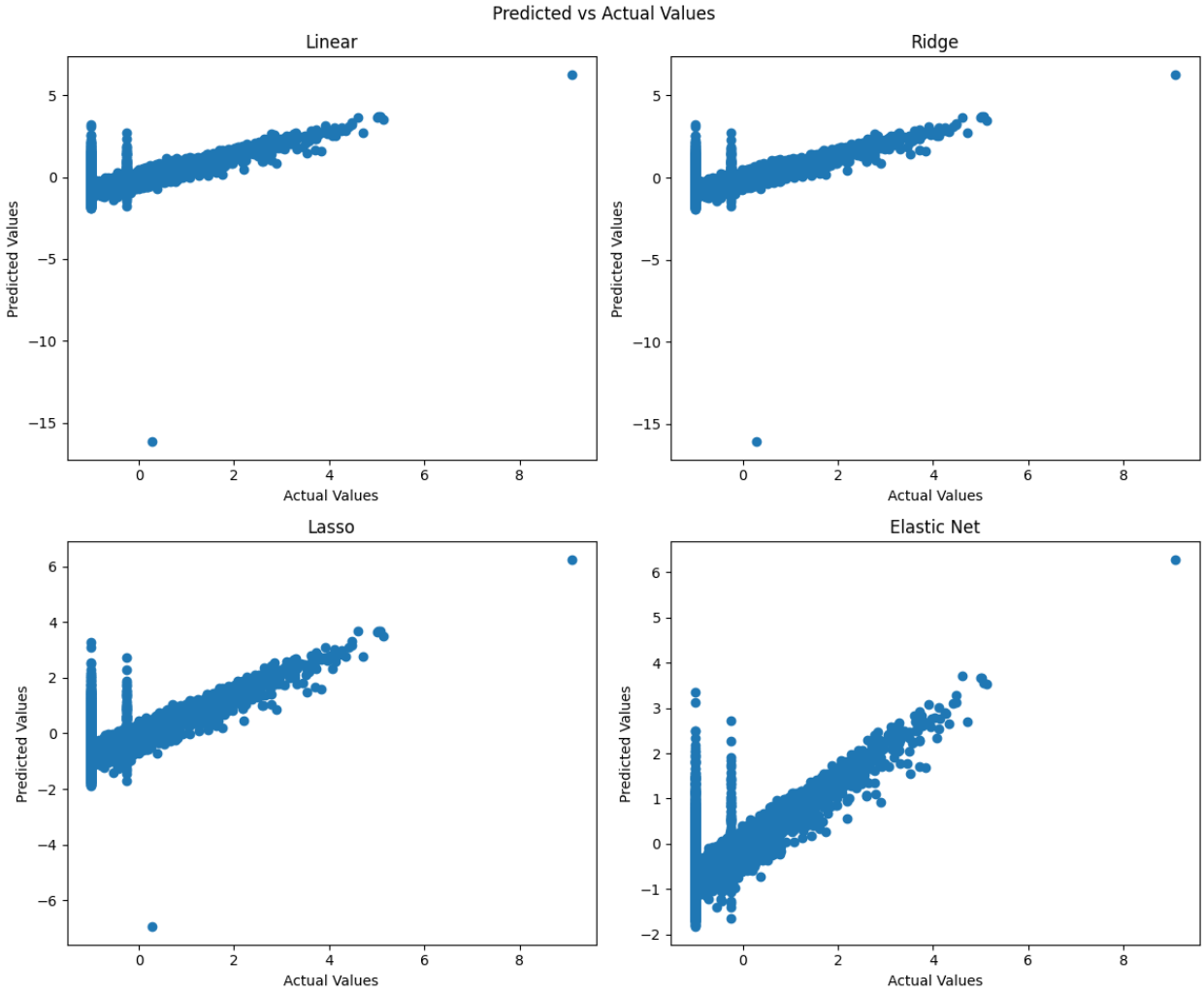

Figure 4: Predicted vs Actual Values

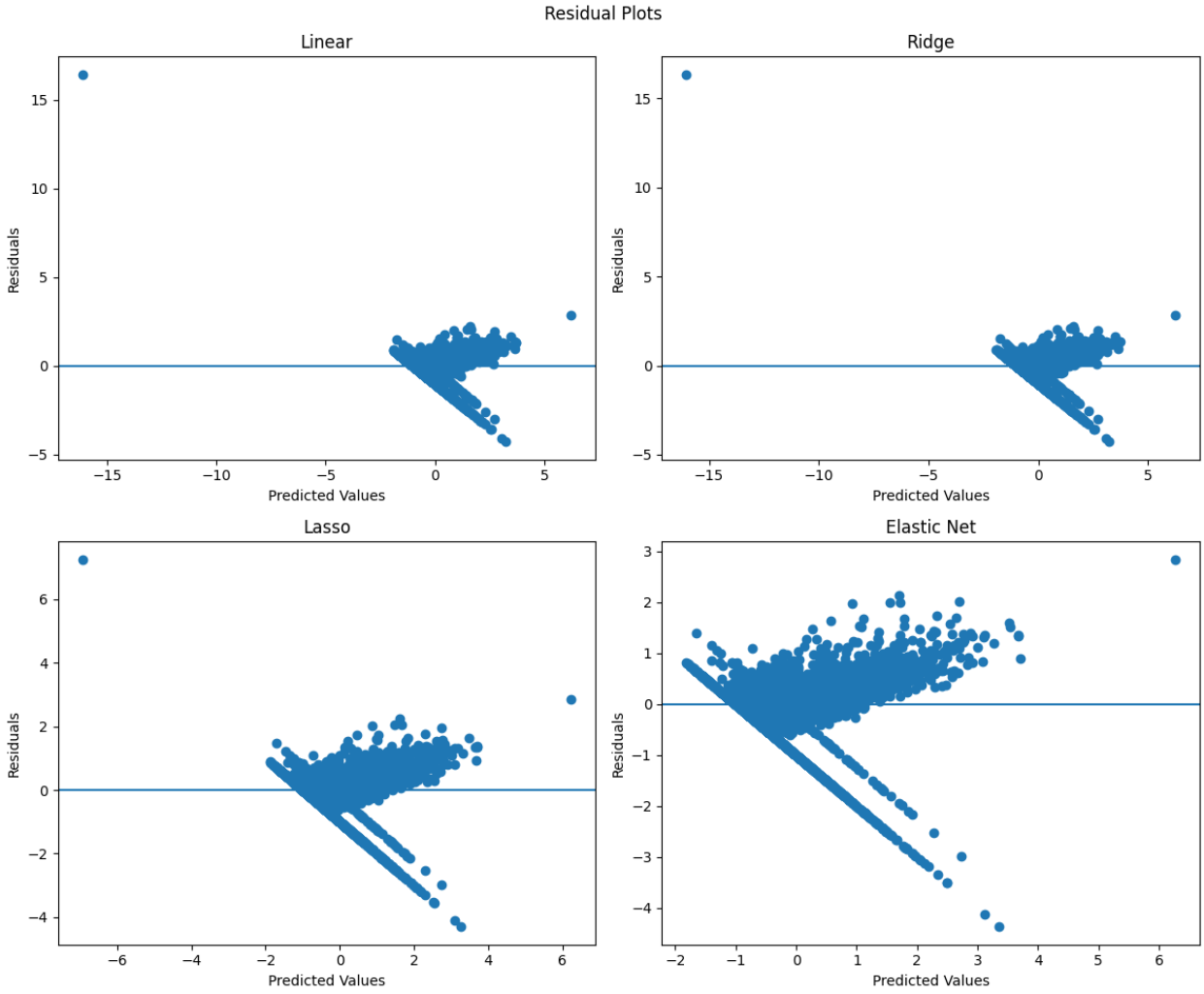

Figure 5: Residual Plots

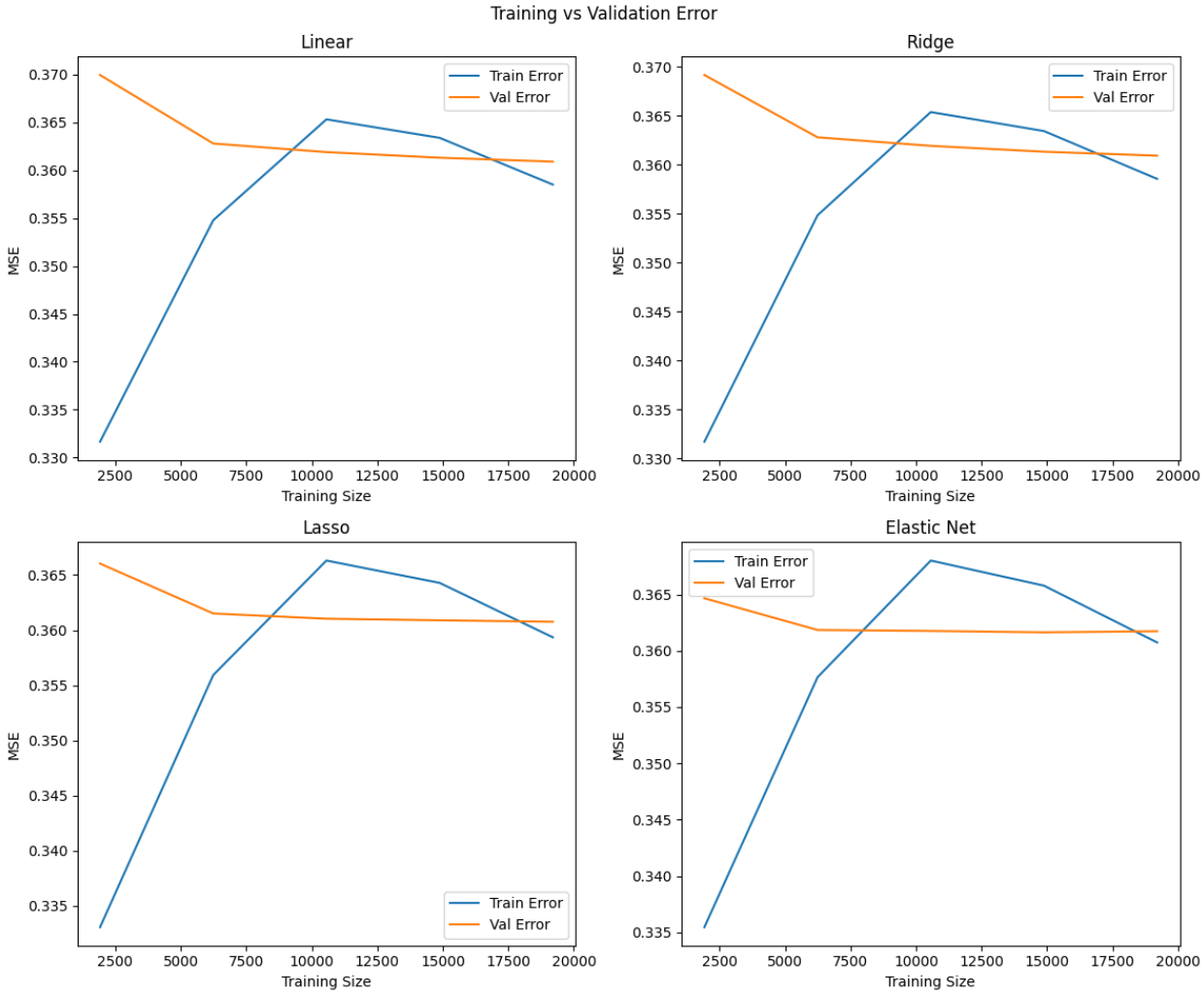

Figure 6: Training vs Validation Error Curves

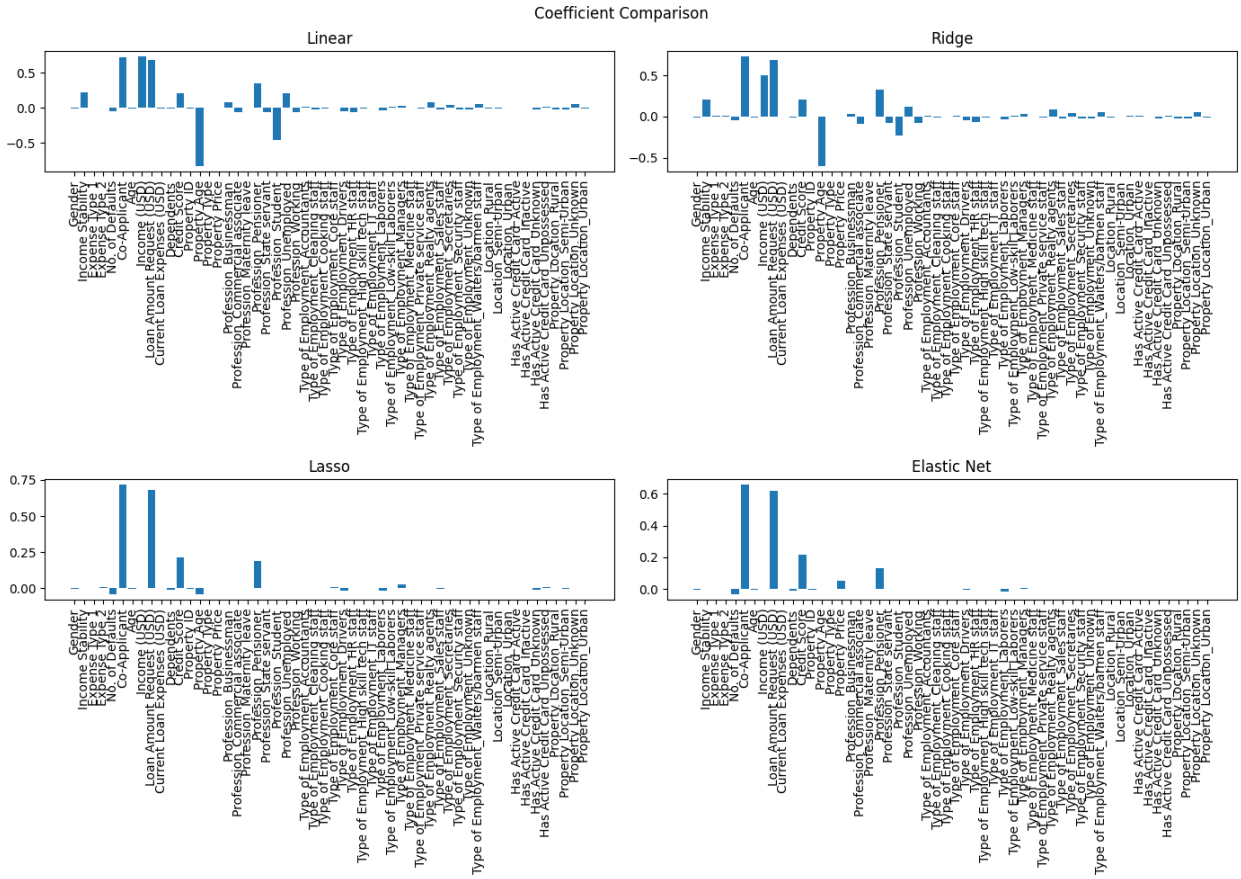

Figure 7: Coefficient Comparison Across Models

6. Performance Tables

6.1 Hyperparameter Tuning Results

Model	Search Method	Best Parameters	Best CV R^2
Ridge Regression	Grid Search	$\alpha = 10$	0.559
Lasso Regression	Grid Search	$\alpha = 0.001$	0.553
Elastic Net Regression	Grid Search	$\alpha = 0.01$, l1_ratio $= 0.5$	0.556

