


Peer-To-Peer Lending

G R E A T Y I E L D S

Prepared By: Shelly Levy, Tom Saacks and Or Liberman

date

13 June 2023



Step Guidelines:

Modeling and Evaluation

- **Recap**
 - Classification model
 - Regression model
 - Selected Models
-

A snapshot of the data

- Data Preparation Completed (flexibility for future changes)
- 326,403 Instances
- 55 Features

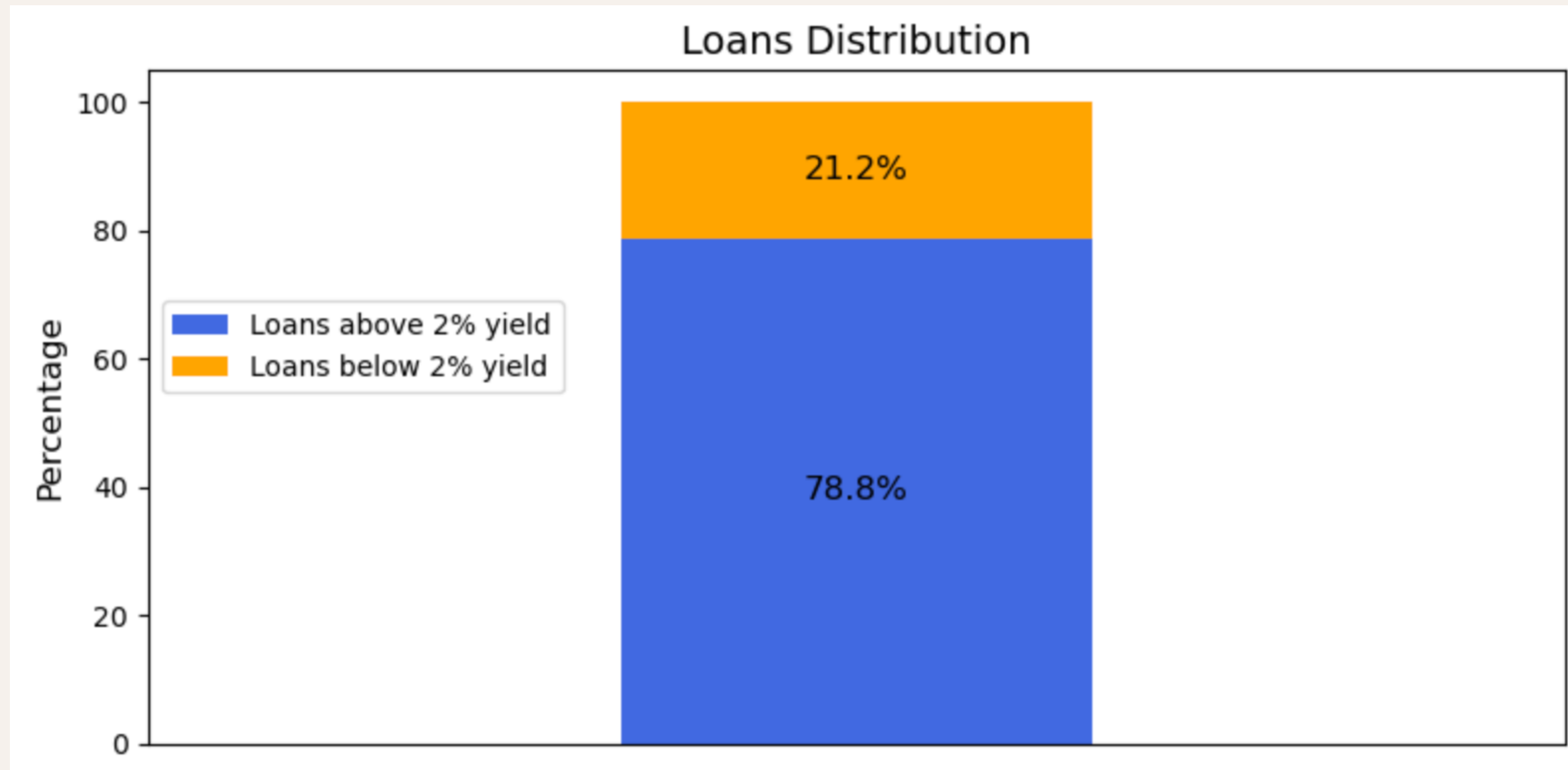


Model Approach

- **Classification model:** Classifying 2% return loans
 - Logistic Regression, XGBoost, AdaBoost with decision tree
- **Regression model:** Numerical prediction of yield
 - Linear Regression, Polynomial Regression
- Select best model for implementation in 2 weeks

