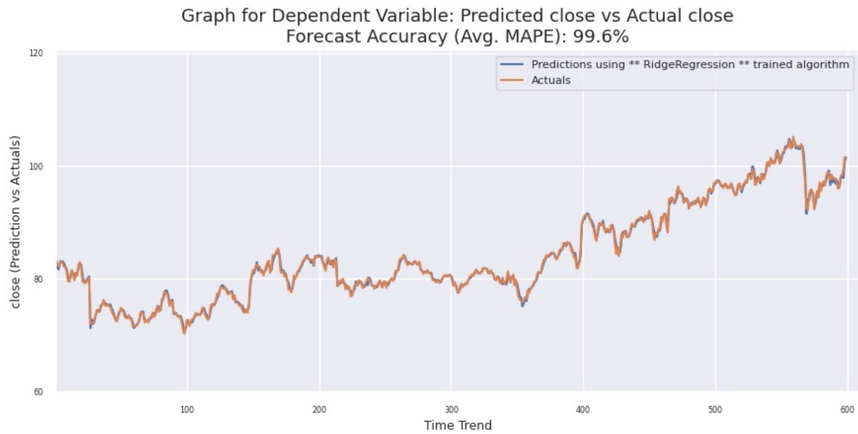
MAADSBML AutoML Report For FIERA CAPITAL

Generated On: 2024-04-24 18:21:35 (UTC)

Best Model(s) Report For admin_stockdata_csv



MODEL DESCRIPTION

Model Trained On: 2024/04/24
Training Start Time: 1817
Training End Time: 1821
Was Data Normalized: Yes
Was Data Shuffled: Yes
Deep Analysis: No
Total Training Data Set: 946
Training Data Percentage: 75%
Total Test Data Set: 314
Total # of Variables: 5
Adjusted for Seasonality: N
Total Algorithms Run: 3600
Removed Outliers: N
Best Distribution FOR ACTUAL Y: VONMISES
Dependent Variable: CLOSE
Independent Variables: ['Open', 'High', 'Low', 'Volume']

PREDICTION VARIABLE STATS

Mean: 84.297
STD: 8.424
Kurtosis: -0.647
Skewness: 0.611
Coef. of Variation: 0.100
Shapiro Test for Normality: 0.934
Jarque-Bera Goodness of Fit: 47.806
Anderson: 14.697
KStat: 71.074
KStatvar: 11.422
Wilcox: 0.000
Theil Slope: 0.043

ACTUAL VARIABLE STATS

Mean: 84.287
STD: 8.449
Kurtosis: -0.658
Skewness: 0.606
Coef. of Variation: 0.100
Shapiro Test for Normality: 0.934
Jarque-Bera Goodness of Fit: 47.497
Anderson: 14.514
KStat: 71.497
KStatvar: 11.462
Wilcox: 0.000
Theil Slope: 0.044

Statistics Showing Comparison Between Prediction and Actuals

Mood(actuals,predictions): 0.157
Pearson(actuals,predictions): 0.998
Kendall Tau(actuals,predictions): 0.957
Ansari(actuals,predictions): 179951.000
Jaccard_distance(actuals,predictions): 1.000
Minkowski_distance(actuals,predictions): 222.692
Euclidean_distance(actuals,predictions): 11.637

IMPORTANT FILE PATHS FOR RAW AND OUTPUT DATA

NOTE: These are DOCKER CONTAINER Paths. You can view these files inside the container by using the command: **docker exec -it {container id} bash If you have re-run the container, these files will be GONE but they exist on your HOST machine. The HOST MACHINE location is based on the volumes you mapped when you ran the Docker container.**
The Docker RUN Volume Mappings are :: (For example here is the docker run command (use multiple -v for multiple mappings):

DOCKER RUN COMMAND: docker run -d -p 5595:5595 -p 5495:5495 -p 10000:10000 -v {HOST MACHINE FOLDER}:{CONTAINER FOLDER}:z --env TRAININGPORT=5595 --env PREDICTIONPORT=5495 --env ABORTPORT=10000 --env COMPANYNAME=MYCOMPANY --env MAXRUNTIME=20 --env MAINHOST=127.0.0.1 maadsdocker/maads-batch-automl-otics

Docker Volume Mappings:
1. {HOST MACHINE FOLDER}/csvuploads:/maads/agentfilesdocker/dist/maadsweb/csvuploads:z
2. {HOST MACHINE FOLDER}/pdfreports:/maads/agentfilesdocker/dist/maadsweb/pdfreports:z
3. {HOST MACHINE FOLDER}/autofeatures:/maads/agentfilesdocker/dist/maadsweb/autofeatures:z
4. {HOST MACHINE FOLDER}/outliers:/maads/agentfilesdocker/dist/maadsweb/outliers:z
5. {HOST MACHINE FOLDER}/sqlloads:/maads/agentfilesdocker/dist/maadsweb/sqlloads:z
6. {HOST MACHINE FOLDER}/networktemp:/maads/agentfilesdocker/dist/maadsweb/networktemp:z
7. {HOST MACHINE FOLDER}/networks:/maads/agentfilesdocker/networks:z
8. {HOST MACHINE FOLDER}/exception:/maads/agentfilesdocker/dist/maadsweb/exception:z
9. {HOST MACHINE FOLDER}/staging:/maads/agentfilesdocker/dist/staging:z

