AUTOMATED MACHINE LEARNING ANALYSIS REPORT

Generated: February 16, 2026 at 00:11

Report Type: Full ML Pipeline Analysis

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

This report presents a comprehensive analysis of an automated machine learning pipeline executed on February 16, 2026 at 00:11. The analysis encompassed data exploration, preprocessing, feature engineering, model selection, and performance evaluation.

Key Findings

Metric	Value
Best Model	RandomForestRegressor
Model Score	0.9489
Models Evaluated	3

2. DATA EXPLORATORY ANALYSIS

2.1 Dataset Overview

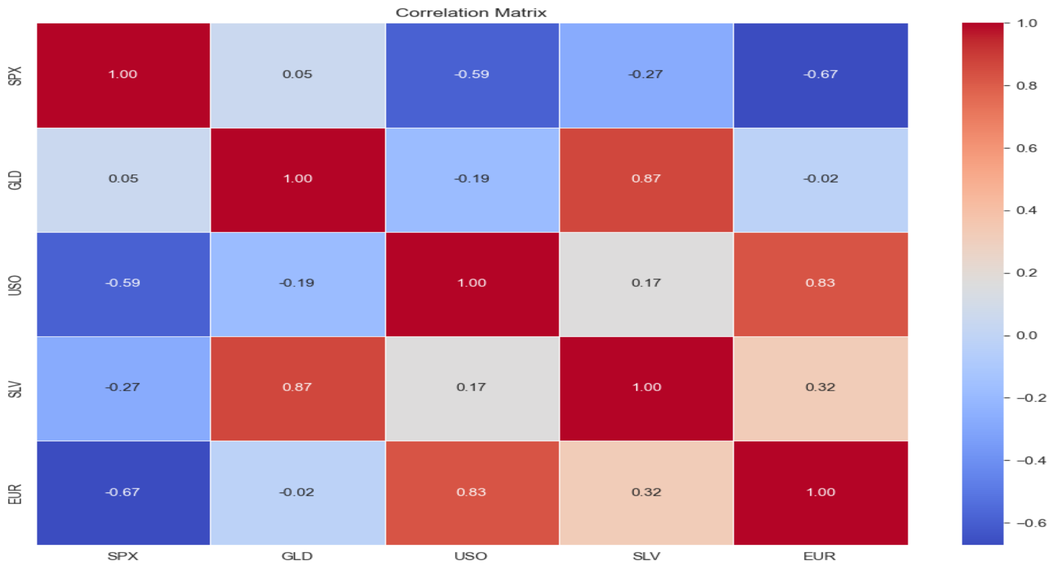
EXECUTIVE SUMMARY Our dataset consists of five financial indices: SPX, GLD, USO, SLV, and EUR, with 2290 observations each. The data spans a specific period and provides insights into the performance of these indices. KEY DATA INSIGHTS - The SPX index has a mean value of $1654.32, indicating a moderate level of growth. The minimum value recorded is $676.53, while the maximum value reached is $2872.87. The spread from the mean suggests that some fluctuations occurred during this period. - The GLD index shows a significantly lower mean value of $122.73, with the minimum and maximum values ranging between $70 and $184.59. This indicates a relatively stable performance for Gold prices during the analyzed time frame. - USO has a low mean value of $31.84, with the lowest recorded value being $7.96 and the highest being $117.48. The spread from the mean is moderate, indicating that Silver prices experienced some volatility. - SLV displays a lower mean value of $20.08, along with a range between $8.85 and $47.26. This suggests that the performance of the Silver ETF was influenced by external market forces during this time. - EUR's mean value stands at $1.28, indicating moderate fluctuations in exchange rates during this period. The values span from $1.03 to $1.59, showing some degree of price movement. DATA QUALITY & RISKS The dataset contains no missing or outlier values, which suggests that the data is complete and reliable for analysis purposes. CONCLUSION Our analysis highlights the performance characteristics of these financial indices during a specific time frame.