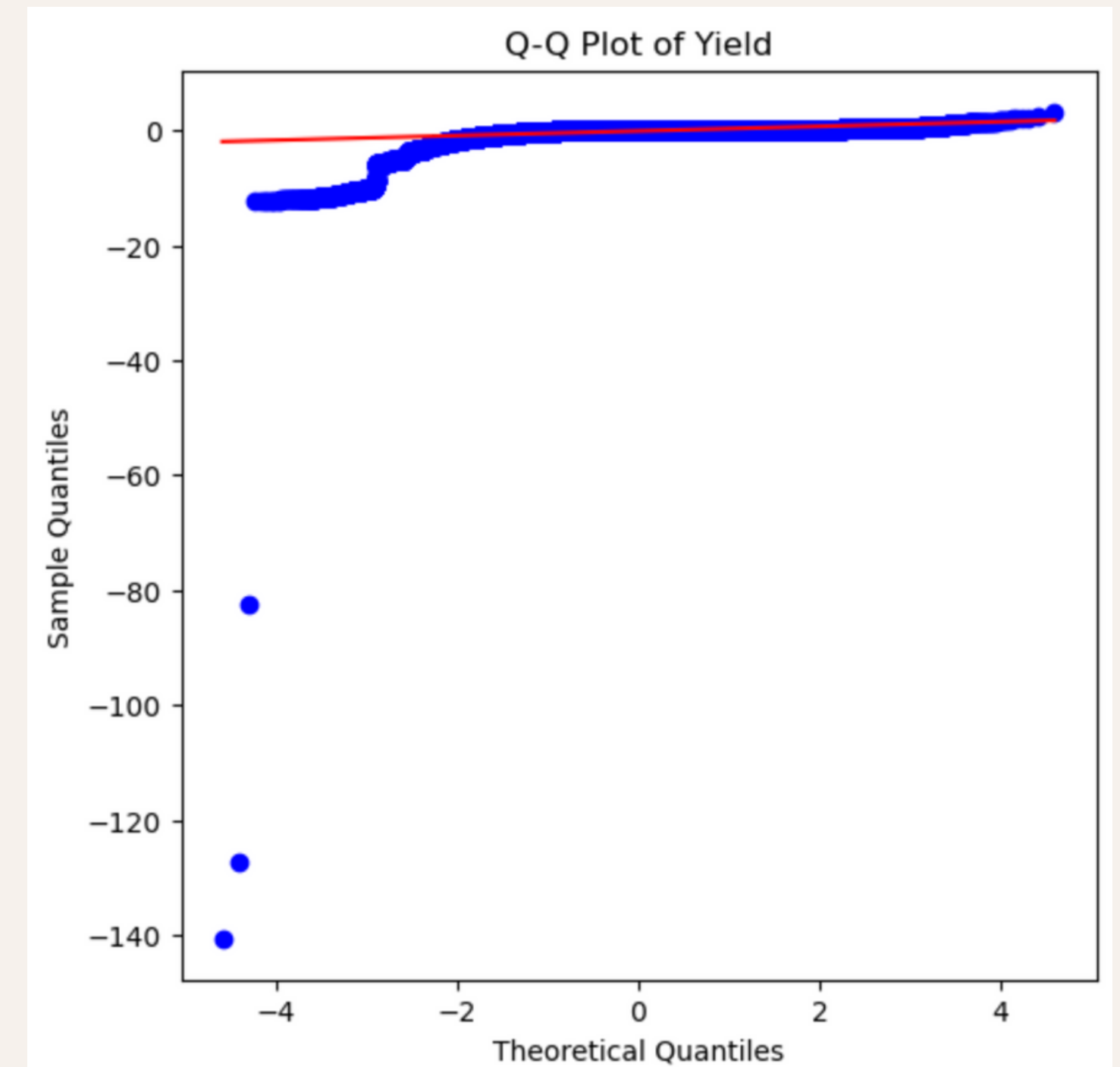