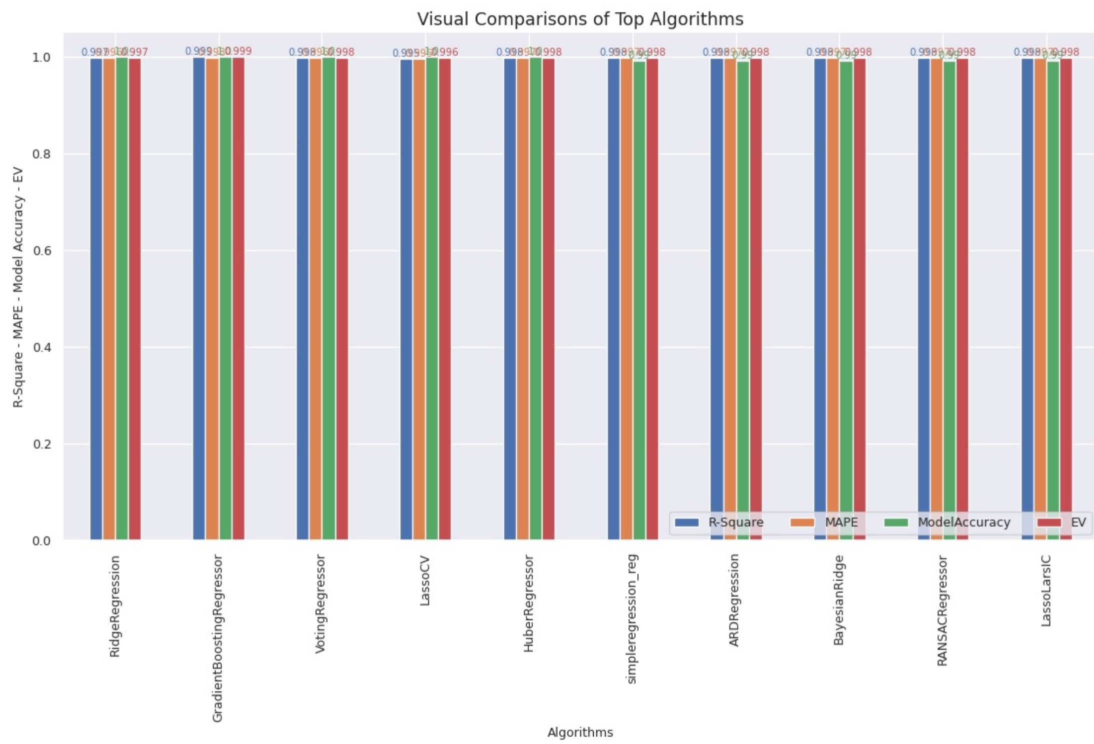
Path for Training Dataset File: /maads/agentfilesdocker/dist/maadsweb/csvuploads/stockdata.csv
Path for PDF Report (i.e. this file): /maads/agentfilesdocker/dist/maadsweb/pdfreports/admin_stockdata_csv_no_seasons.pdf
Path for AutoFeature File: /maads/agentfilesdocker/dist/maadsweb/autofeatures/admin_stockdata_csv_.csv
Path for Outliers File: /maads/agentfilesdocker/dist/maadsweb/outliers/admin_stockdata_csv.csv
Path for Algo JSON File: /maads/agentfilesdocker/dist/maadsweb/exception/admin_stockdata_csv_trained_algo_no_seasons.json
Folder Path for MySQL Scripts: /maads/agentfilesdocker/dist/maadsweb/sqlloads/
Path for Detailed Prediction File: /maads/agentfilesdocker/dist/maadsweb/csvuploads/admin_stockdata_csv_prediction_details.csv
Path for Algorithm Zip File (i.e pickle files): /maads/agentfilesdocker/dist/maadsweb/networktemp/admin_stockdata_csv.zip
Path for Algorithm Pickle Files:
1. /maads/agentfilesdocker/networks/Alberta-Electric-System-Operator_AESO)_ADMIN_STOCKDATA_CSVALLSEASON_AG1_4_RidgeRegression_normal_1.00000000_946_.pkl
2. /maads/agentfilesdocker/networks/Alberta-Electric-System-Operator_AESO)_ADMIN_STOCKDATA_CSVALLSEASON_AG1_4_RidgeRegression_normal_1.00000000_946_scalerx_.pkl
3. /maads/agentfilesdocker/networks/Alberta-Electric-System-Operator_AESO)_ADMIN_STOCKDATA_CSVALLSEASON_AG1_4_RidgeRegression_normal_1.00000000_946_scalery_.pkl

DESCRIPTIVE STATISTICS

Variables	T-Statistic	Count	Mean	STD	MIN	25%	50%	75%	MAX
Open	1221.412	946.0	94.354	70.58	15.464	80.4	93.29	107.905	125.25
High	1233.139	946.0	95.083	70.75	15.58	80.892	93.925	109.128	125.72
Low	1.521615424968217e+16	946.0	93.697	69.0	15.379	79.803	92.285	107.398	123.68
Volume	-9.308	946.0	2723209.302	686200.0	1440702.602	1984825.0	2433550.0	3089250.0	28284300.0
CLOSE	NA	946.0	94.437	70.28	15.504	80.43	93.295	108.152	125.4