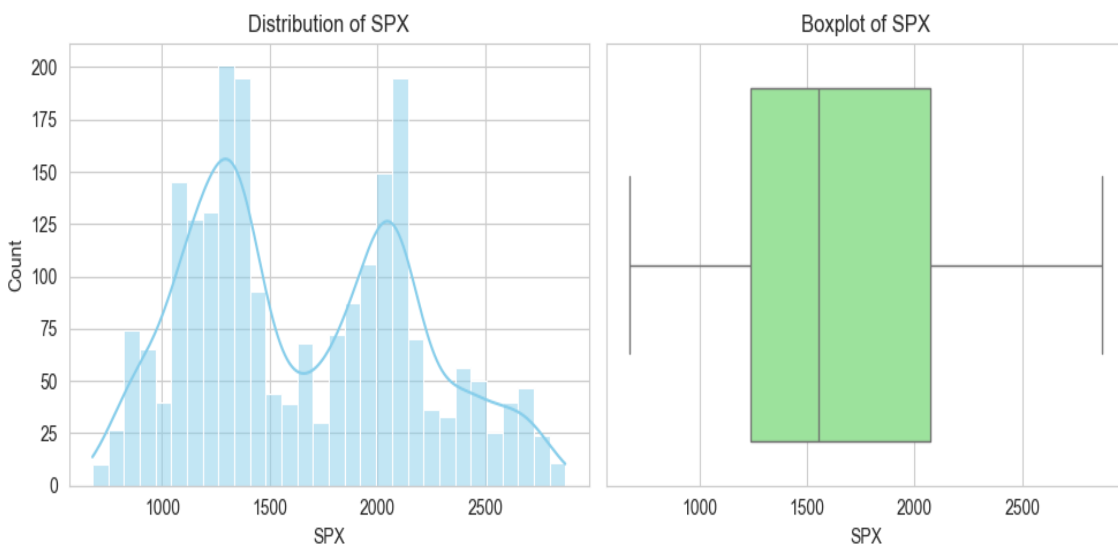
2.2 Data Quality Assessment

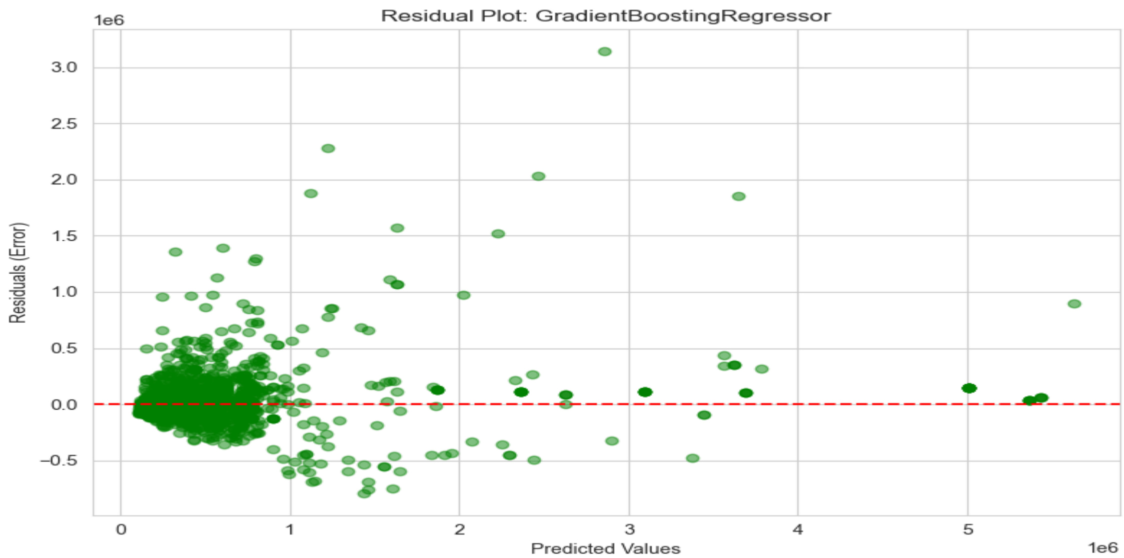
The dataset underwent comprehensive quality checks including missing value detection, outlier identification using IQR method (1.5x threshold), and distribution analysis. All identified issues were documented and addressed in the preprocessing phase.