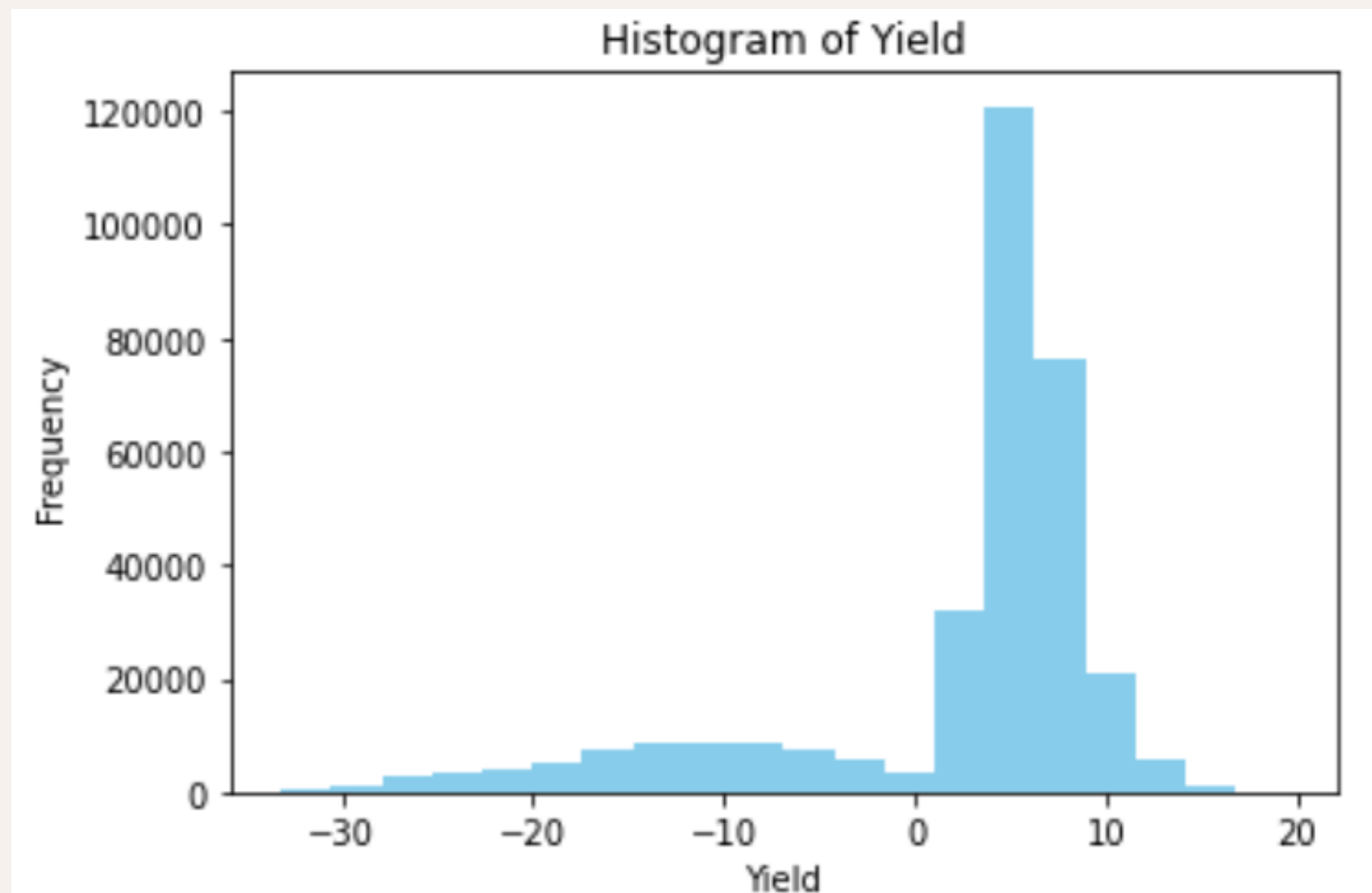
➔ Skewed Target Data Distribution