BEST ALGORITHM FOUND FOR THIS DATASET (Note: This trained model will be used to predict CLOSE)					
Algorithm	**Description**	**Model Results**	**MAPE Accuracy**	**Forecast Months**	**Season**
RidgeRegression	Ridge Regression: Linear least squares with l2 regularization.	Ridge(alpha=0.14549464103297374, random_state=48, solver='saga') Coefficient: [0.0776637186596356, 0.4807798231629039, 0.4366212011905196, 0.008674525867625061] **R-square:** 0.997 **Mean Squared Error (MSE):** 0.226 **Skewness:** -0.355 **Kurtosis:** 3.979 **Mean Square Model (MSM):** 42573.293 **F-Statistic (F):** 188641.861 **Jarque-Bera (JB):** 36.558 **Explained Variance (EV):** 0.997 **Multicolinearity Test (Avg. VIF):** inf **Heteroscedasticity Test (Avg P-Value):** -1.000 **(Based on White Test, there seems to be heteroscedasticity in the model)** **Autocorrelation (Durbin-Watson) Test:** 1.973 **(Based on DW Test - there seems to be autocorrelation in your model)**	0.996	1 - 12	allseason

TOP 10 ALGORITHMS FOR ALLSEASON

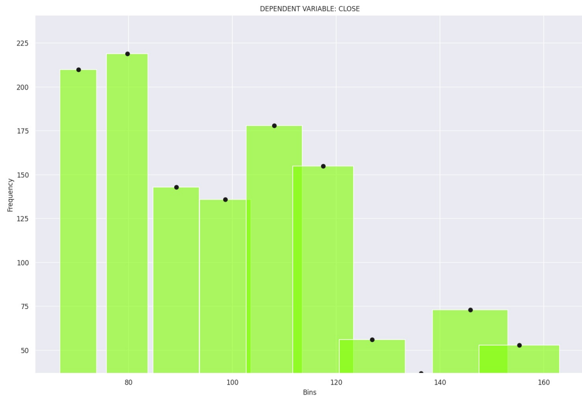


Num	Algorithm	Model Accuracy	Details	Season	Description
1	RidgeRegression	0.9960	**R-square:** 0.997 **MAPE:** 0.996 **Explained Variance (EV):** 0.997 **MSE:** 0.226 **MSM:** 42573.293 **Skewness:** -0.355 **Kurtosis:** 3.979 **F:** 188641.861 **DW:** 1.973 **JB:** 36.558	allseason	RIDGE REGRESSION: Linear least squares with l2 regularization.
2	GradientBoostingRegressor	0.9957	**R-square:** 0.999 **MAPE:** 0.998 **Explained Variance (EV):** 0.999 **MSE:** 0.081 **MSM:** 42498.172 **Skewness:** -2.203 **Kurtosis:** 10.673 **F:** 522015.718 **DW:** 1.723 **JB:** 1957.217	allseason	GRADIENT BOOSTING FOR REGRESSION.: GB builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. In each stage a regression tree is fit on the negative gradient of the given loss function.
3	VotingRegressor	0.9956	**R-square:** 0.998 **MAPE:** 0.996 **Explained Variance (EV):** 0.998 **MSE:** 0.177 **MSM:** 42201.548 **Skewness:** -0.028 **Kurtosis:** 3.649 **F:** 237926.228 **DW:** 1.881 **JB:** 10.62	allseason	VOTINGREGRESSOR:
4	LassoCV	0.9955	**R-square:** 0.995 **MAPE:** 0.995 **Explained Variance (EV):** 0.996 **MSE:** 0.362 **MSM:** 40451.365 **Skewness:** -1.673 **Kurtosis:** 4.421 **F:** 111849.886 **DW:** 1.128 **JB:** 330.376	allseason	LASSO CROSS VALIDATION: Lasso linear model with iterative fitting along a regularization path
5	HuberRegressor	0.9950	**R-square:** 0.998 **MAPE:** 0.997 **Explained Variance (EV):** 0.998 **MSE:** 0.138 **MSM:** 42535.281 **Skewness:** -0.021 **Kurtosis:** 4.145 **F:** 308842.817 **DW:** 2.064 **JB:** 32.817	allseason	HUBER REGRESSOR: Linear regression model that is robust to outliers.
			R-square: 0.998 **MAPE:** 0.997 **Explained Variance (EV):** 0.998 **MSE:** 0.138		

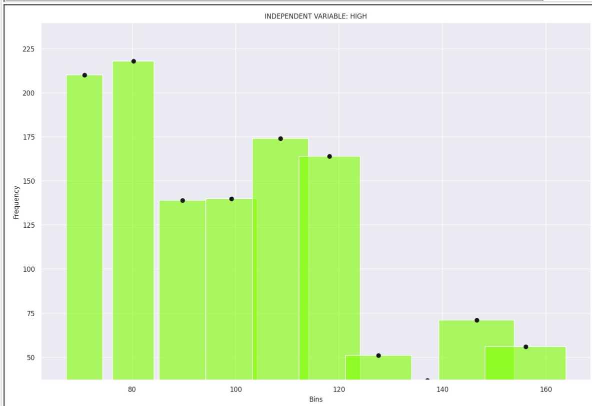
6	simpleregression_reg	0.9950	**MSM:** 42521.698 **Skewness:** 0.092 **Kurtosis:** 4.302 **F:** 308608.598 **DW:** 2.07 **JB:** 43.237	allseason	LINEAR REGRESSION: Simple multivariate linear regression
7	ARDRegression	0.9949	**R-square:** 0.998 **MAPE:** 0.997 **Explained Variance (EV):** 0.998 **MSE:** 0.138 **MSM:** 42525.09 **Skewness:** 0.044 **Kurtosis:** 4.211 **F:** 308290.472 **DW:** 2.068 **JB:** 36.864	allseason	BAYESIAN ARD: Fit the weights of a regression model, using an ARD prior. The weights of the regression model are assumed to be in Gaussian distributions. Also estimate the parameters lambda (precisions of the distributions of the weights) and alpha (precision of the distribution of the noise). The estimation is done by an iterative procedures (Evidence Maximization)
8	BayesianRidge	0.9949	**R-square:** 0.998 **MAPE:** 0.997 **Explained Variance (EV):** 0.998 **MSE:** 0.138 **MSM:** 42524.375 **Skewness:** 0.048 **Kurtosis:** 4.218 **F:** 308418.724 **DW:** 2.069 **JB:** 37.305	allseason	BAYESIAN RIDGE REGRESSION: Fit a Bayesian ridge model and optimize the regularization parameters lambda (precision of the weights) and alpha (precision of the noise).
9	RANSACRegressor	0.9948	**R-square:** 0.998 **MAPE:** 0.997 **Explained Variance (EV):** 0.998 **MSE:** 0.138 **MSM:** 42521.698 **Skewness:** 0.092 **Kurtosis:** 4.302 **F:** 308608.598 **DW:** 2.07 **JB:** 43.237	allseason	RANSAC (RANDOM SAMPLE CONSENSUS) ALGORITHM.: RANSAC is an iterative algorithm for the robust estimation of parameters from a subset of inliers from the complete data set. More information can be found in the general documentation of linear models.
10	LassoLarsIC	0.9948	**R-square:** 0.998 **MAPE:** 0.997 **Explained Variance (EV):** 0.998 **MSE:** 0.138 **MSM:** 42521.698 **Skewness:** 0.092 **Kurtosis:** 4.302 **F:** 308608.598 **DW:** 2.07 **JB:** 43.237	allseason	LASSO MODEL FIT WITH LARS USING BIC OR AIC FOR MODEL SELECTION: Lasso model fit with Lars using BIC or AIC for model selection