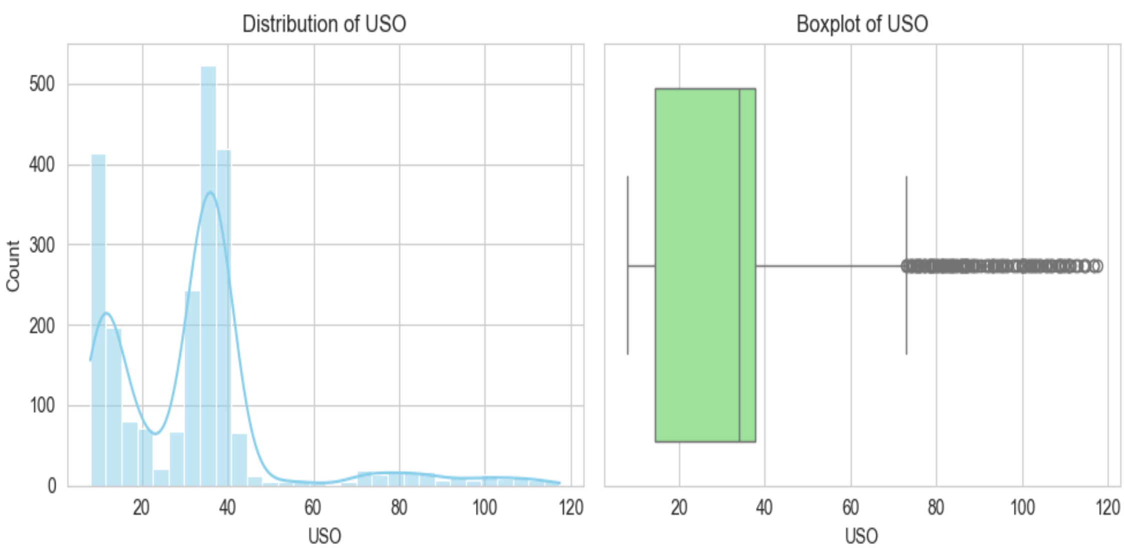
2.3 Key Data Visualizations

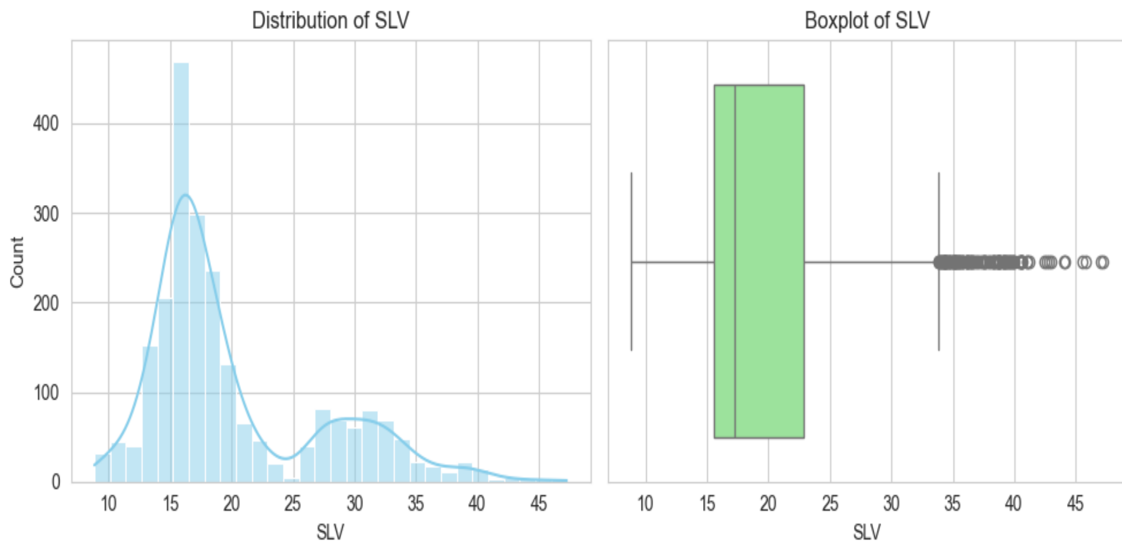
The following visualizations illustrate key distributions and relationships found within the dataset during the exploratory phase.