- **Classification model:**




➔ Skewed Target Data Distribution

- **Regression model:**
 - Skewness: Highly skewed (-40.14)
 - Shapiro-Wilk: Non-normal ($p \sim 0.0$)



Data Normalization

- Min-Max Normalization
 - Outliers Challenge
- Standard Normalization
 - Non-normal Features Challenge
- Robust Normalization

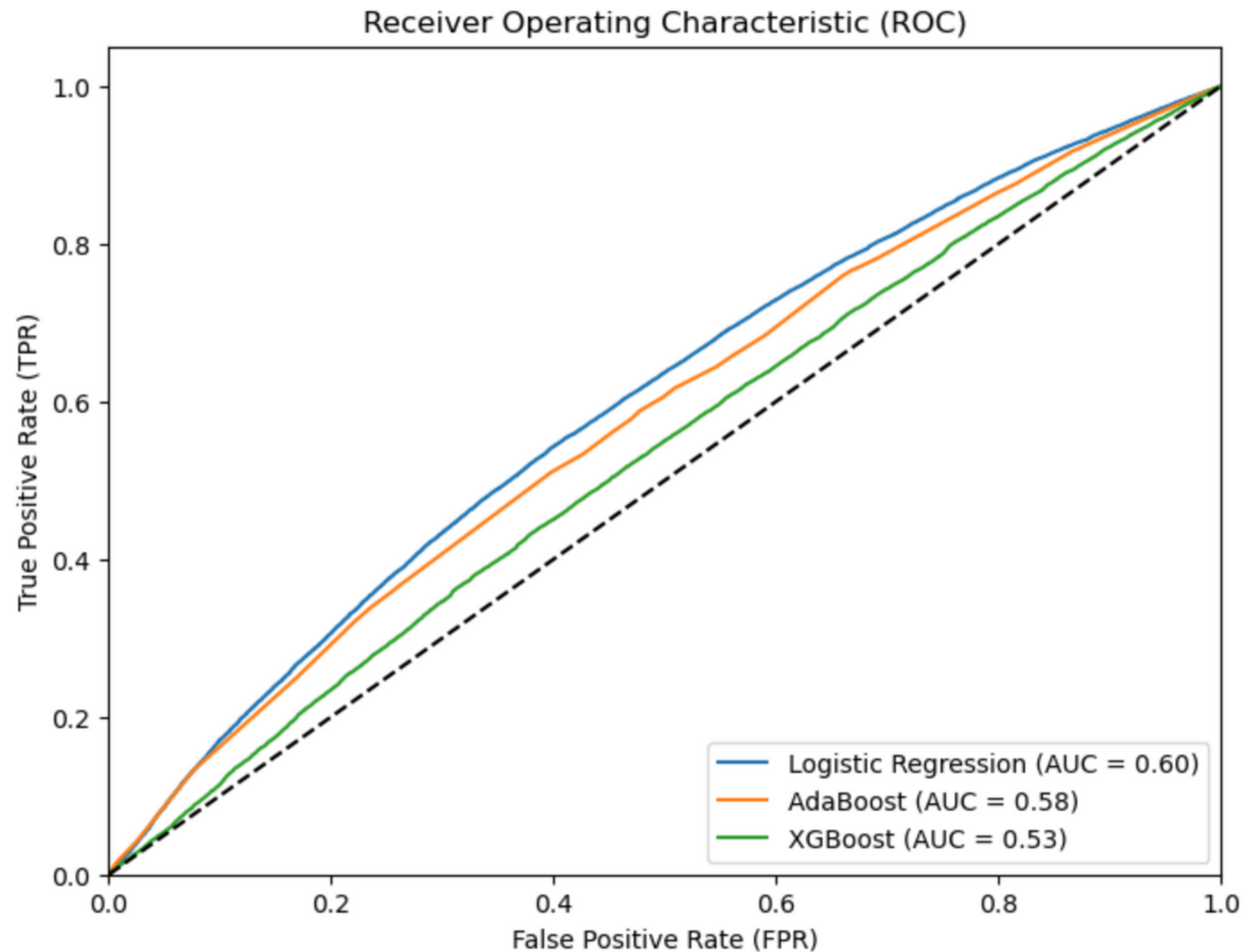


Step Guidelines:

Modeling and Evaluation

- Recap
 - **Classification model**
 - Regression model
 - Selected Models
-

Classification model:



	LG	AdaBoost	XGBoost
F1	0.722	0.604	0.586
Recall	0.714	0.604	0.581
Precision	0.710	0.6045	0.581
Accuracy	0.741	0.604	0.586