Detailed Histograms of Training and Test Data Sets

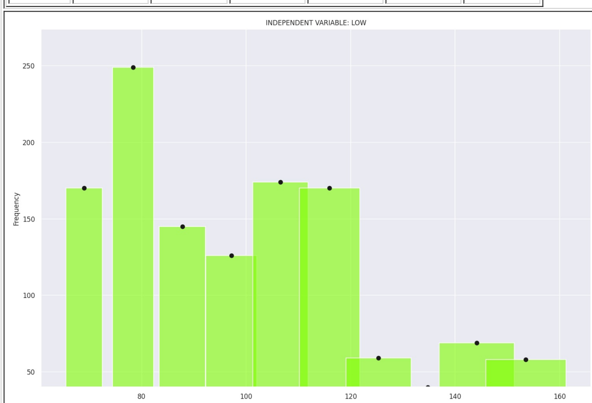
TRAINING VARIABLES



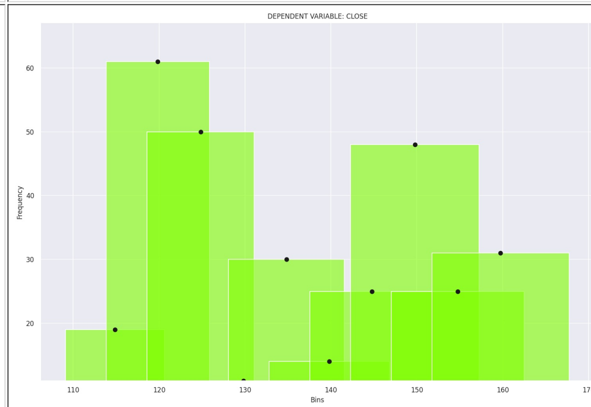
Bins	[7.03e+01, 7.97e+01]	[7.97e+01, 8.92e+01]	[8.92e+01, 9.86e+01]	[9.86e+01, 1.08e+02]	[1.08e+02, 1.18e+02]	[1.18e+02, 1.27e+02]
Count	210	219	143	136	178	155
Share	17.0%	17.0%	11.0%	11.0%	14.0%	12.0%
Total Rows	1260	1260	1260	1260	1260	1260
Min	7.00e+01	7.00e+01	7.00e+01	7.00e+01	7.00e+01	7.00e+01
Max	1.64e+02	1.64e+02	1.64e+02	1.64e+02	1.64e+02	1.64e+02
Number of Bins	6	6	6	6	6	6



Bins	[7.08e+01, 8.02e+01]	[8.02e+01, 8.97e+01]	[8.97e+01, 9.92e+01]	[9.92e+01, 1.09e+02]	[1.09e+02, 1.18e+02]	[1.18e+02, 1.28e+02]
Count	210	218	139	140	174	164
Share	17.0%	17.0%	11.0%	11.0%	14.0%	13.0%
Total Rows	1260	1260	1260	1260	1260	1260
Min	7.00e+01	7.00e+01	7.00e+01	7.00e+01	7.00e+01	7.00e+01
Max	1.65e+02	1.65e+02	1.65e+02	1.65e+02	1.65e+02	1.65e+02
Number of Bins	6	6	6	6	6	6

