Anomaly Detection in Univariate Time Series: Rule-Based, Unsupervised, Regression and Conformal Prediction Approaches

Date : 2024/12/05

Ajeet Kumar Rai
Department of Applied Mathematics and Statistics, Stony Brook University
ajeetkumar.rai@stonybrook.edu

INTRODUCTION

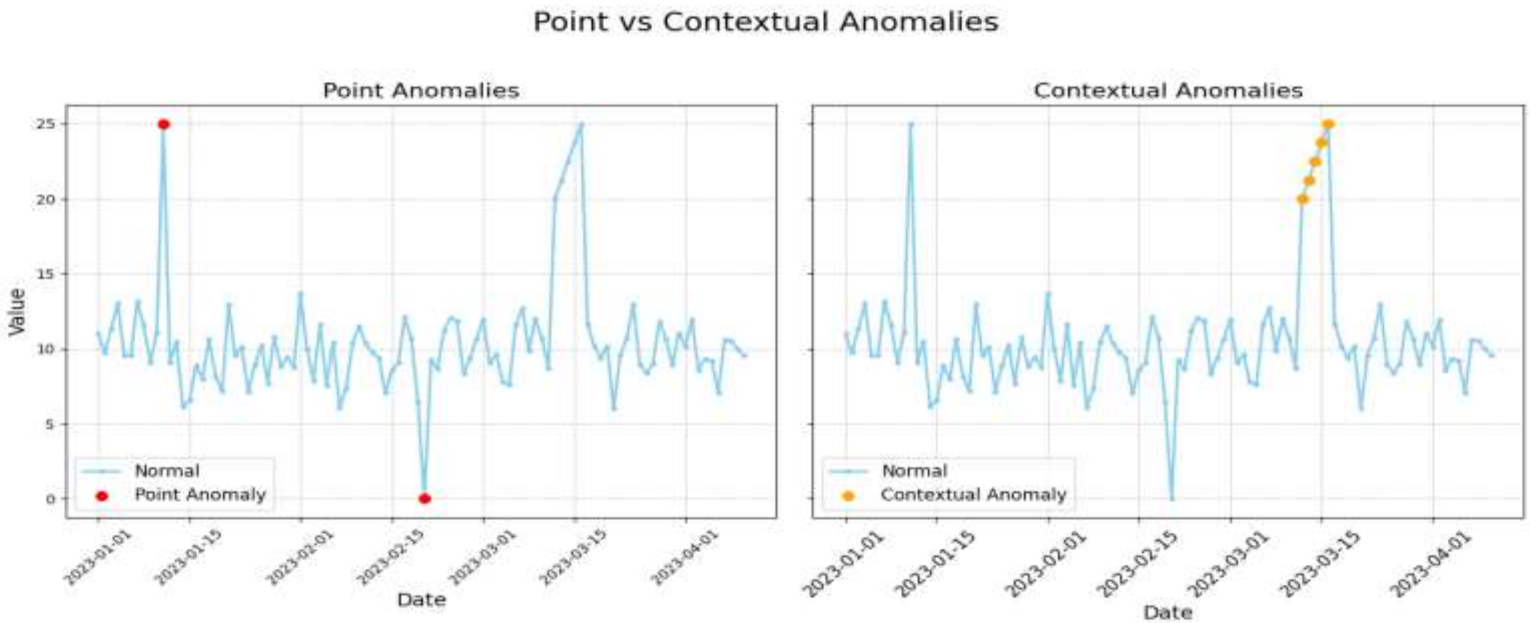
- Anomalies are data points that deviate significantly from the overall time series trend.
- These deviations often reflect unique or special events within the time series data.
- Such anomalies are rare and occur rarely.
- In financial time series, anomalies often result in significant costs if ignored.
- With effective modeling, these anomalies can also present opportunities for profit.

ANOMALY

- Anomaly is something that is not normal.