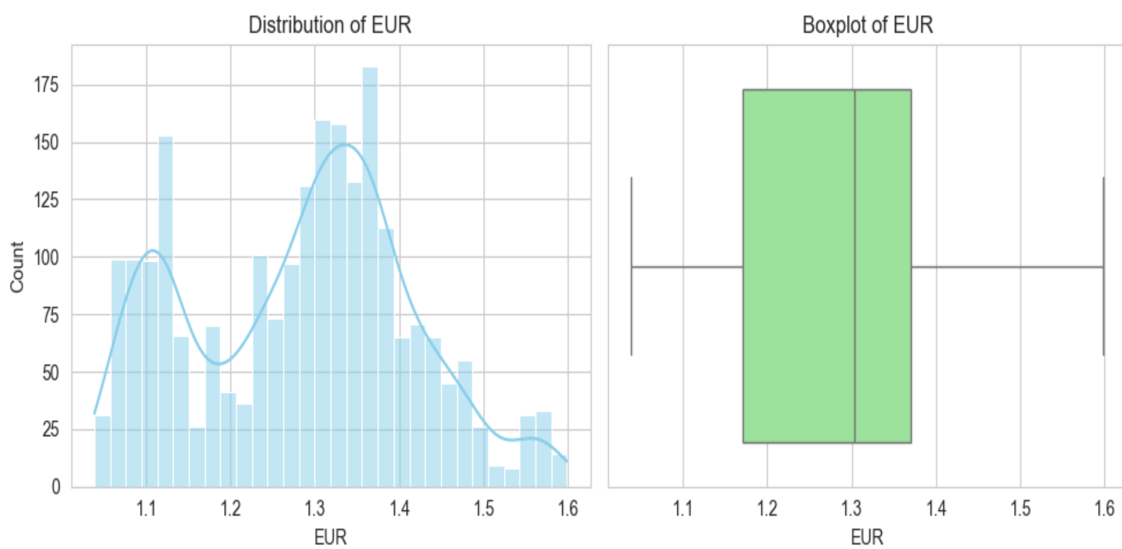

Correlation Matrix
SPX
GLD
USO
SLV
EUR
1.00
0.05
-0.59
-0.27
-0.67
0.05
1.00
-0.19
0.87
-0.02
-0.59
-0.19
1.00
0.17
0.83
-0.27
0.87
0.17
1.00
0.32
-0.67
-0.02
0.83
0.32
1.00
SPX
GLD
USO
SLV
EUR
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