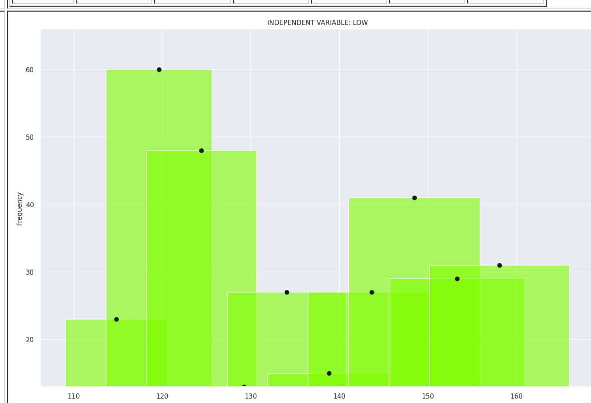


TEST VARIABLES



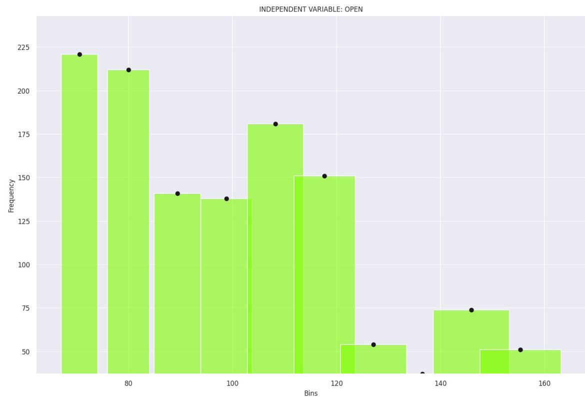
Bins	[1.15e+02, 1.20e+02]	[1.20e+02, 1.25e+02]	[1.25e+02, 1.30e+02]	[1.30e+02, 1.35e+02]	[1.35e+02, 1.40e+02]	[1.40e+02, 1.45e+02]
Count	19	61	50	11	30	14
Share	6.0%	19.0%	16.0%	4.0%	10.0%	4.0%
Total Rows	314	314	314	314	314	314
Min	1.14e+02	1.14e+02	1.14e+02	1.14e+02	1.14e+02	1.14e+02
Max	1.64e+02	1.64e+02	1.64e+02	1.64e+02	1.64e+02	1.64e+02
Number of Bins	6	6	6	6	6	6

Bins	[1.16e+02, 1.21e+02]	[1.21e+02, 1.26e+02]	[1.26e+02, 1.31e+02]	[1.31e+02, 1.36e+02]	[1.36e+02, 1.41e+02]	[1.41e+02, 1.46e+02]
Count	19	64	51	7	30	16
Share	6.0%	20.0%	16.0%	2.0%	10.0%	5.0%
Total Rows	314	314	314	314	314	314
Min	1.15e+02	1.15e+02	1.15e+02	1.15e+02	1.15e+02	1.15e+02
Max	1.65e+02	1.65e+02	1.65e+02	1.65e+02	1.65e+02	1.65e+02
Number of Bins	6	6	6	6	6	6

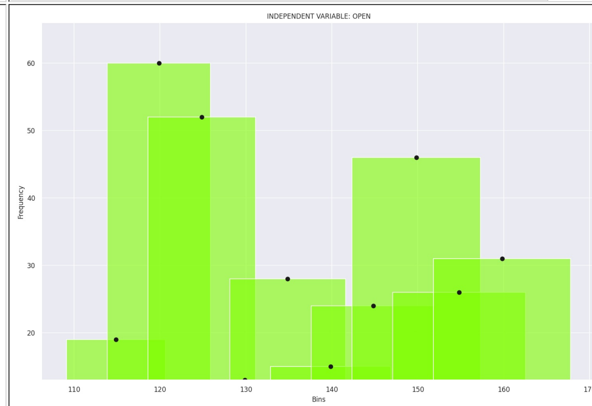


Bins	[6.90e+01, 7.84e+01]	[7.84e+01, 8.78e+01]	[8.78e+01, 9.72e+01]	[9.72e+01, 1.07e+02]	[1.07e+02, 1.16e+02]	[1.16e+02, 1.25e+02]
Count	170	249	145	126	174	170
Share	**13.0%**	**20.0%**	**12.0%**	**10.0%**	**14.0%**	**13.0%**
Total Rows	1260	1260	1260	1260	1260	1260
Min	6.90e+01	6.90e+01	6.90e+01	6.90e+01	6.90e+01	6.90e+01
Max	1.62e+02	1.62e+02	1.62e+02	1.62e+02	1.62e+02	1.62e+02
Number of Bins	6	6	6	6	6	6

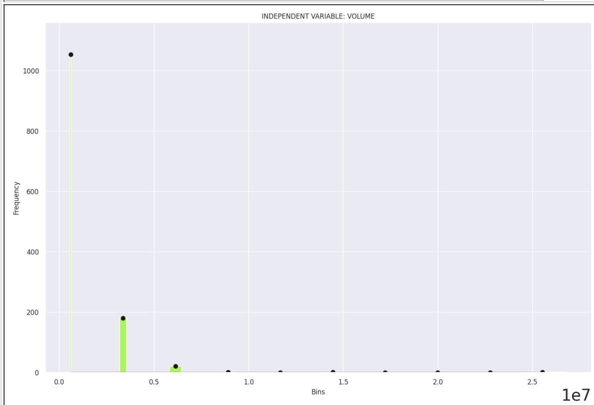
Bins	[1.15e+02, 1.20e+02]	[1.20e+02, 1.24e+02]	[1.24e+02, 1.29e+02]	[1.29e+02, 1.34e+02]	[1.34e+02, 1.39e+02]	[1.39e+02, 1.44e+02]
Count	23	60	48	13	27	15
Share	**7.0%**	**19.0%**	**15.0%**	**4.0%**	**9.0%**	**5.0%**
Total Rows	314	314	314	314	314	314
Min	1.14e+02	1.14e+02	1.14e+02	1.14e+02	1.14e+02	1.14e+02
Max	1.62e+02	1.62e+02	1.62e+02	1.62e+02	1.62e+02	1.62e+02
Number of Bins	6	6	6	6	6	6



