DataDriven
PORTFOLIO SOLUTIONS

Peer-To-Peer Lending

GREATYIELDS

Prepared By: Shelly Levy, Tom Saacks and Or Liberman

date

30 May 2023

A snapshot of the data

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

DataDriven
PORTFOLIO SOLUTIONS

Step Guidelines:

Model Selection and Setup

- Model approach
- Model Preprocessing
- Models Selection and Performance

Step Guidelines:

Model Selection and Setup

- **Model approach**
- Model Preprocessing
- Models Selection and Performance

Model Approach

- **Classification model:** Classifying 2% return loans
  - Logistic Regression, XGBoost, AdaBoost
- **Regression model:** Numerical prediction of yield
  - Linear Regression, Polynomial Regression
- Select best model for implementation in 4 weeks

DataDriven
PORTFOLIO SOLUTIONS

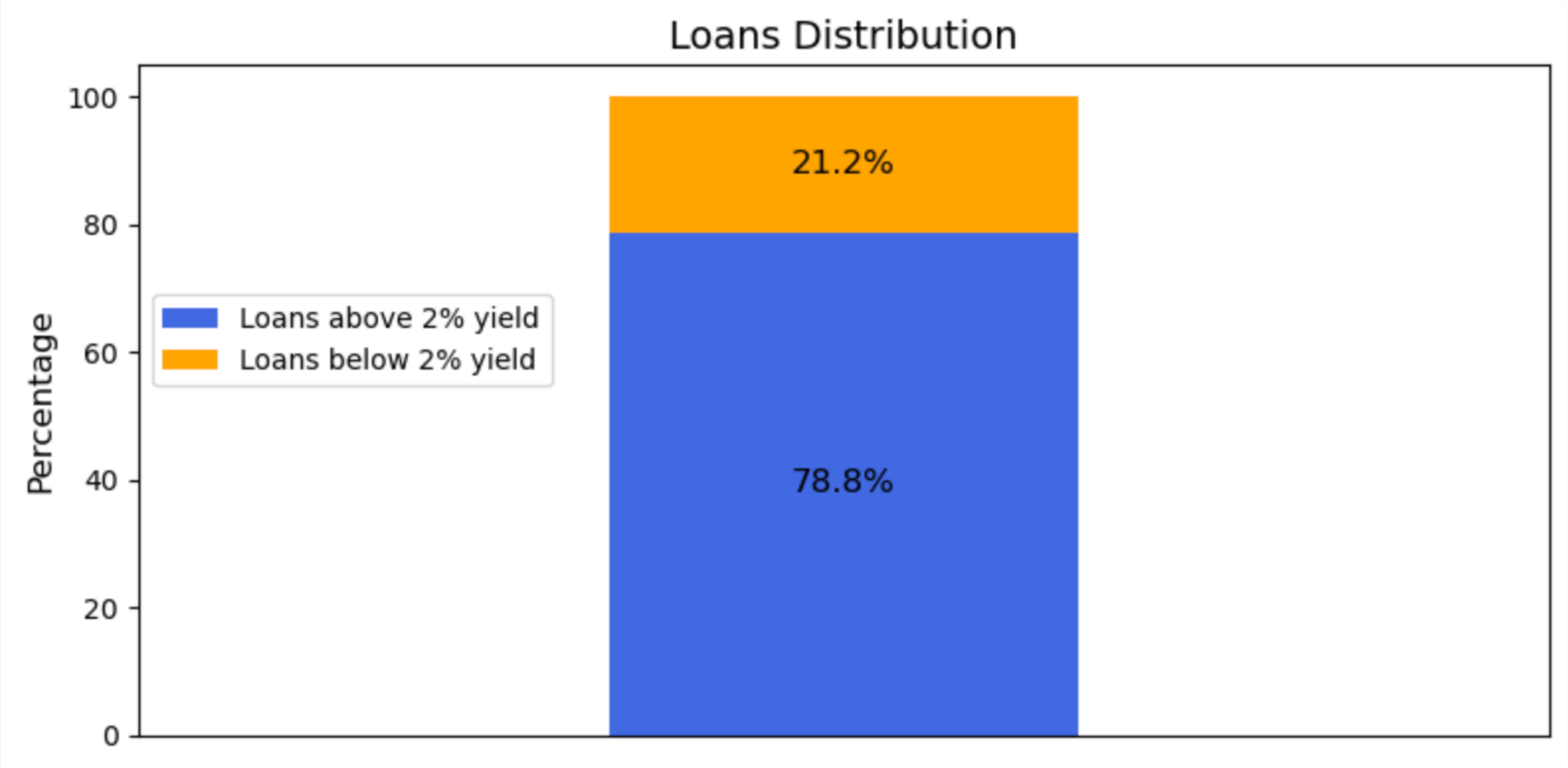
Step Guidelines:

Model Selection and Setup

- Model approach
- **Model Preprocessing**
  - **Skewed Target Data Distribution**
  - Data Normalization
- Models Selection and Performance

Skewed Target Data Distribution

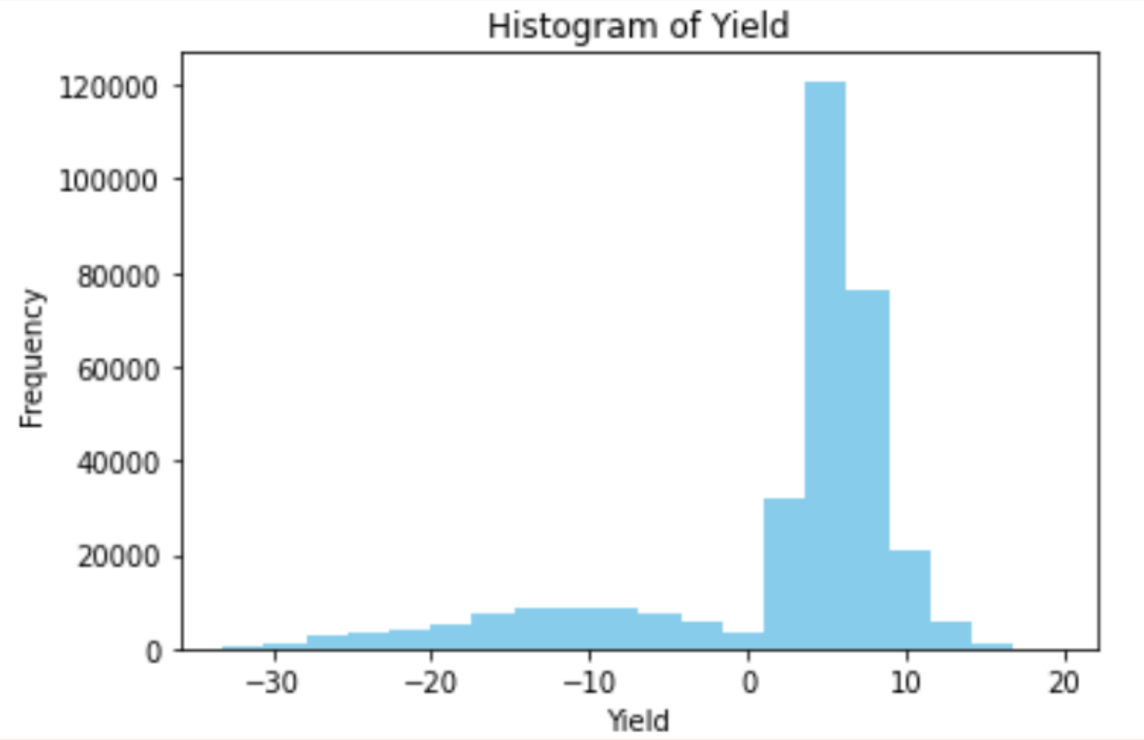
- **Classification model:**

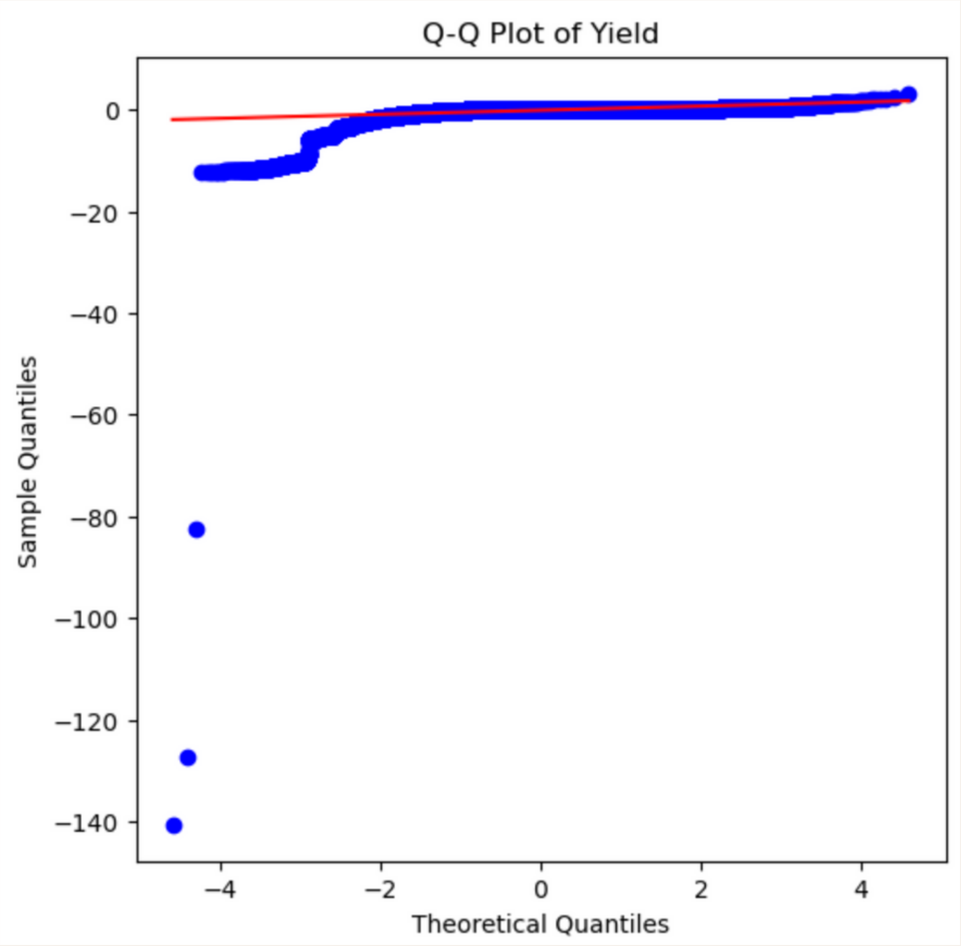


DataDriven
PORTFOLIO SOLUTIONS

Skewed Target Data Distribution

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





Step Guidelines:

Model Selection and Setup

- Model approach
- **Model Preprocessing**
  - Skewed Target Data Distribution
  - **Data Normalization**
- Models Selection and Performance

Data Normalization

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

DataDriven
PORTFOLIO SOLUTIONS

Step Guidelines:

Model Selection and Setup

- Model approach
- Model Preprocessing
- **Models Selection and Performance**
  - **Classification model**
  - Regression model

Classification model:

Applied Models:

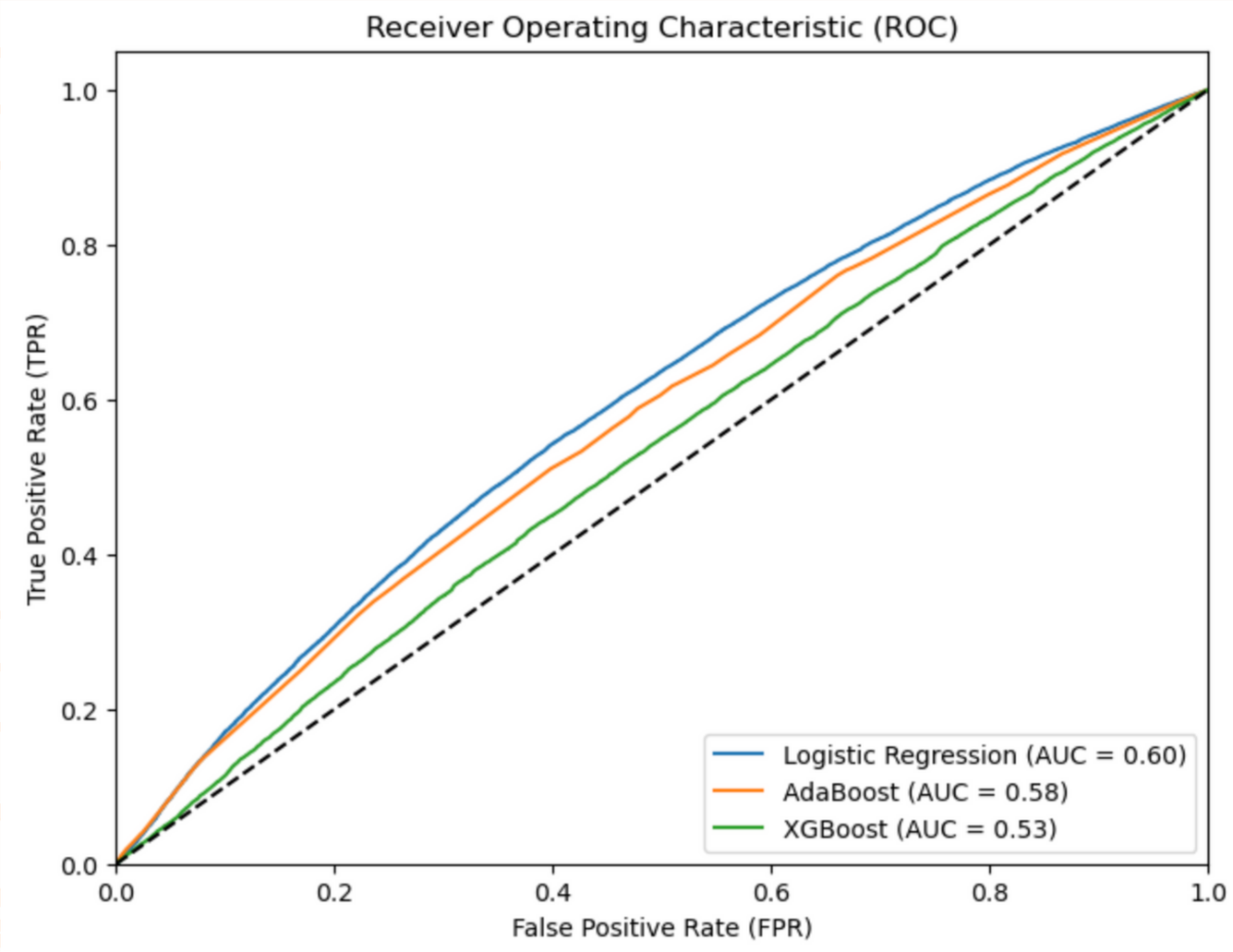
- Logistic Regression
- XGBoost
- AdaBoost

Classification model:

General Steps:

- Feature Selection
- Class Weighting
- Hyperparameter Tuning: Exploring Diverse Combinations
- Evaluation Metric

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:

- The model results are not yet optimized
- Ongoing hyperparameter tuning to enhance models.

Step Guidelines:

Model Selection and Setup

- Model approach
- Model Preprocessing
- **Models Selection and Performance**
  - Classification model
  - **Regression model**

Regression model:

Applied Models:

- Linear Regression
- Polynominal Regression

Regression model:

General Steps:

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

Regression model:

Linear Regression:

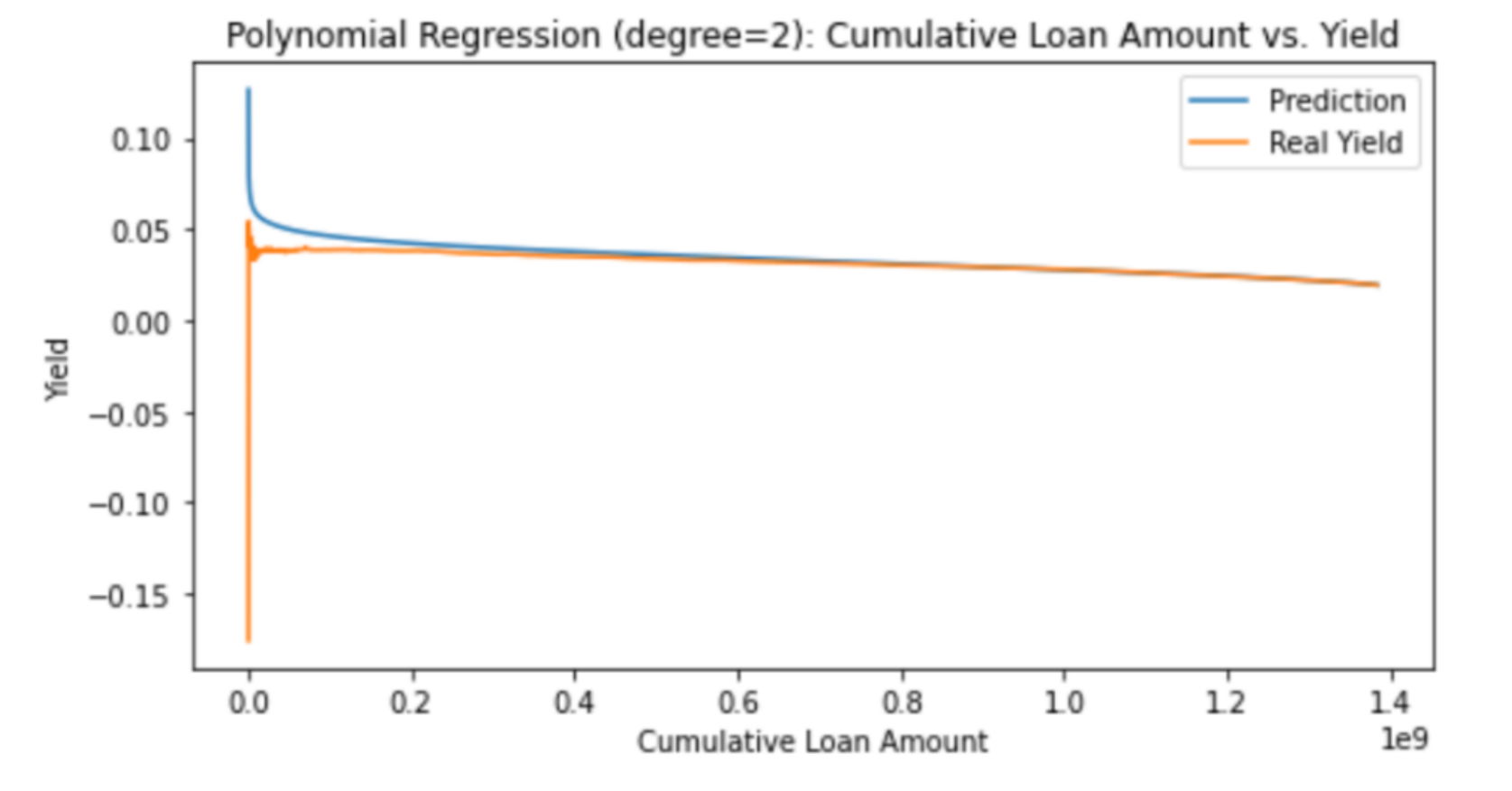
	Normalization	Split	MSE	R^2 Score
0	RobustScaler	80:20	0.007101	0.023337
1	RobustScaler	70:30	0.007098	0.023385
2	min-max	70:30	0.007203	0.212900
3	min-max	80:20	0.007303	0.202900
4	standard	70:30	0.007156	0.232400
5	standard	80:20	0.007178	0.231000

Polynominal 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:

Polynominal Regression :



Potential Pitfalls:

- Failure to balance skewed distribution while maintaining interpretability
- Ignoring market dynamics and external factors.
- Failure to address potential overfitting issues

DataDriven
PORTFOLIO SOLUTIONS

Next steps

- Hyperparameter Investigation
- Model Evaluations and Comparisons
- Addressing Skewed Distribution
- Financial Potential Analysis
- Model Selection

DataDriven
PORTFOLIO SOLUTIONS

Concluding Remarks and Q&A

THANK **YOU**

DataDriven
PORTFOLIO SOLUTIONS