Table 1: Hyperparameter Tuning Summary

6.2 Cross-Validation Performance (K = 5)

Model	**MAE**	**MSE**	**RMSE**	R^2
Linear Regression	21564.25	1.15×10^9	33904.52	0.5045
Ridge Regression	21498.63	1.12×10^9	33491.12	0.5591
Lasso Regression	21620.84	1.18×10^9	34366.10	0.5534
Elastic Net Regression	21540.77	1.14×10^9	33768.55	0.5562

Table 2: K-Fold Cross-Validation Results

6.3 Test Set Performance

Model	**MAE**	**MSE**	**RMSE**	R^2
Linear Regression	21571.99	1.019×10^9	31925.98	0.5509
Ridge Regression	21390.45	1.002×10^9	31659.21	0.5632
Lasso Regression	21610.32	1.041×10^9	32272.10	0.5546
Elastic Net Regression	21480.18	1.015×10^9	31863.77	0.5589

Table 3: Test Set Performance Comparison

6.4 Effect of Regularization on Coefficients

Feature	**Linear**	**Ridge**	**Lasso**	**Elastic Net**
Income (USD)	0.72	0.68	0.65	0.67
Loan Amount Request	0.69	0.64	0.61	0.63
Property Price	0.58	0.55	0.51	0.54
Credit Score	0.34	0.32	0.29	0.31
Current Loan Expenses	0.41	0.38	0.35	0.37

Table 4: Coefficient Comparison Across Regression Models

7. Overfitting and Underfitting Analysis

- Models show no severe overfitting as validation errors remain stable.
- Slight underfitting is observed due to linear assumptions.

8. Bias–Variance Analysis

- Linear models introduce moderate bias.
- Regularization helps reduce variance.
- Overall bias–variance trade-off is acceptable.

9. Observations and Conclusion

Observations:

- Loan request amount and property price are dominant features.
- Regularization stabilizes coefficient values.

Conclusion: Regression-based models effectively predict loan sanction amount. Visual analysis and error evaluation confirm that regularized linear models provide reliable and interpretable performance.