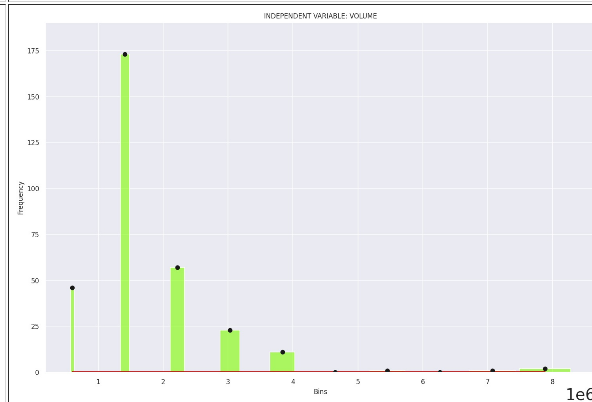
Bins	[7.06e+01, 8.00e+01]	[8.00e+01, 8.94e+01]	[8.94e+01, 9.88e+01]	[9.88e+01, 1.08e+02]	[1.08e+02, 1.18e+02]	[1.18e+02, 1.27e+02]
Count	221	212	141	138	181	151
Share	**18.0%**	**17.0%**	**11.0%**	**11.0%**	**14.0%**	**12.0%**
Total Rows	1260	1260	1260	1260	1260	1260
Min	7.00e+01	7.00e+01	7.00e+01	7.00e+01	7.00e+01	7.00e+01
Max	1.64e+02	1.64e+02	1.64e+02	1.64e+02	1.64e+02	1.64e+02
Number of Bins	6	6	6	6	6	6



Bins	[1.15e+02, 1.20e+02]	[1.20e+02, 1.25e+02]	[1.25e+02, 1.30e+02]	[1.30e+02, 1.35e+02]	[1.35e+02, 1.40e+02]	[1.40e+02, 1.45e+02]
Count	19	60	52	13	28	15
Share	**6.0%**	**19.0%**	**17.0%**	**4.0%**	**9.0%**	**5.0%**
Total Rows	314	314	314	314	314	314
Min	1.14e+02	1.14e+02	1.14e+02	1.14e+02	1.14e+02	1.14e+02
Max	1.64e+02	1.64e+02	1.64e+02	1.64e+02	1.64e+02	1.64e+02
Number of Bins	6	6	6	6	6	6

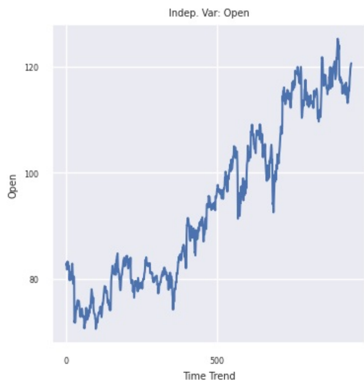


Bins	[6.02e+05, 3.37e+06]	[3.37e+06, 6.14e+06]	[6.14e+06, 8.91e+06]	[8.91e+06, 1.17e+07]	[1.17e+07, 1.44e+07]	[1.44e+07, 1.72e+07]
Count	1054	181	21	2	0	1
Share	**84.0%**	**14.0%**	**2.0%**	**0.0%**	**0.0%**	**0.0%**
Total Rows	1260	1260	1260	1260	1260	1260
Min	6.02e+05	6.02e+05	6.02e+05	6.02e+05	6.02e+05	6.02e+05
Max	2.83e+07	2.83e+07	2.83e+07	2.83e+07	2.83e+07	2.83e+07
Number of Bins	6	6	6	6	6	6

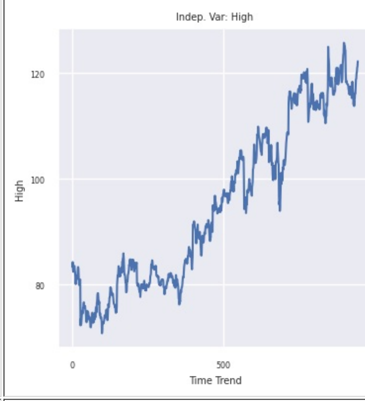


Bins	[6.02e+05, 1.41e+06]	[1.41e+06, 2.22e+06]	[2.22e+06, 3.03e+06]	[3.03e+06, 3.84e+06]	[3.84e+06, 4.65e+06]	[4.65e+06, 5.46e+06]
Count	46	173	57	23	11	0
Share	**15.0%**	**55.0%**	**18.0%**	**7.0%**	**4.0%**	**0.0%**
Total Rows	314	314	314	314	314	314
Min	6.02e+05	6.02e+05	6.02e+05	6.02e+05	6.02e+05	6.02e+05
Max	8.69e+06	8.69e+06	8.69e+06	8.69e+06	8.69e+06	8.69e+06
Number of Bins	6	6	6	6	6	6

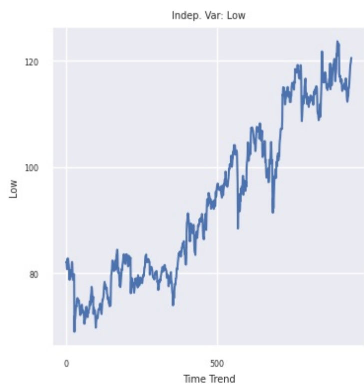
Detailed Graphs of Variables Against Time



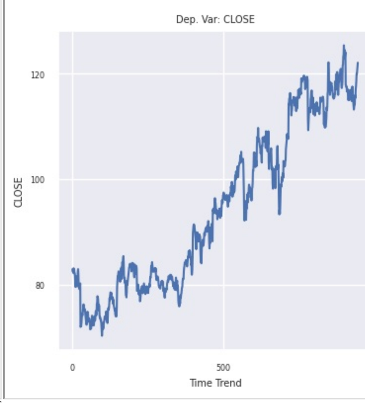
	Stats
Min	70.58
Max	125.25
Median	93.29
Mean	94.354
STD	15.456
Kurtosis	-1.346
Skewness	0.261
Coeff. Variation	0.164
Shapiro Normality	0.918