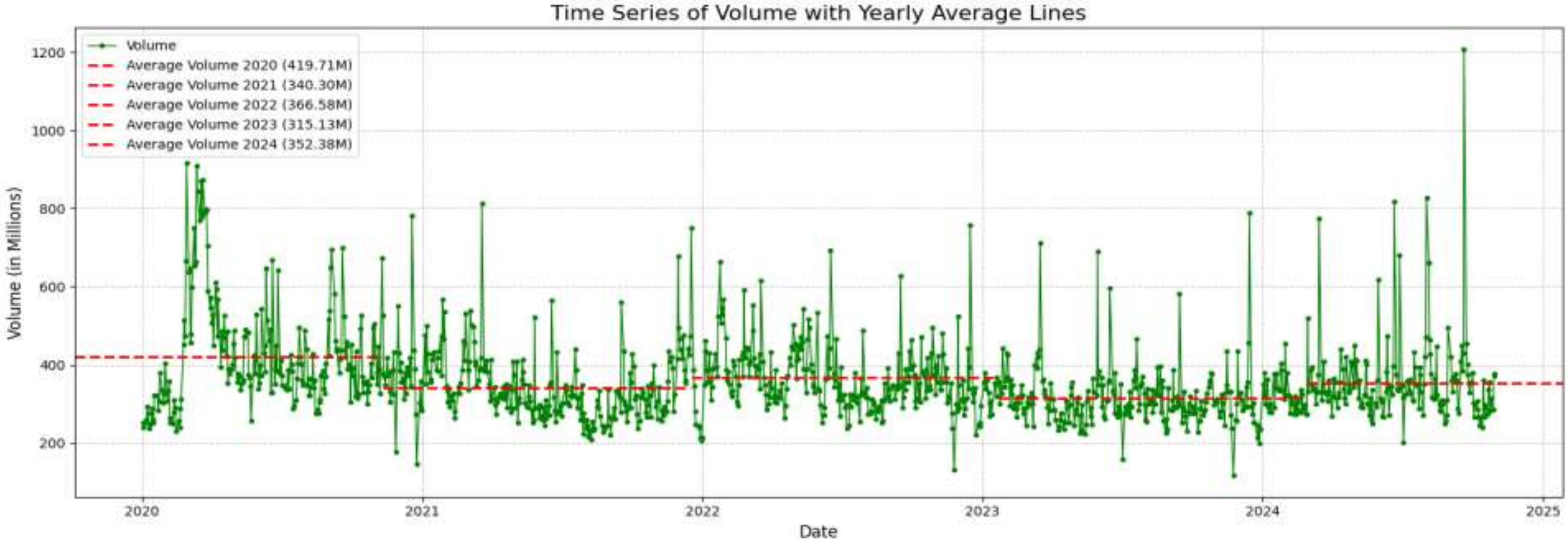
- **Point anomalies**: A data point is considered an anomaly because it differs substantially from the rest of the dataset, and no context is needed for the anomaly to become apparent.

- **Contextual anomalies**: A data point may appear normal on a global scale but becomes abnormal when viewed in its specific context



INDEX VOLUME

- The average trading volume of the Dow Jones index fluctuates every year.
- Significant higher/lower peaks in trading volume occur several times throughout the year.
- Above average peaks in volume are observed more often than below-average peaks.



Stony Brook University

HIGH/LOW VOLUME

- Total number of stocks or contracts traded during a given period.

- **High Volume Strategy**
- Breakout Strategy: High volume during a breakout confirms the strength and sustainability of the price movement.

