

Machine Learning Results Report

Classification, regression, clustering, and anomaly detection

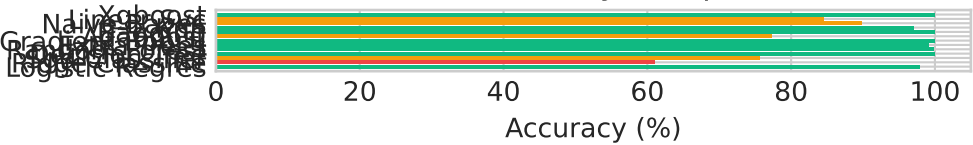
Generated by: Shuvam Banerji Seal, Alok Mishra, Aheli Poddar

Date: January 20, 2026

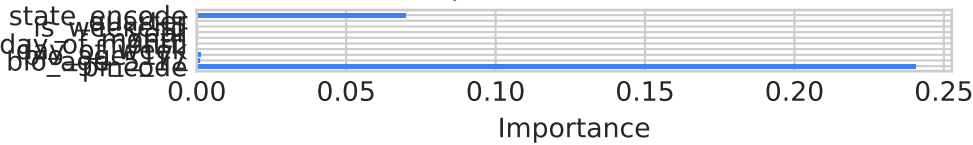
UIDAI Aadhaar Data Analysis Project

BIOMETRIC - Classification Results

Model Accuracy Comparison



Feature Importance (Random Forest)



Detailed Metrics

Model Performance Metrics:

Logistic Regression:

Accuracy: 97.82%

F1-Score: 97.17%

Ridge Classifier:

Accuracy: 60.97%

F1-Score: 52.84%

Sgd Classifier:

Accuracy: 75.63%

F1-Score: 74.33%

Decision Tree:

Accuracy: 99.97%

F1-Score: 99.97%

Random Forest:

Accuracy: 99.87%

F1-Score: 99.86%

Extra Trees:

Accuracy: 99.10%

F1-Score: 98.86%

Gradient Boosting:

Accuracy: 99.97%

F1-Score: 99.97%

Adaboost:

Accuracy: 77.30%

F1-Score: 68.93%

Bagging:

Accuracy: 99.97%

F1-Score: 99.97%

Knn:

Accuracy: 97.02%

F1-Score: 96.84%

Naive Bayes:

Accuracy: 89.80%

F1-Score: 91.18%

Linear Svc:

Accuracy: 84.52%

F1-Score: 82.05%

Xgboost:

Accuracy: 99.97%

F1-Score: 99.97%

Task Summary

Classification Task Summary:

Target Variable: region

Dataset: biometric

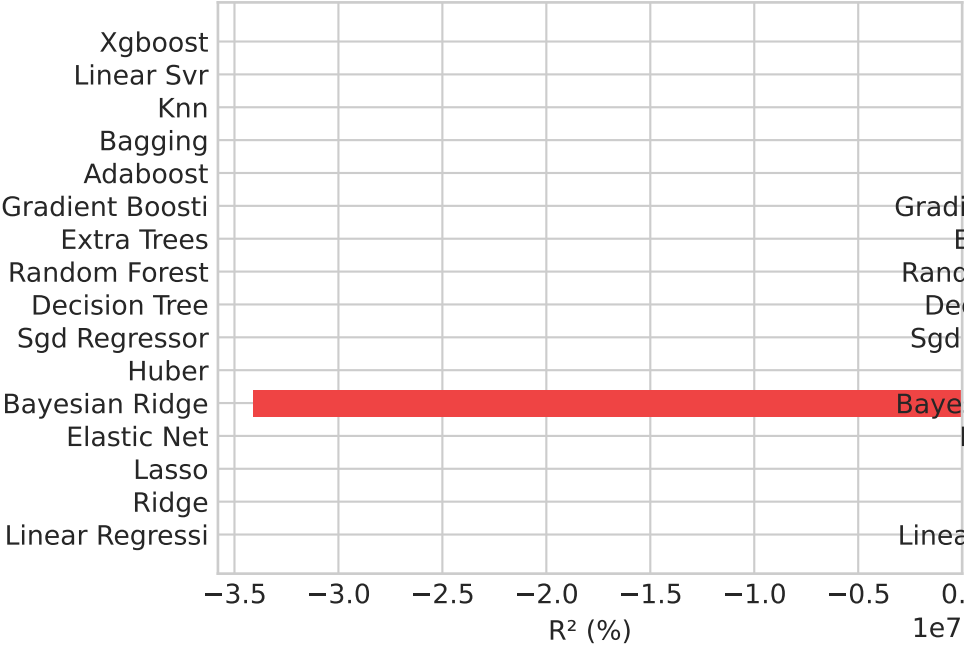
Records Used: 30,000

Best Model: Decision Tree

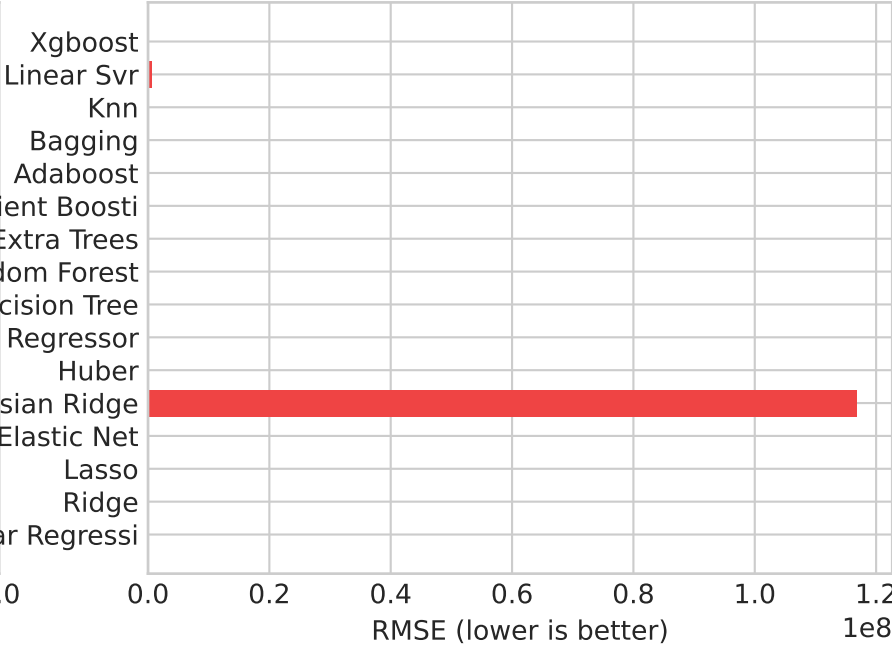
Best Accuracy: 99.97%

BIOMETRIC - Regression Results

Model R² Comparison



Model RMSE Comparison



Regression Metrics:

Linear Regression:

R²: 1.0000
RMSE: 0.00
MAE: 0.00

Ridge:

R²: 1.0000
RMSE: 430.47
MAE: 364.11

Lasso:

R²: 1.0000
RMSE: 135.90
MAE: 111.64

Elastic Net:

R²: 0.9772
RMSE: 30,178.28
MAE: 25,937.68

Bayesian Ridge:

R²: -340904.0106
RMSE: 116,817,472.59
MAE: 98,475,200.72

Huber:

R²: 1.0000
RMSE: 0.00
MAE: 0.00

Sgd Regressor:

R²: 1.0000
RMSE: 958.46
MAE: 810.14

Decision Tree:

R²: 1.0000
RMSE: 96.21
MAE: 55.87

Random Forest:

R²: 1.0000
RMSE: 124.89
MAE: 42.88

Extra Trees:

R²: 1.0000
RMSE: 197.00
MAE: 150.03

Gradient Boosting:

R²: 1.0000
RMSE: 163.94
MAE: 101.16

Adaboost:

R²: 0.9922
RMSE: 17,726.45
MAE: 14,670.21

Bagging:

R²: 1.0000
RMSE: 125.06
MAE: 5.98

Knn:

R²: 0.9964
RMSE: 12,007.81
MAE: 5,901.34

Linear Svr:

R²: -6.2293
RMSE: 537,947.76
MAE: 499,467.74

Xgboost:

R²: 1.0000
RMSE: 436.76
MAE: 306.54

Regression Interpretation:

Best R² achieved: 1.0000

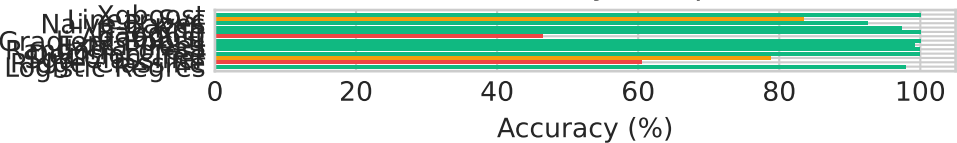
R² Interpretation:

- R² < 0.2: Weak predictive power
- R² 0.2-0.5: Moderate predictive power
- R² > 0.5: Good predictive power

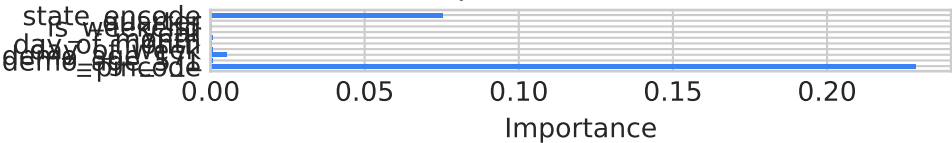
The regression task shows moderate predictability. Additional features may improve results.

DEMOGRAPHIC - Classification Results

Model Accuracy Comparison



Feature Importance (Random Forest)



Detailed Metrics

Model Performance Metrics:

Logistic Regression:

Accuracy: 97.92%

F1-Score: 97.35%

Ridge Classifier:

Accuracy: 60.50%

F1-Score: 52.20%

Sgd Classifier:

Accuracy: 78.77%

F1-Score: 78.07%

Decision Tree:

Accuracy: 99.95%

F1-Score: 99.95%

Random Forest:

Accuracy: 99.93%

F1-Score: 99.93%

Extra Trees:

Accuracy: 99.20%

F1-Score: 99.03%

Gradient Boosting:

Accuracy: 100.00%

F1-Score: 100.00%

Adaboost:

Accuracy: 46.40%

F1-Score: 35.60%

Bagging:

Accuracy: 99.97%

F1-Score: 99.97%

Knn:

Accuracy: 97.40%

F1-Score: 97.27%

Naive Bayes:

Accuracy: 92.53%

F1-Score: 92.54%

Linear Svc:

Accuracy: 83.45%

F1-Score: 81.47%

Xgboost:

Accuracy: 100.00%

F1-Score: 100.00%

Task Summary

Classification Task Summary:

Target Variable: region

Dataset: demographic

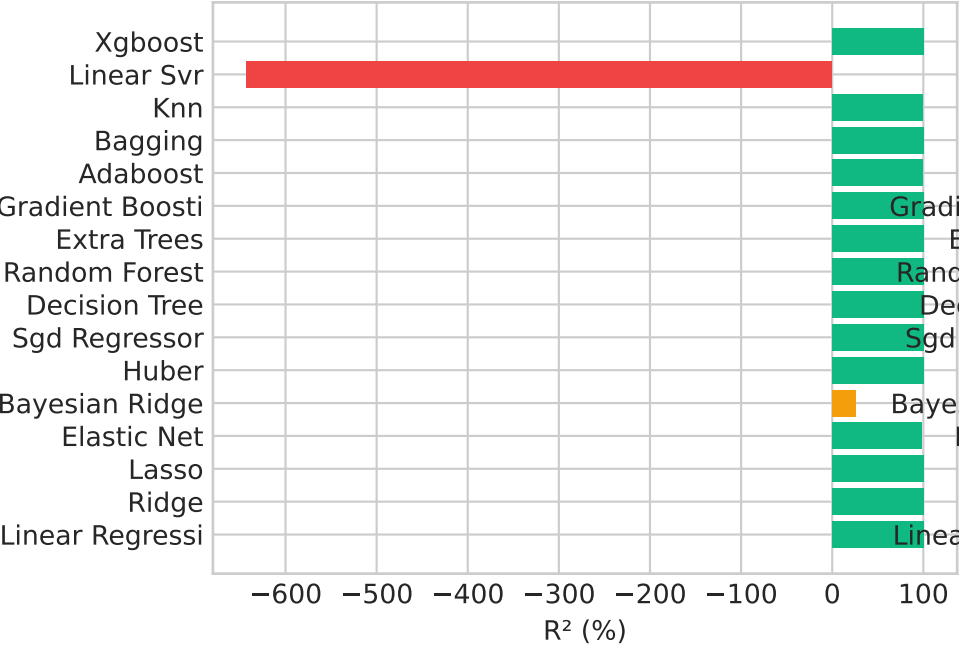
Records Used: 30,000

Best Model: Gradient Boosting

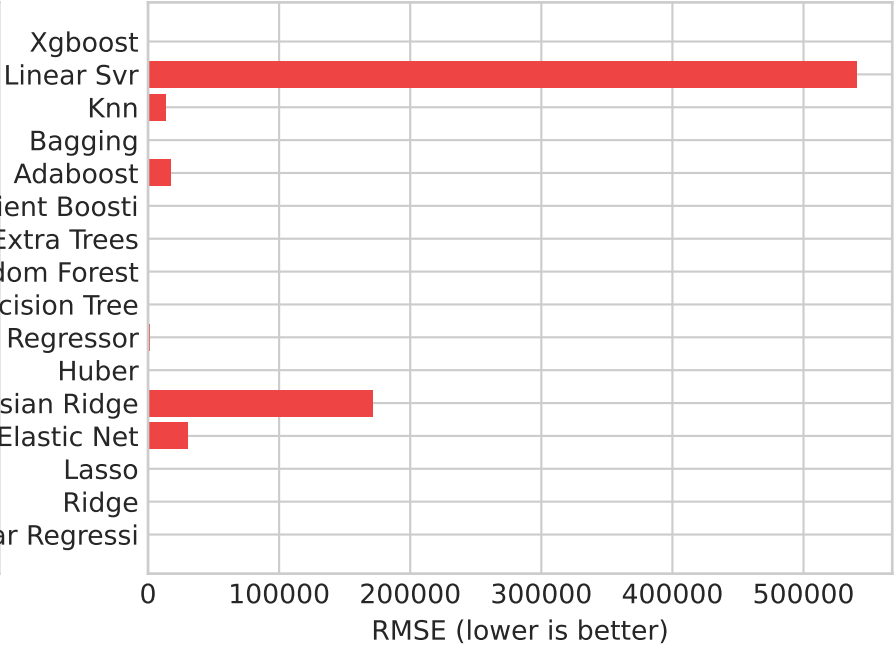
Best Accuracy: 100.00%

DEMOGRAPHIC - Regression Results

Model R² Comparison



Model RMSE Comparison



Regression Metrics:

Linear Regression:

R²: 1.0000
RMSE: 0.00
MAE: 0.00

Ridge:

R²: 1.0000
RMSE: 427.28
MAE: 360.35

Lasso:

R²: 1.0000
RMSE: 62.09
MAE: 49.71

Elastic Net:

R²: 0.9769
RMSE: 30,131.35
MAE: 25,868.35

Bayesian Ridge:

R²: 0.2535
RMSE: 171,432.25
MAE: 132,908.89

Huber:

R²: 1.0000
RMSE: 0.00
MAE: 0.00

Sgd Regressor:

R²: 1.0000
RMSE: 959.74
MAE: 808.30

Decision Tree:

R²: 1.0000
RMSE: 91.33
MAE: 53.30

Random Forest:

R²: 1.0000
RMSE: 69.26
MAE: 42.22

Extra Trees:

R²: 1.0000
RMSE: 187.42
MAE: 146.17

Gradient Boosting:

R²: 1.0000
RMSE: 159.09
MAE: 97.08

Adaboost:

R²: 0.9927
RMSE: 16,952.60
MAE: 13,940.62

Bagging:

R²: 1.0000
RMSE: 19.29
MAE: 4.91

Knn:

R²: 0.9955
RMSE: 13,260.38
MAE: 6,122.70

Linear Svr:

R²: -6.4251
RMSE: 540,651.92
MAE: 502,980.21

Xgboost:

R²: 1.0000
RMSE: 454.64
MAE: 316.11

Regression Interpretation:

Best R² achieved: 1.0000

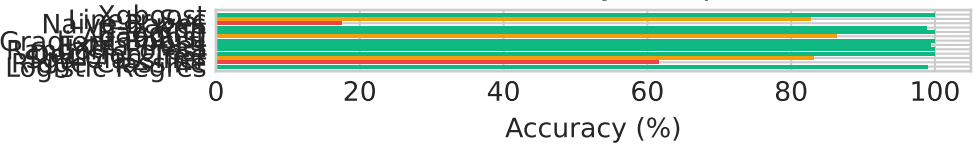
R² Interpretation:

- R² < 0.2: Weak predictive power
- R² 0.2-0.5: Moderate predictive power
- R² > 0.5: Good predictive power

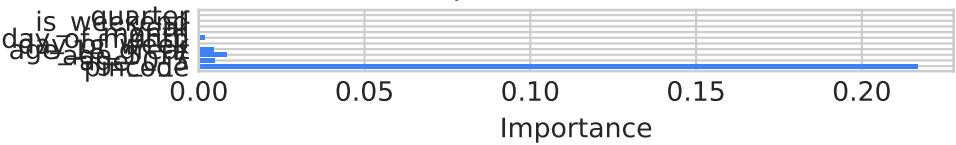
The regression task shows moderate predictability. Additional features may improve results.

ENROLLMENT - Classification Results

Model Accuracy Comparison



Feature Importance (Random Forest)



Detailed Metrics

Model Performance Metrics:

Logistic Regression:

Accuracy: 98.94%

F1-Score: 98.52%

Ridge Classifier:

Accuracy: 61.62%

F1-Score: 55.54%

Sgd Classifier:

Accuracy: 83.08%

F1-Score: 82.83%

Decision Tree:

Accuracy: 100.00%

F1-Score: 100.00%

Random Forest:

Accuracy: 99.96%

F1-Score: 99.96%

Extra Trees:

Accuracy: 99.40%

F1-Score: 99.24%

Gradient Boosting:

Accuracy: 100.00%

F1-Score: 100.00%

Adaboost:

Accuracy: 86.38%

F1-Score: 82.59%

Bagging:

Accuracy: 100.00%

F1-Score: 100.00%

Knn:

Accuracy: 98.83%

F1-Score: 98.82%

Naive Bayes:

Accuracy: 17.40%

F1-Score: 23.13%

Linear Svc:

Accuracy: 82.67%

F1-Score: 81.02%

Xgboost:

Accuracy: 100.00%

F1-Score: 100.00%

Task Summary

Classification Task Summary:

Target Variable: region

Dataset: enrollment

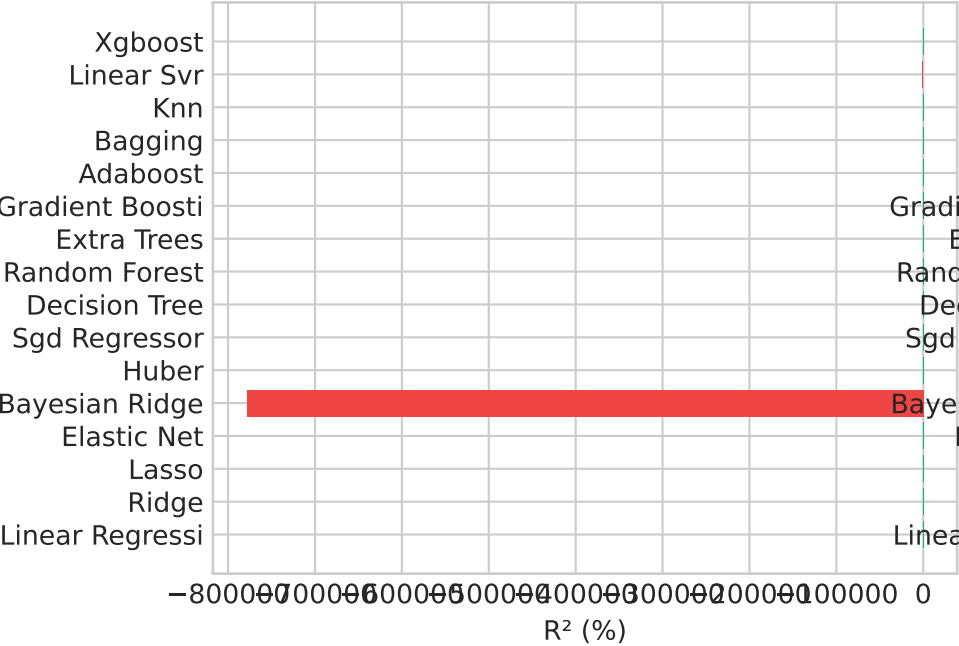
Records Used: 26,029

Best Model: Decision Tree

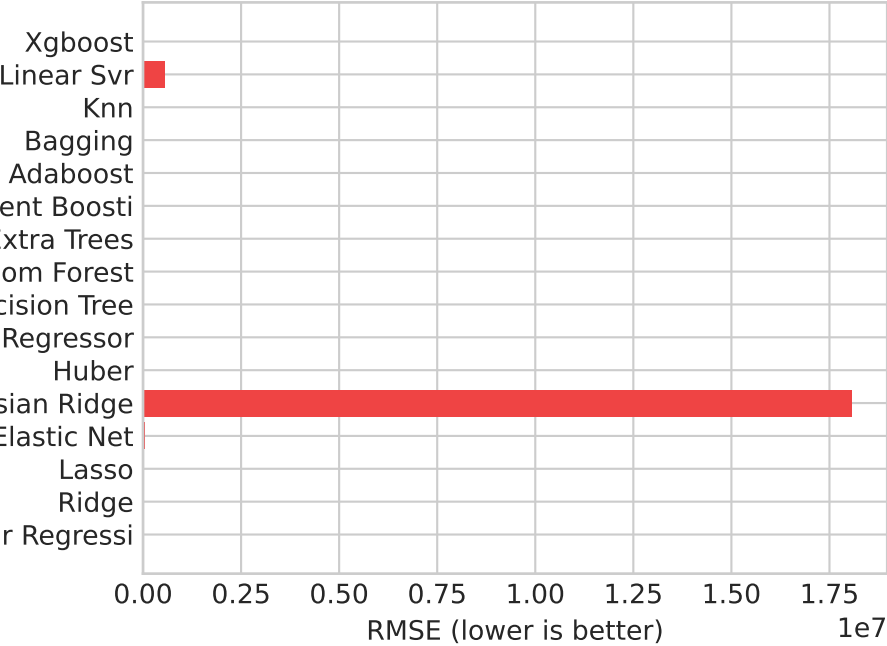
Best Accuracy: 100.00%

ENROLLMENT - Regression Results

Model R² Comparison



Model RMSE Comparison



Regression Metrics:

Linear Regression:

R²: 1.0000
RMSE: 0.00
MAE: 0.00

Ridge:

R²: 1.0000
RMSE: 495.76
MAE: 419.88

Lasso:

R²: 1.0000
RMSE: 142.72
MAE: 122.01

Elastic Net:

R²: 0.9775
RMSE: 30,712.04
MAE: 26,643.25

Bayesian Ridge:

R²: -7784.9929
RMSE: 18,069,411.58
MAE: 13,816,958.89

Huber:

R²: 1.0000
RMSE: 0.00
MAE: 0.00

Sgd Regressor:

R²: 1.0000
RMSE: 1,004.35
MAE: 850.55

Decision Tree:

R²: 1.0000
RMSE: 93.83
MAE: 55.78

Random Forest:

R²: 1.0000
RMSE: 70.21
MAE: 42.54

Extra Trees:

R²: 1.0000
RMSE: 210.86
MAE: 163.96

Gradient Boosting:

R²: 1.0000
RMSE: 158.83
MAE: 95.94

Adaboost:

R²: 0.9946
RMSE: 15,035.21
MAE: 11,623.17

Bagging:

R²: 1.0000
RMSE: 15.66
MAE: 3.91

Knn:

R²: 0.9960
RMSE: 12,871.14
MAE: 2,465.18

Linear Svr:

R²: -6.4485
RMSE: 558,884.59
MAE: 519,811.91

Xgboost:

R²: 1.0000
RMSE: 430.75
MAE: 310.52

Regression Interpretation:

Best R² achieved: 1.0000

R² Interpretation:

- R² < 0.2: Weak predictive power
- R² 0.2-0.5: Moderate predictive power
- R² > 0.5: Good predictive power

The regression task shows moderate predictability. Additional features may improve results.