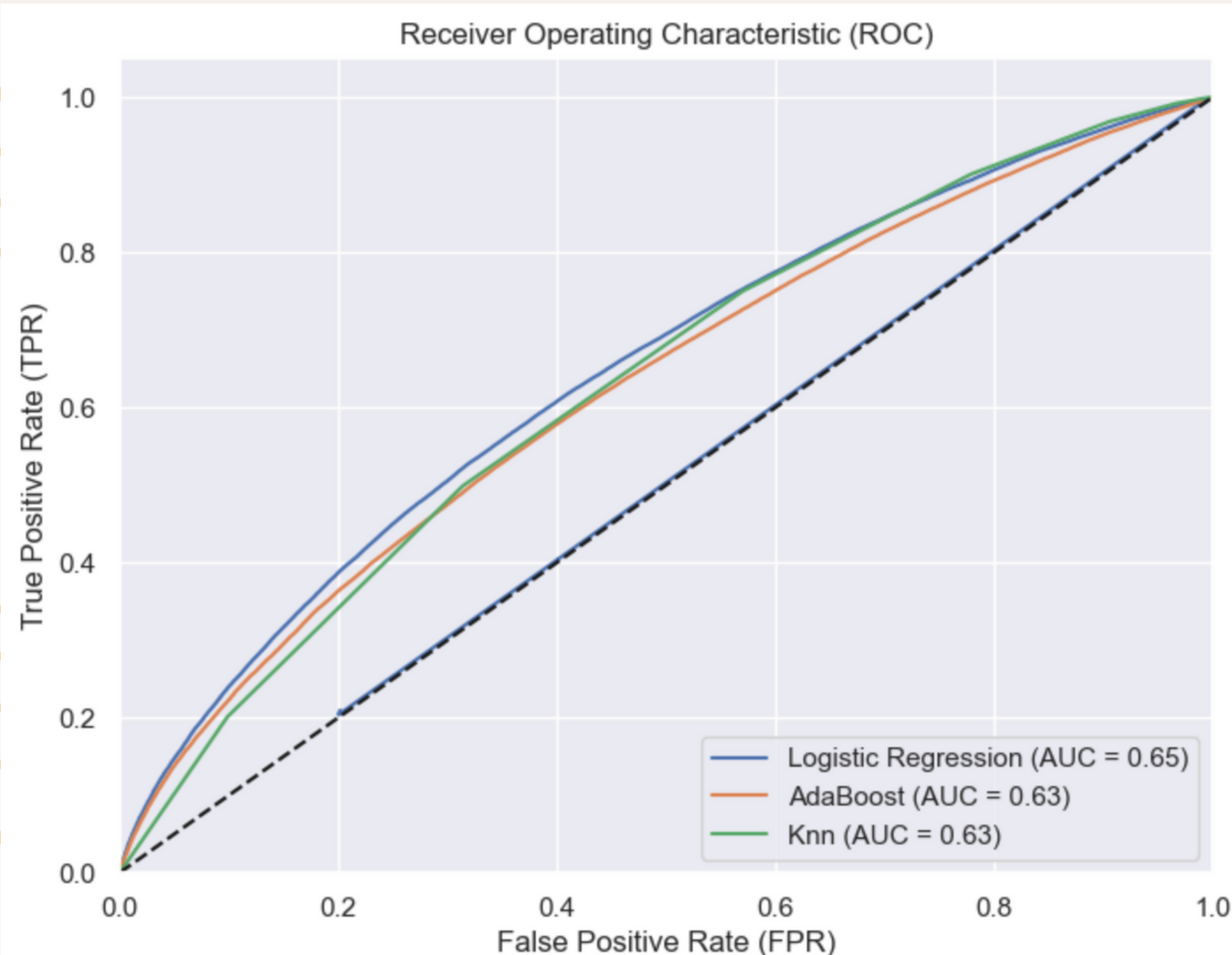


Classification model:


General Steps:

- Hyperparameter Tuning
- Features selection methods
- Skewness Handling
- AdaBoost with Logistic Regression
- Knn model

Classification model:



	LG	AdaBoost with LG	Knn
F1	0.725021	0.695193	0.724328
Recall	0.739447	0.788454	0.784289
Precision	0.727746	0.621663	0.725432
Accuracy	0.791447	0.788454	0.784289



Step Guidelines:

Modeling and Evaluation

- Recap
 - Classification model
 - **Regression model**
 - Selected Models
-



Regression model:

Polynomial Regression

	Model			MSE	R ²
0	PR	Degree-2	(70:30)	0.007032	0.030277
1	PR	Degree-3	(70:30)	0.007140	0.015245
2	PR	Degree-2	(80:20)	0.007042	0.030021
3	PR	Degree-3	(80:20)	0.007182	0.011088



Regression model:



General Steps:

- Hyperparameter Tuning
- Feature Selection
- Model training
- Model Evaluation
- Model Improvement
- Evaluation Metric