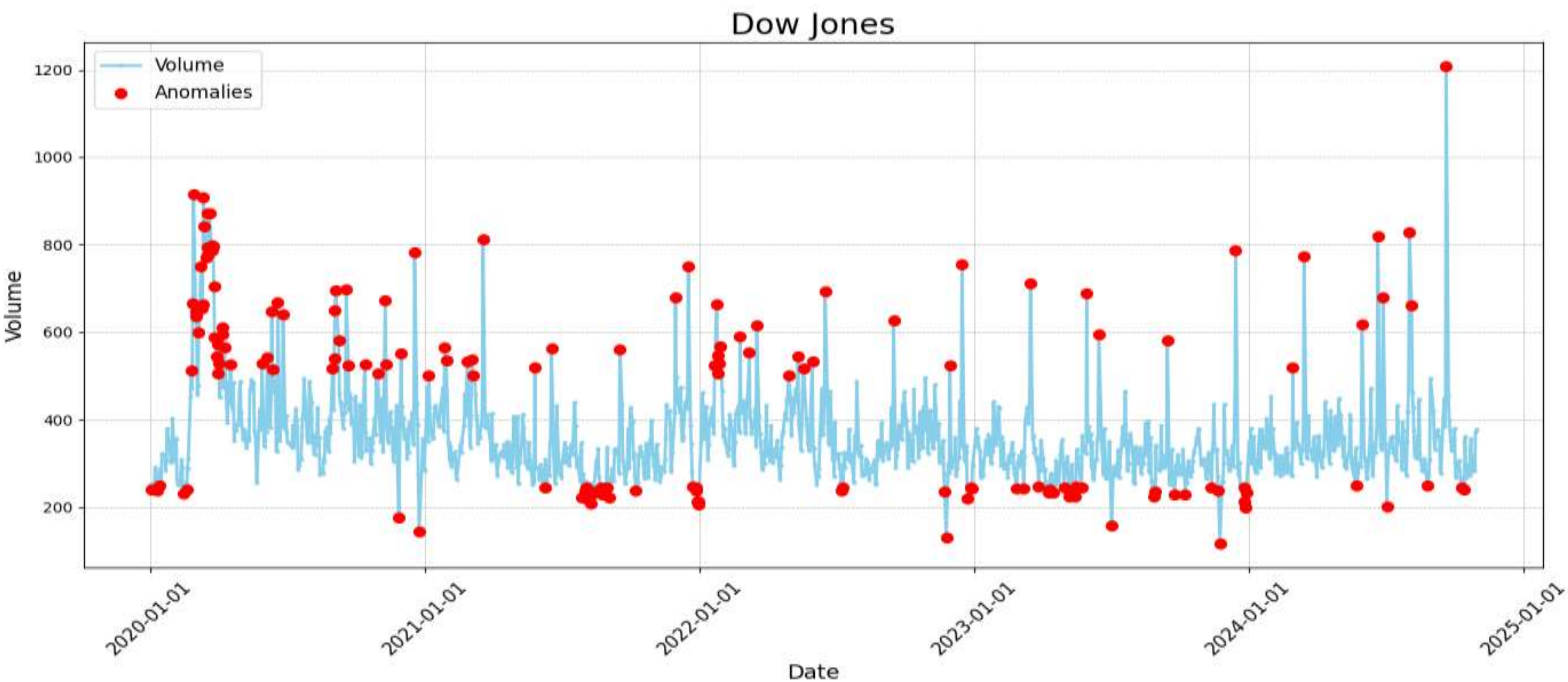
- Volume-Weighted Average Price (VWAP) Strategy: Average price at which most trades have occurred over a given period, weighted by volume.

- **Low Volume Strategy**
- Mean Reversion Strategy : Indicating a lack of interest or strong directional momentum.

- Support and Resistance Trading : Price tends to move within a defined range between established support and resistance levels.

VOLUME AND ANOMALIES

- Volumes above 500 million and below 250 million were flagged as anomaly.



ANOMALY DETETCTION METHOD

Rule-Based

- Quantiles
- Generalized Extreme Studentized Deviate

Unsupervised Models

- Isolation Forest
- One Class - SVM

Conformal Anomaly Detection

- Unsupervised base models

Regression

- Quantile Regression

RULE – BASED AD

- Calculate lower and upper thresholds.
- Use statistical methods/ test statistics.
- Mark the data points as anomalies if they fall outside the thresholds.

QUANTILE

- Quantiles or Value at Risk are cut points that divide the range of a probability distribution into continuous intervals with equal probabilities.