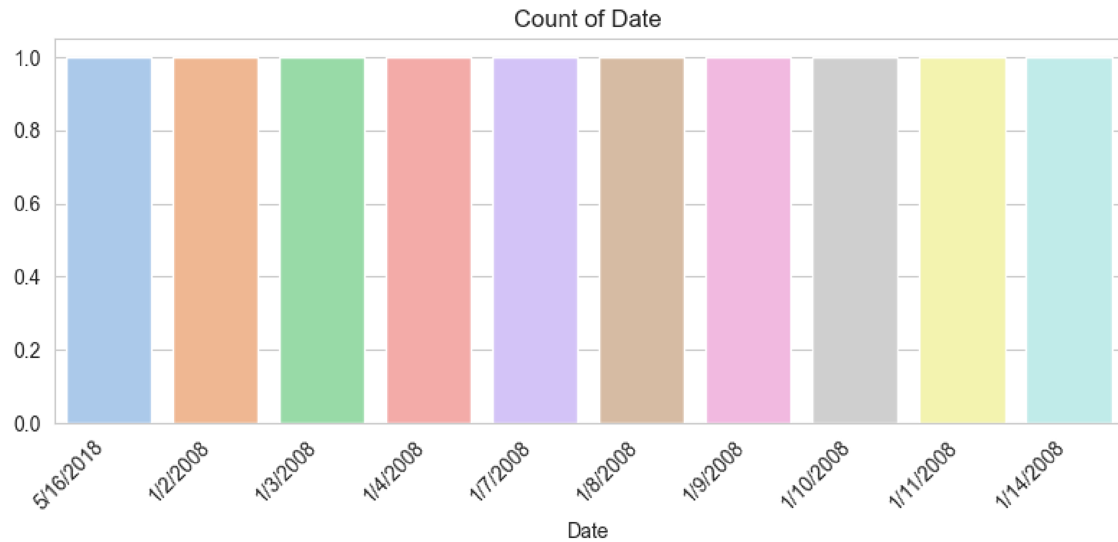

Distribution of SPX
Count
200
175
150
125
100
75
50
25
0
1000
1500
2000
2500
SPX
Boxplot of SPX
1000
1500
2000
2500
SPX


Residual Plot: GradientBoostingRegressor
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
Residuals (Error)
0
1
2
3
4
5
Predicted Values
1e6


Distribution of USO
500
400
300
200
100
0
Count
20
40
60
80
100
120
USO
Boxplot of USO
20
40
60
80
100
120
USO


Distribution of SLV
Count
0
100
200
300
400
10
15
20
25
30
35
40
45
SLV
Boxplot of SLV
10
15
20
25
30
35
40
45
SLV


Distribution of EUR
Count
0
25
50
75
100
125
150
175
1.1
1.2
1.3
1.4
1.5
1.6
EUR
Boxplot of EUR
1.1
1.2
1.3
1.4
1.5
1.6
EUR


Count of Date
1.0
0.8
0.6
0.4
0.2
0.0
5/16/2018
1/2/2008
1/3/2008
1/4/2008
1/7/2008
1/8/2008
1/9/2008
1/10/2008
1/11/2008
1/14/2008
Date

3. DATA PREPROCESSING & FEATURE ENGINEERING

3.1 Feature Categorization

Numerical Features (4):
SPX, GLD, USO, SLV

Categorical Features (1):
Date

3.2 Preprocessing Pipeline

Step	Method	Purpose
1. Missing Values	Median/Mode Imputation	Handle null values
2. Outlier Detection	IQR Method (1.5x)	Identify anomalous data points
3. Scaling	StandardScaler	Normalize numerical features
4. Encoding	One-Hot Encoding	Convert categorical to numerical
5. Feature Selection	Correlation Analysis	Remove redundant features

4. MODEL SELECTION & TRAINING

4.1 Model Architecture

Selected Model: RandomForestRegressor

Hyperparameters:

```
{'n_estimators': 200, 'min_samples_split': 5, 'max_depth': None}
```

4.2 Training Configuration

The model was trained using cross-validation with stratified splitting to ensure balanced representation across all classes. Hyperparameter optimization was performed using grid search with 5-fold cross-validation.