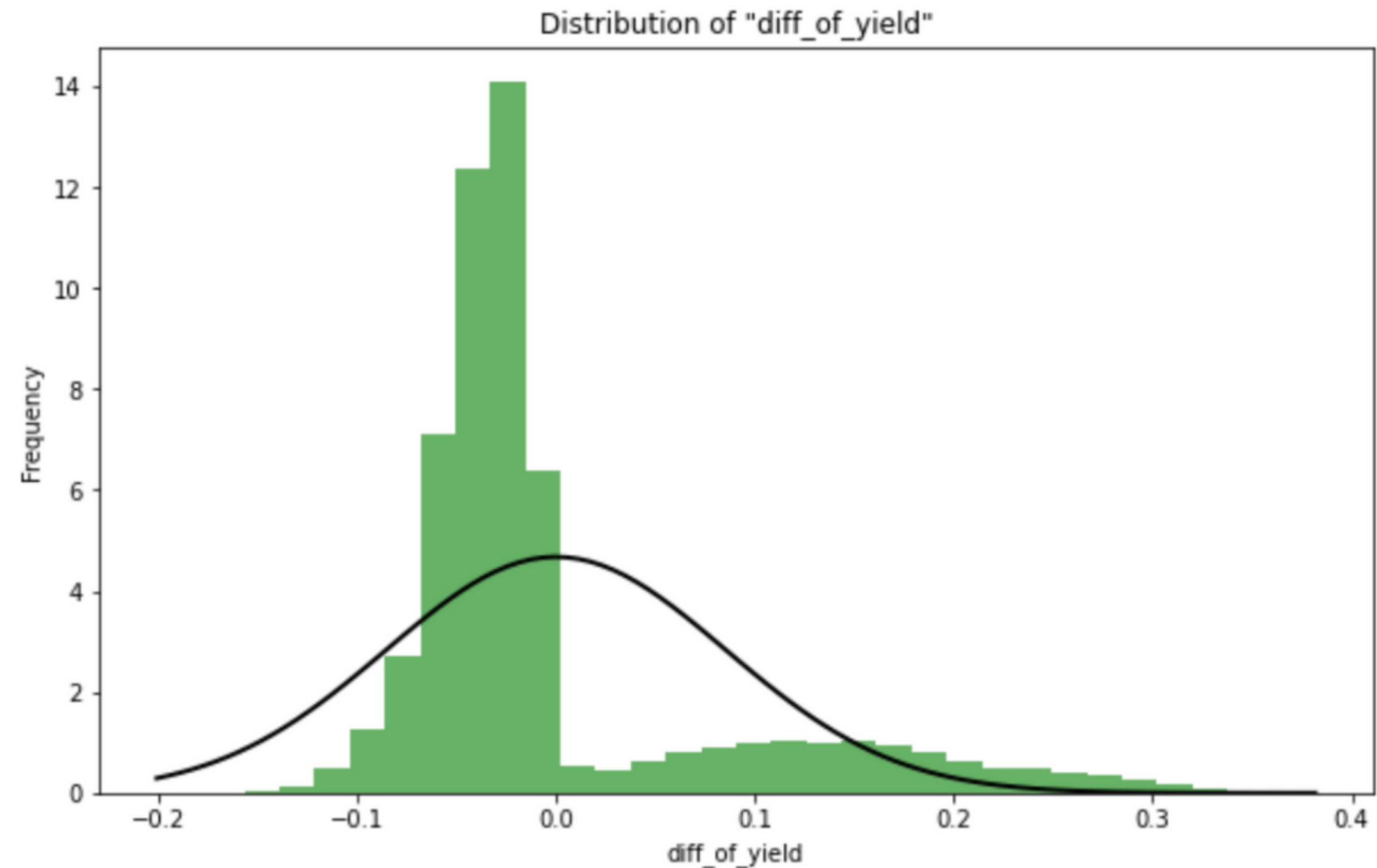
Regression model:

Polynomial Regression

Model_Name	MSE	R2_Score	Regularization	Alpha
Polynomial Regression (degree=2) - all featuers	0.00706	0.027781	Ridge	10
Polynomial Regression (degree=2) - all featuers	0.00702	0.030425	Lasso	0.01
Polynomial Regression (degree=2) - best featuers	0.00733	-0.00882	Ridge	10
Polynomial Regression (degree=2) - best featuers	0.00726	-0.00029	Lasso	10

Regression model: Polynomial Regression

Mean of 'diff': $-1.8535593212225408 \times 10^{-6}$
Standard Deviation of 'diff': 0.08528777623813089