- Quantiles are defined mathematically as the inverse of the cumulative distribution function (CDF): $q(\alpha) = F^{-1}(\alpha)$

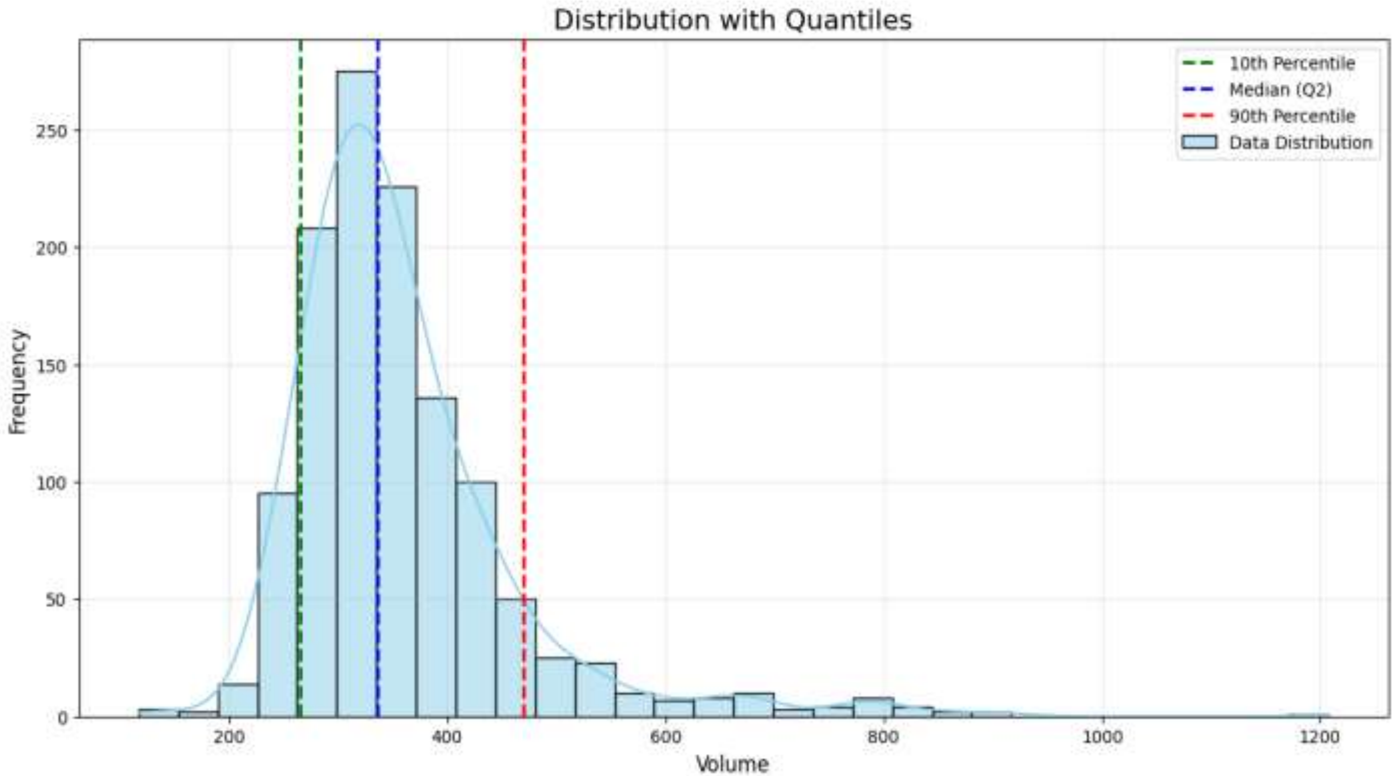
- Quantile $q(\alpha)$ is a solution to the equation : $F(x) = \alpha$ or $F(q(\alpha)) = \alpha$

- Quantile can be represented in terms of the complementary probability: $p(x)=1-\alpha$ or $p(q(\alpha))=1-\alpha$