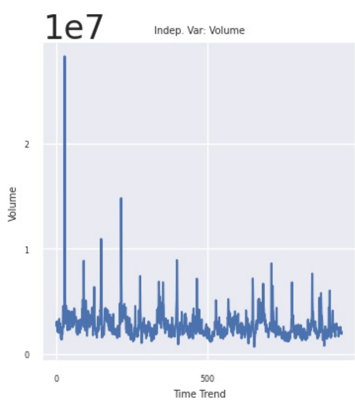
	Stats
Min	70.75
Max	125.72
Median	93.925
Mean	95.083
STD	15.572
Kurtosis	-1.35
Skewness	0.26
Coeff. Variation	0.164
Shapiro Normality	0.917



	Stats
Min	69.0
Max	123.68
Median	92.285
Mean	93.697
STD	15.371
Kurtosis	-1.341
Skewness	0.268
Coeff. Variation	0.164
Shapiro Normality	0.918

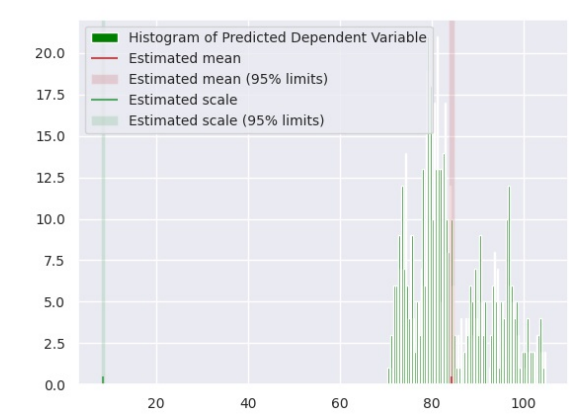


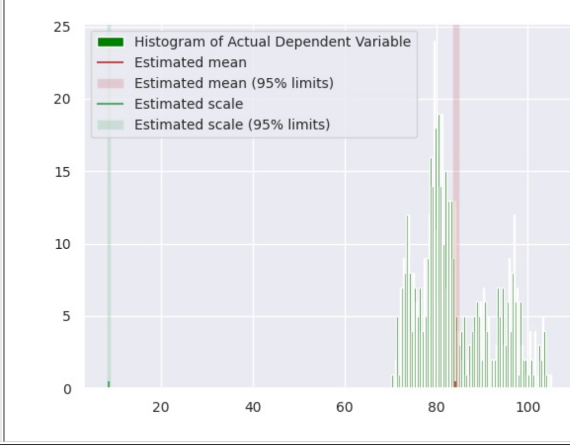
	Stats
Min	70.28
Max	125.4
Median	93.295
Mean	94.437
STD	15.496
Kurtosis	-1.347
Skewness	0.262
Coeff. Variation	0.164
Shapiro Normality	0.918

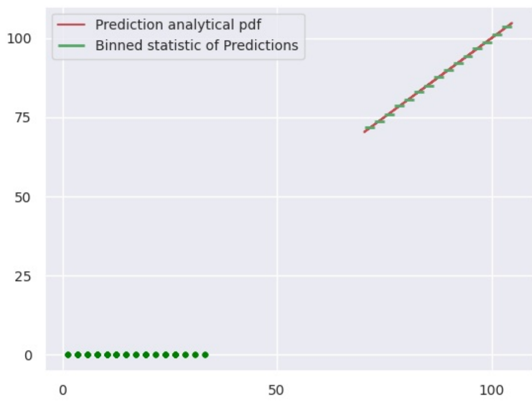
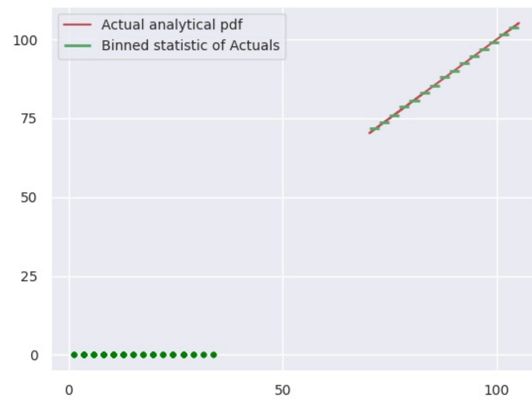


	Stats
Min	686200.0
Max	28284300.0
Median	2433550.0
Mean	2723209.302
STD	1439940.93
Kurtosis	110.956
Skewness	7.476
Coeff. Variation	0.529
Shapiro Normality	0.611