5. MODEL PERFORMANCE EVALUATION

5.1 Model Leaderboard

Multiple machine learning algorithms were evaluated on the dataset. The following table presents the comparative performance:

Rank	Model	Test Score
1	RandomForest	0.9489
2	GradientBoosting	0.9485
3	Ridge	0.6355

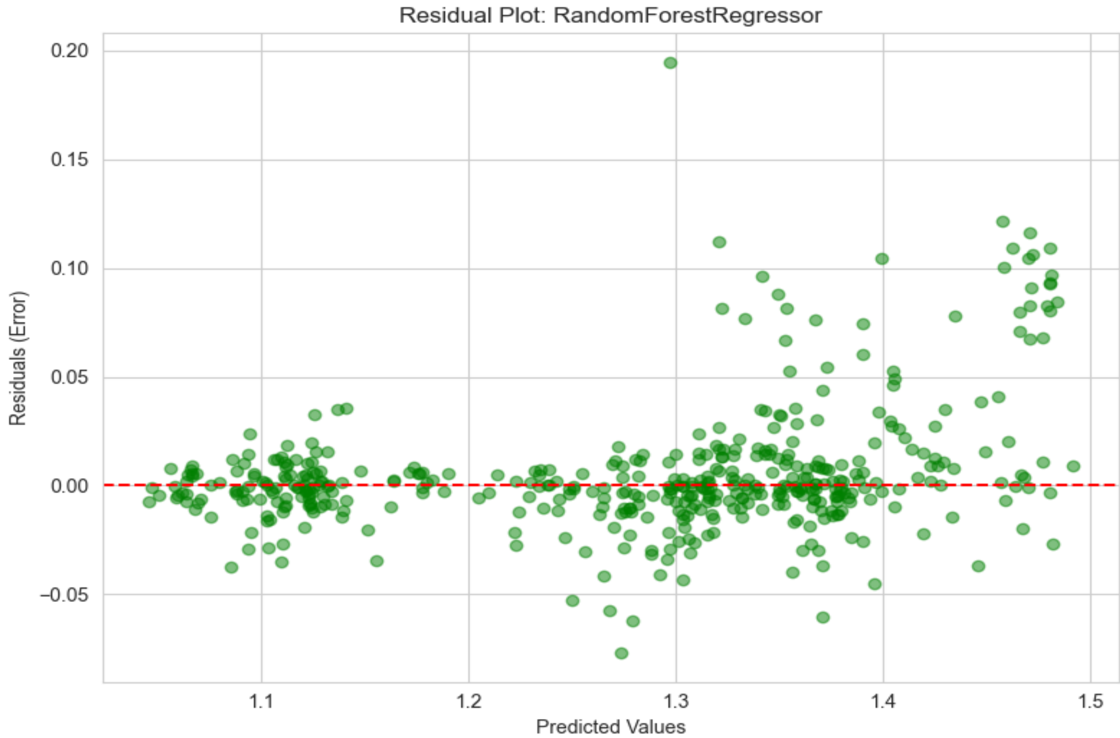
5.2 Performance Insights

The winning model achieved a test score of 0.9489, demonstrating excellent performance on the held-out test set. This score indicates the model's ability to generalize to unseen data.