Stony Brook University

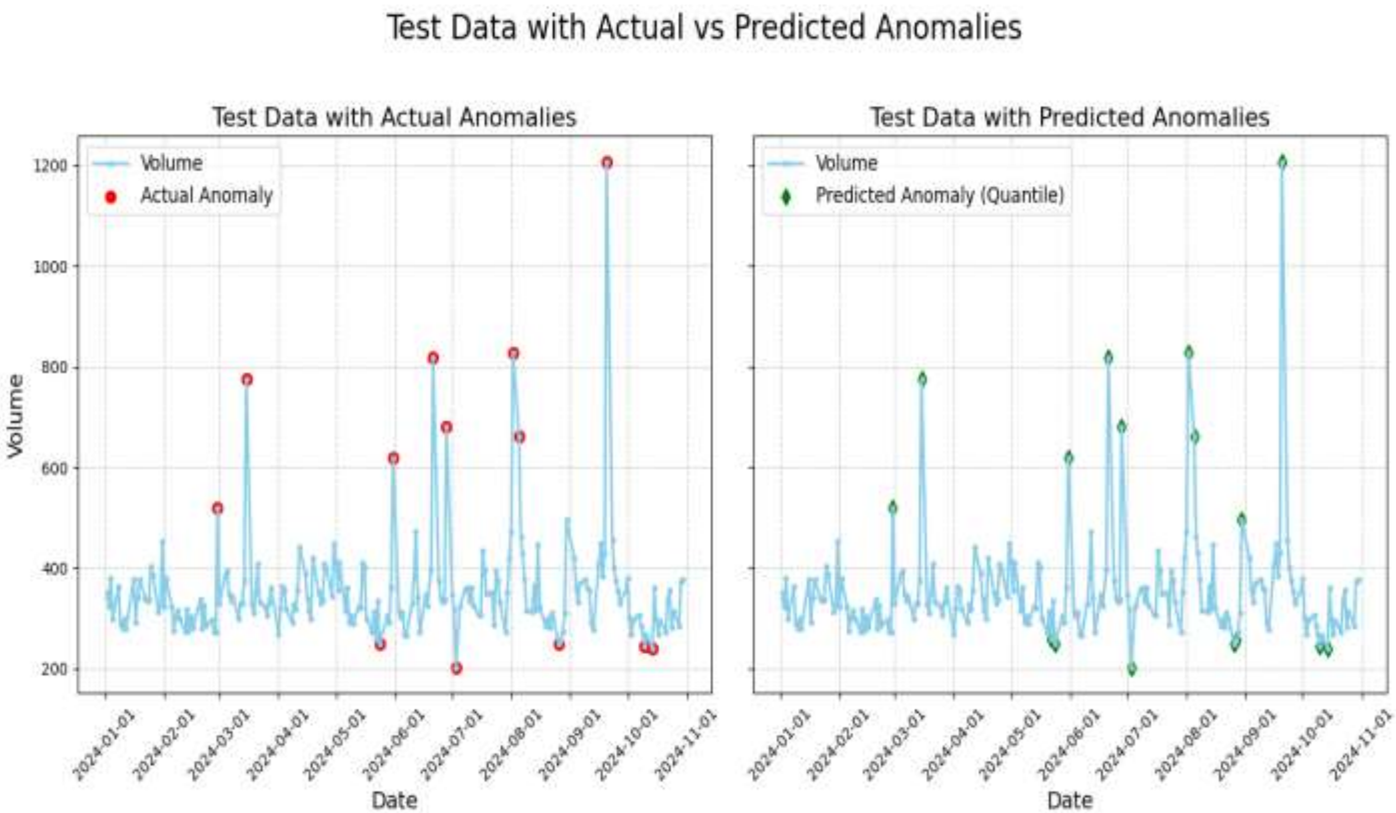
QUANTILE

- Upper threshold quantile (α = 0.9) and a lower threshold quantile (α = 0.1) were used.