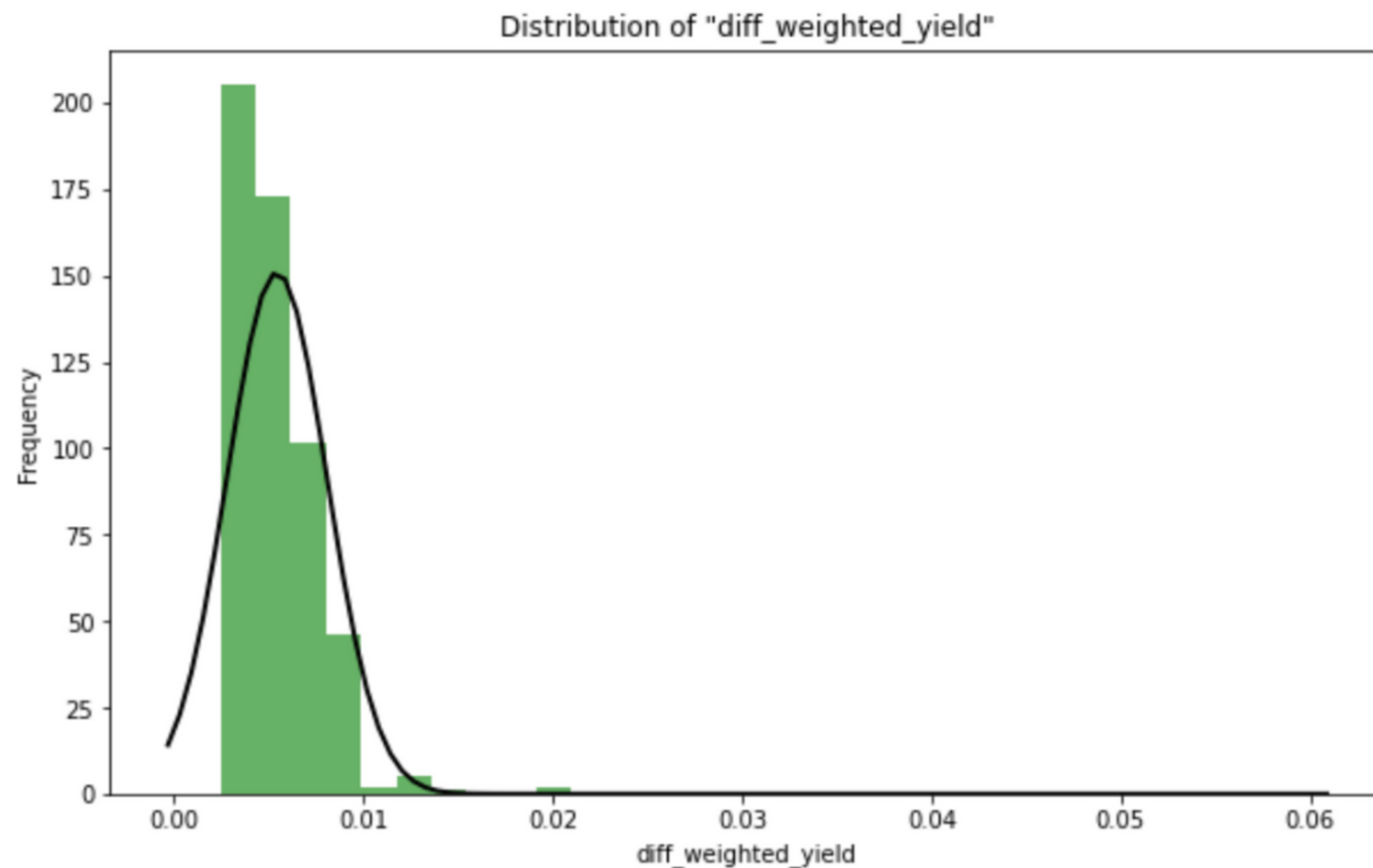




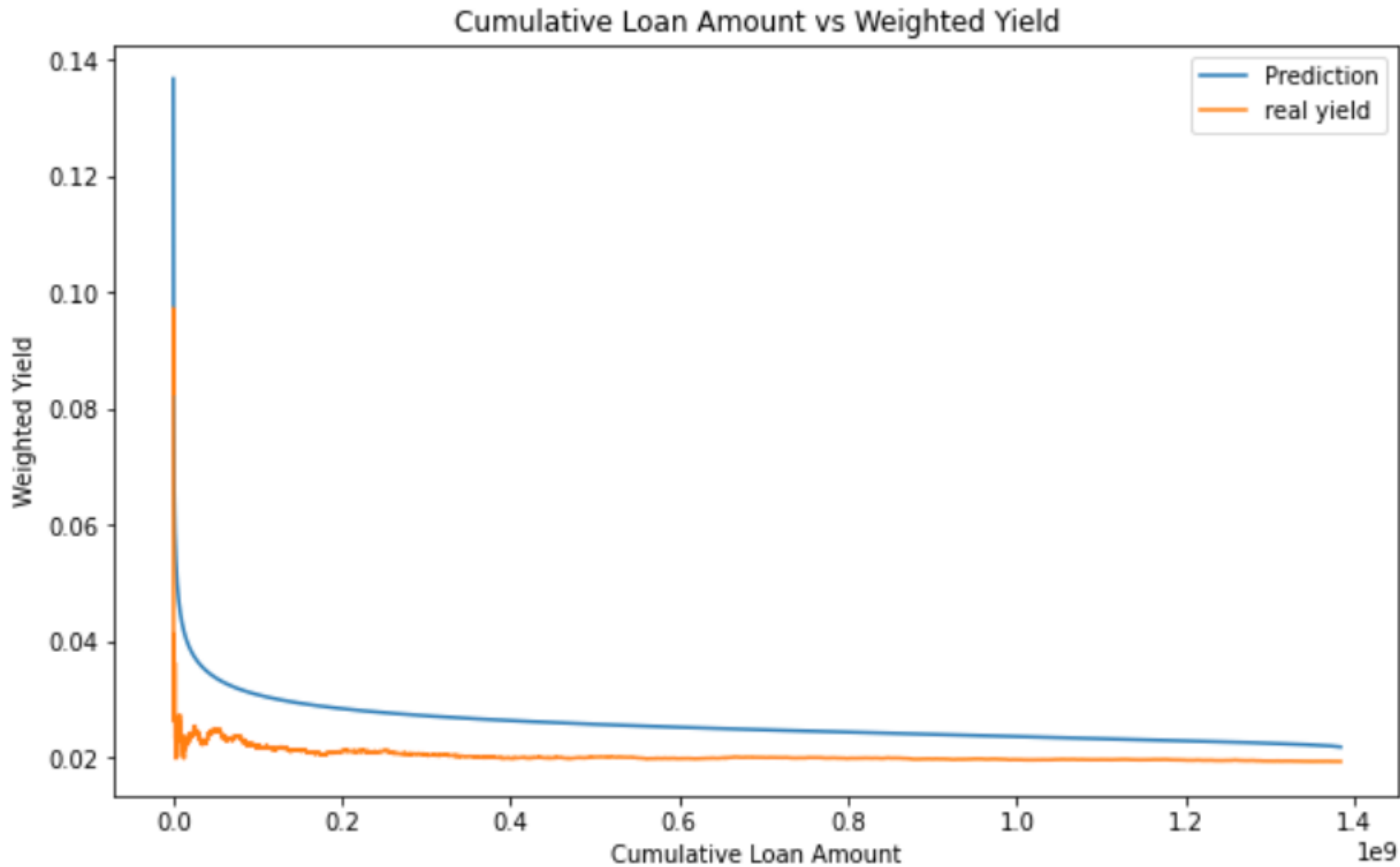
Regression model: Polynomial Regression


Mean of 'diff_weighted': 0.0054587335295935

Standard Deviation of 'diff_weighted': 0.0026440640031357757



Regression model: Polynomial Regression





Step Guidelines:

Modeling and Evaluation

- Recap
 - Classification model
 - Regression model
 - **Selected Models**
-

Selected Models:

- Model Selection: Logistic Regression (Classification)
Polynomial Regression (Regression)
- Two-step Modeling Approach
- Deeper Insights, Accurate Predictions
- **Business-Driven Determination:** Opting for Single or Dual Model Usage

Potential Pitfalls:

- Failure to balance skewed distribution while maintaining interpretability
 - Ignoring market dynamics and external factors.
 - Failure to address potential overfitting/underfitting issues
 - Computational power limitations
 - Model Assumptions
-



Next steps

- Business Perspective
- Financial Potential Analysis
- Existing Grade Model Comparison
- Addressing CFO Queries
- Best Model Selection



Concluding Remarks and Q&A

THANK
YOU