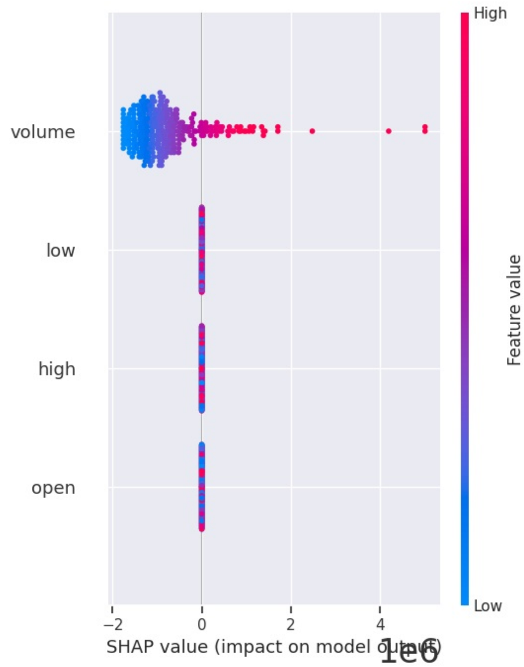
Bayesian Distribution





Binned Statistic	
	

MODEL EXPLANATION

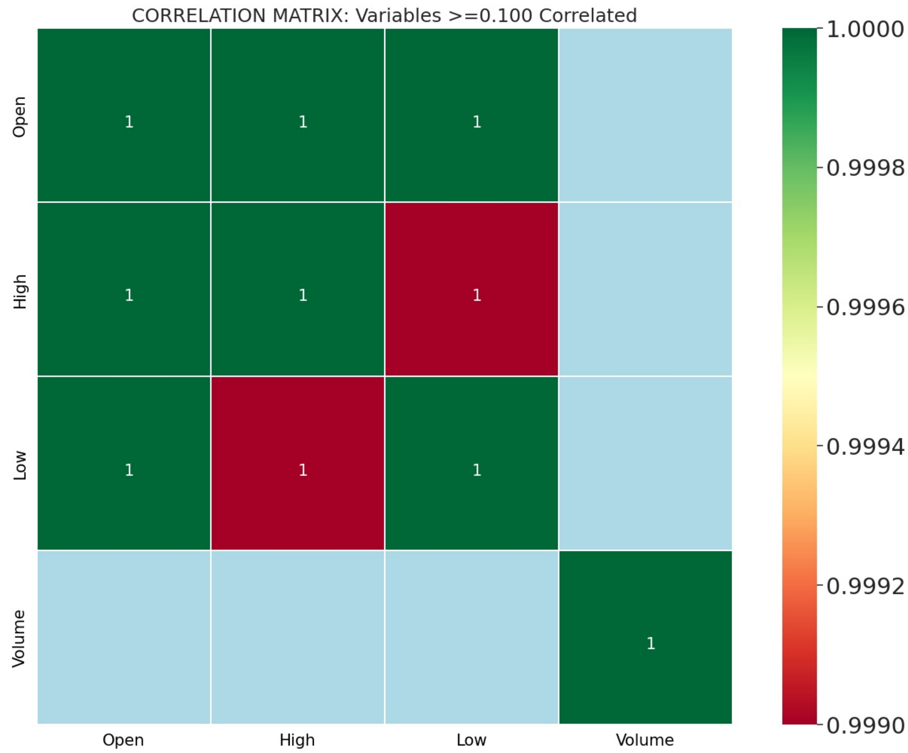


- The x-axis represents the model's output values of **CLOSE**
- The plot is centered on the x-axis at explainer.expected_value.
- All values are relative to the model's expected value like a linear model's effects are relative to the intercept.
- The y-axis lists the model's features. By default, the features are ordered by descending importance.
- The importance is calculated over the observations plotted. _This is usually different than the importance ordering for the entire dataset.
- In addition to feature importance ordering, the decision plot also supports hierarchical cluster feature ordering and user-defined feature ordering.
- Each observation's prediction is represented by a colored line.
- At the top of the plot, each line strikes the x-axis at its corresponding observation's predicted value. This value determines the color of the line on a spectrum.
- Moving from the bottom of the plot to the top, SHAP values for each feature are added to the model's base value.
- This shows how each feature contributes to the overall prediction.
- At the bottom of the plot, the observations converge at explainer.expected_value.
- The points in the graph are the values of the feature in the training dataset.

FEATURE SELECTION	
RFE Variable (Most important to Least Important)	**Value**
Close	0.414
High	0.206
Low	0.192
Open	0.128
Volume	0.059
Best Variable(s) From Genetic Algorithm	
Close	
Volume	
Volume	
Excluded Variable(s)	
High	
Low	
Open	
PCA for Best Variable(s)	Value
Close_pca_1	-0.707
Close_pca_2	0.707
Close_pca_3	-0.708
High_pca_1	-0.673
High_pca_2	0.218
High_pca_3	0.706
Volume_pca_1	0.707
Volume_pca_2	0.707
Volume_pca_3	-0.006
PCA Explained Variance	Value
PCA1	0.627
PCA2	0.373
PCA3	0.000

- Feature selection shows which variables were more influential than other variables
- It uses two core algorithms: Recursive Feature Elimination (RFE) and Genetic Algorithm to determine influence
- It also performs PCA (principal component analysis) analysis to determine the influence of the best variables in the model
- These results should be used in conjunction with other information as well as theory to establish relevance and confidence in the chosen model formulation

CORRELATION MATRIX



CORRELATED FEATURES

	Feature(s)	Feature(s)	Correlation >= 0.100
0	High	Low	0.999
1	Open	Open	NaN

SUGGESTED CORRELATED FEATURES TO DELETE

	1 Feature(s) to Delete	Correlation
0	High	0.999

END OF REPORT

MAADSBML Python Library:https://pypi.org/project/maadsbml/
MAADSBML Docker Container For Windows: https://hub.docker.com/r/maadsdocker/maads-batch-automl-otics
MAADSBML Docker Container For MAC: https://hub.docker.com/r/maadsdocker/maads-batch-automl-otics-arm64
MAADSBML Sample Code and Setup: https://github.com/smaurice101/raspberrypi/tree/main/maadsbml

MAADSBML
Developed and Maintained by: Otics Advanced Analytics, Inc.
Toronto, Ontario, Canada
https://www.otics.ca
Email: support@otics.ca