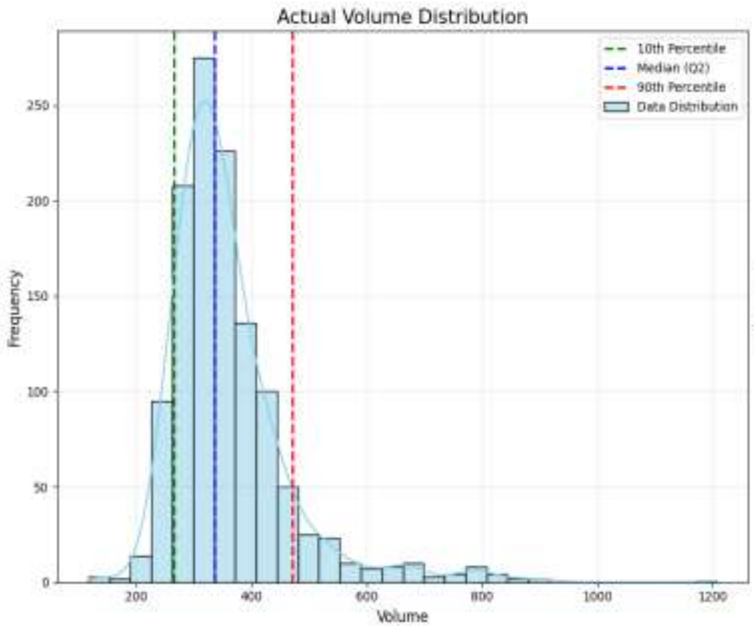
QUANTILE

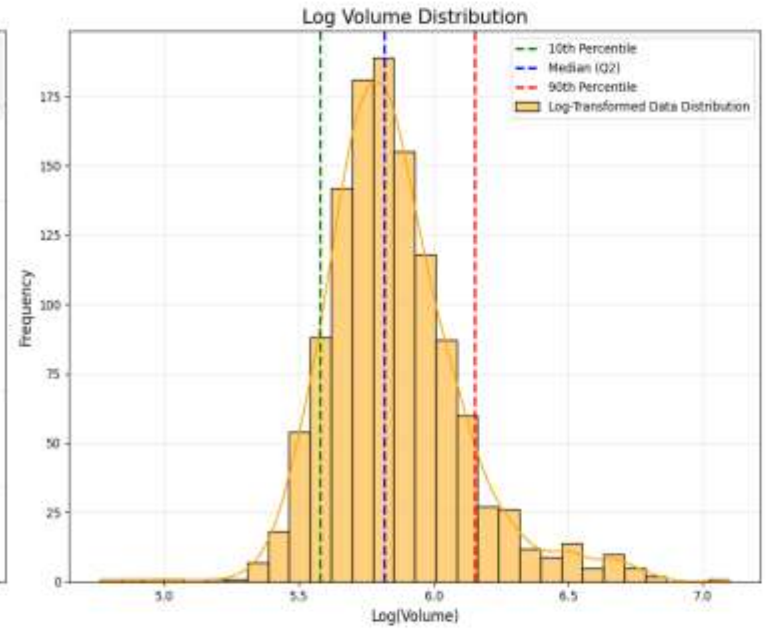
- As expected, Quantile AD can identify most of the outliers.



GENERALIZED EXTREME STUDENTIZED DEVIATE

- Iterative test identifies one outlier at a time and removes it, recalculating the test statistic and significance level for the remaining data in each step.
- Data should approximately follow a normal distribution
- Test statistic R_i exceeds the critical value λ_i, the observation corresponding to R_i is considered an outlier.