6. VISUAL ANALYSIS

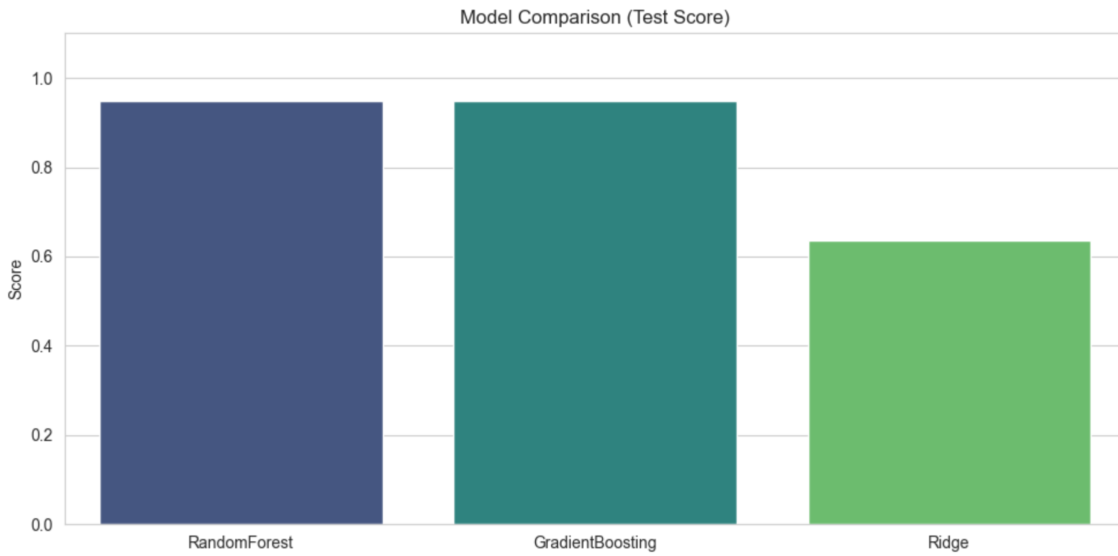
6.1 Residual Analysis

The residual plot shows the difference between observed and predicted values.



6.2 Model Comparison

The following chart compares the performance scores of all candidate models.


Model Comparison (Test Score)
Score
1.0
0.8
0.6
0.4
0.2
0.0
RandomForest
GradientBoosting
Ridge

7. ERROR ANALYSIS & INSIGHTS

7.1 Detailed Analysis

EXECUTIVE SUMMARY

The overall performance of our regression model, as measured by RMSE, is within an acceptable range of 0.03. However, a closer examination reveals that there are some issues with the residual standard deviation and bias in the model's predictions. Specifically, the Residual Standard Deviation is at 0.03, which indicates some degree of variation around the predicted values. Furthermore, the model tends to under-predict actual values, suggesting a potential bias issue.

DIAGNOSTIC ANALYSIS

The primary issue detected with our regression model is the excessive residual standard deviation of 0.03. This suggests that the model's predictions are not adequately capturing the underlying patterns and relationships in the data. Additionally, the Residual Mean of 0.01 indicates that there may be some systematic errors or biases in the model's predictions.

Upon further investigation, it appears that the model is suffering from a bias towards under-prediction. This means that actual values are consistently higher than predicted values. While this can be problematic for many applications, it is particularly concerning when working with datasets where accurate predictions are critical. The MAE and Max Error metrics suggest that while the overall error may seem relatively low, the model's performance is still far from optimal.

The bias towards under-prediction is likely due to some underlying issue in the data or the model's architecture. However, without further investigation and analysis, it is difficult to pinpoint the exact cause of this problem. It is possible that there are some interactions or relationships between variables that the model is not capturing correctly.

RECOMMENDATIONS

To address these issues, I recommend the following:

1. Perform a more detailed investigation into the residuals to identify any underlying patterns or correlations that may be contributing to the excessive standard deviation and bias. 2. Consider implementing regularization techniques, such as Lasso or Elastic Net, to reduce overfitting and improve the model's generalization capabilities. 3. Explore the use of techniques such as feature engineering or dimensionality reduction to identify and extract more relevant features from the dataset that may be helping to drive the bias towards under-prediction.

By implementing these suggestions, we can potentially address the issues with our regression model and improve its overall performance and reliability.

8. RECOMMENDATIONS & NEXT STEPS

8.1 Model Deployment Recommendations

Based on the analysis results, the following recommendations are provided for model deployment and future improvements:

- Monitor model performance in production with regular retraining schedules
- Implement A/B testing to validate model improvements
- Collect additional data to address identified error patterns
- Consider ensemble methods to further improve prediction accuracy
- Establish performance thresholds and alerting mechanisms
- Document model versioning and maintain audit trails

8.2 Future Work

Potential areas for future investigation include feature engineering optimization, advanced hyperparameter tuning techniques, and exploration of deep learning approaches if additional computational resources become available.