Machine Learning Results Summary

Sainath Raja

June 9, 2025

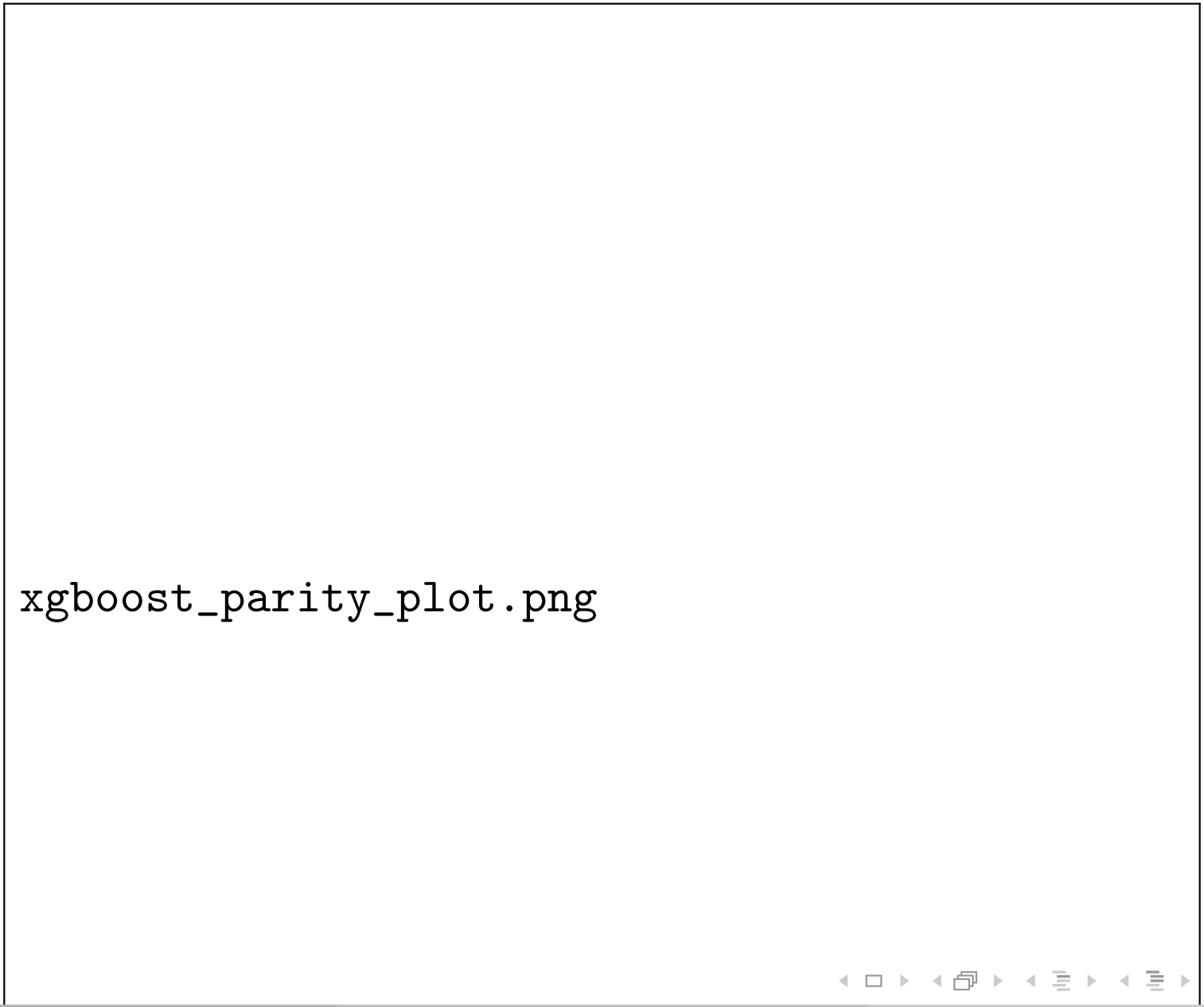
XGBoost Regressor (Best Model)

Hyperparameter Tuning Summary

Fitting 10 folds for each of 36 candidates, totalling 360 fits

- **Best Parameters:**
  - learning_rate = 0.01
  - max_depth = 5
  - n_estimators = 100
  - subsample = 0.8
- **Train R^2:** 0.6351
- **Test R^2:** -3.6986
- **Train RMSE:** 6.6189
- **Test RMSE:** 11.3658
- **Train Relative RMSE:** 0.4209
- **Test Relative RMSE:** 0.8666

Parity Plot - XGBoost



Features Used in CIN Prevalence

- Low CIN-16-prevalence
- High CIN-16-prevalence
- Low CIN-18-prevalence
- High CIN-18-prevalence

Combined Screening Coverage Statistics

Metric	Value
Count	186
Mean	41.0502
Std Dev	31.2920
Min	0.0000
25%	10.0417
50% (Median)	37.8333
75%	74.9167
Max	89.1667

Random Forest after DBSCAN Filtering

- **Train R^2:** 0.8212
- **Test R^2:** -0.7274
- **Train MSE:** 11.5526
- **Test MSE:** 47.4938

Parity Plot - Random Forest

rf_parity_plot.png