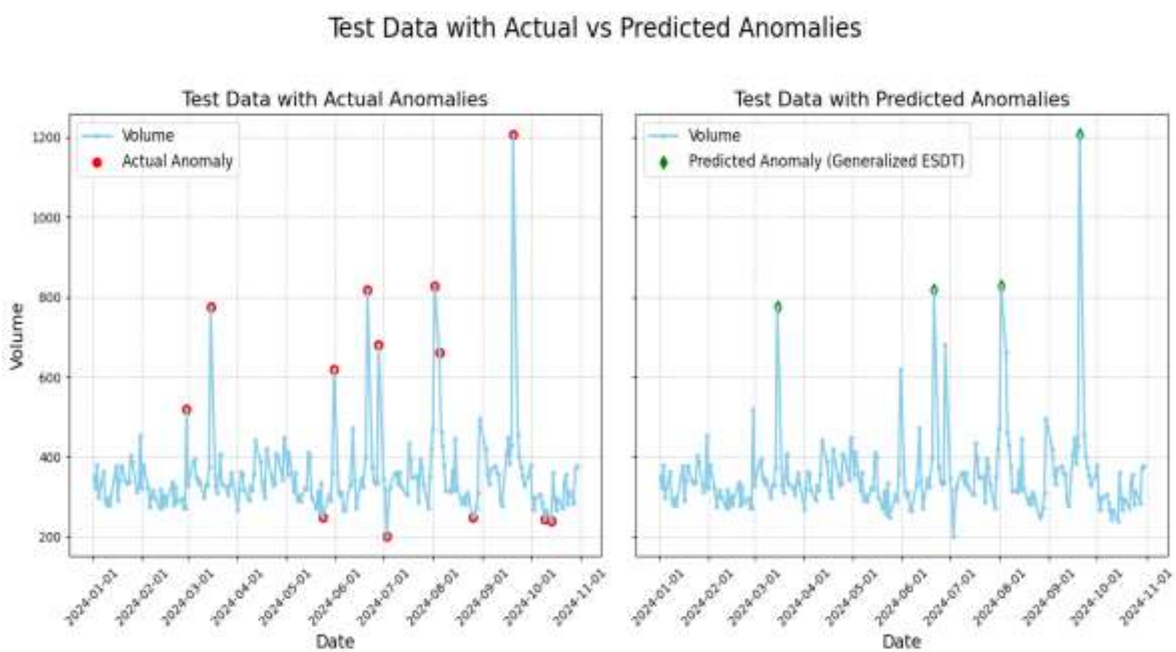


GENERALIZED EXTREME STUDENTIZED DEVIATE

- The test statistic R_i is computed as: $R_i = \frac{\max(|x_i - \bar{x}|)}{s}$
- x_i is the data point, $\bar{x}$ is the mean of data set, s is the standard deviation of the dataset
- **Tail Distribution** : The critical value λ_i is calculated as: $\lambda_i = \frac{(n-i)t_{p,n-i-1}}{\sqrt{\left(n-i-1+t_{p,n-i-1}^2\right)(n-i+1)}}$
- n is the total number of observations
- i is the index of the current suspected outlier (starting from 1),
- $t_{p,n-i-1}$ is the critical value from the $t-$distribution with $(n-i-1)$ degrees of freedom,
- $p = 1 - \frac{\alpha}{2(n-i+1)}$ is the adjusted significance level
- The procedure stops when $R_i \leq \lambda_i$, indicating that no further outliers exist in the data.

GENERALIZED EXTREME STUDENTIZED DEVIATE

- Generalized Extreme Studentized Deviate fails to detect lower anomalies, as it only works for the upper tail.



UNSUPERVISED AD

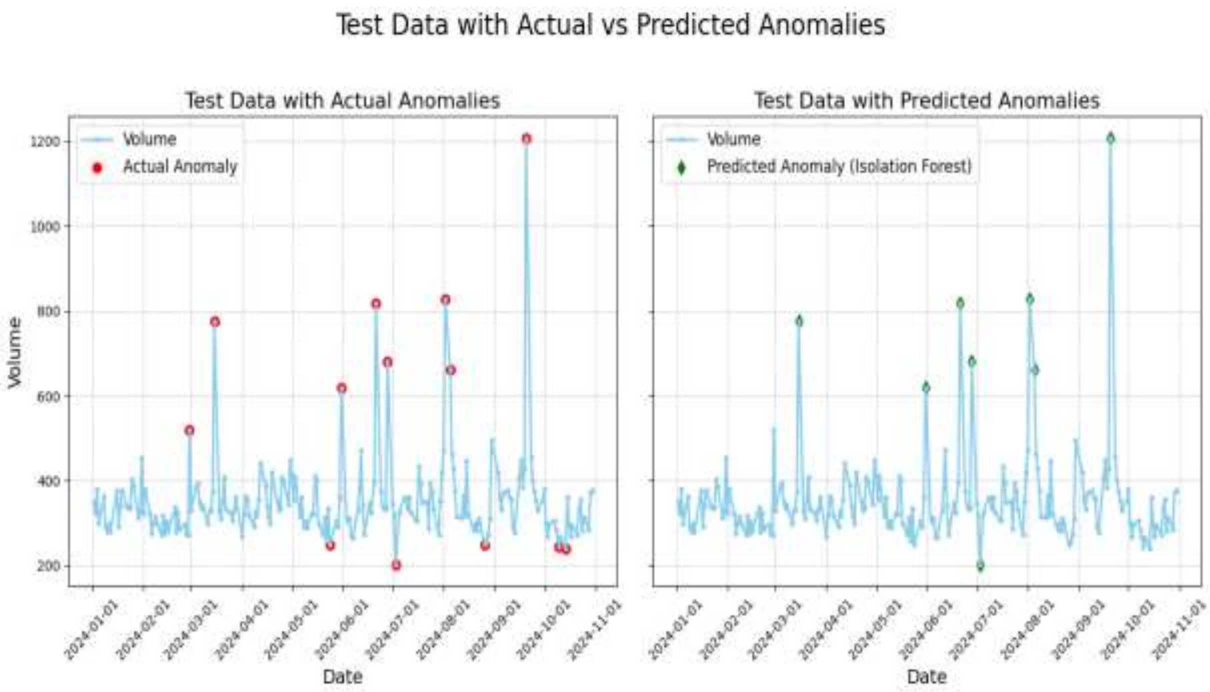
- Identify unusual patterns or anomalies in data without relying on labeled examples of anomalies.
- Computational efficiency for large datasets.
- Suitable for environments where normal behavior changes over time.
- Identifying deviations without requiring balanced datasets.

ISOLATION FOREST

- Anomaly detection using binary trees.
- **Assumption** : Anomalies are few and different from other data, they can be isolated using a few partitions.
- Use the training dataset to build a number of Isolation Trees (iTrees).
- For each data point in the test set:
- (a) Pass it through all the iTrees, counting the path length for each tree.
- (b) Assign an anomaly score to the instance.
- (c) Label the point as an anomaly if its score exceeds a predefined threshold, which depends on the domain.

ISOLATION FOREST

➢ Fewer outliers in lower anomalies.



Stony Brook University

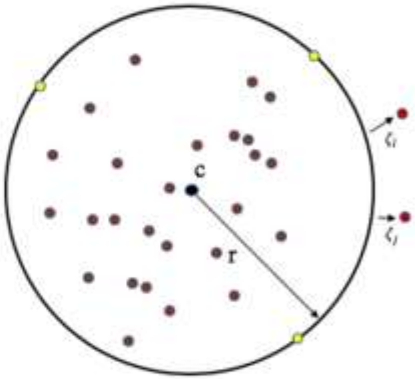
ONE CLASS - SUPPORT VECTOR MACHINE

- SVM-based one-class classification (OCC) relies on identifying the smallest hypersphere, with radius r and center c, consisting of all the data points.

- Objective is to find the center and radius such that all data points lie within or on the surface of this sphere

- Model trained on only normal data points

- Anomalies predicted if data points look different than trained data points.



The hypersphere containing the target data having center c and radius r. Objects on the boundary are support vectors, and two objects lie outside the boundary having slack greater than 0.

ONE CLASS - SUPPORT VECTOR MACHINE

- The original problem can be defined as :

$$\min_{r,c} \; r^2$$

$$\text{subject to} \quad ||\Phi(x_i) - c||^2 \leq r^2 \quad \forall i = 1,2,\dots,n$$

- $||\Phi(x_i) - c||^2$ is the squared distance between the data point and the center of the hypersphere.

- To allow for the presence of outliers, a more flexible formulation is given as:

$$\min_{r,c,\zeta} r^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\zeta_i$$

$$\text{subject to} \quad ||\Phi(x_i) - c||^2 \leq r^2 + \zeta_i \quad \forall i = 1,2,\dots,n$$

- ζ_i are the slack variables that allow some data points to lie outside the hypersphere.
- ν is number of outliers, n is total data points.

ONE CLASS - SUPPORT VECTOR MACHINE

- ***Karush-Kuhn-Tucker Conditions for Optimality***

$$c = \sum_{i=1}^{n} \alpha_i \Phi(x_i)$$

- Where the α_i's are the solution to the following optimization problem is:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i \kappa(x_i, x_i) - \sum_{i,j=1}^{n} \alpha_i \alpha_j \kappa(x_i, x_j)$$

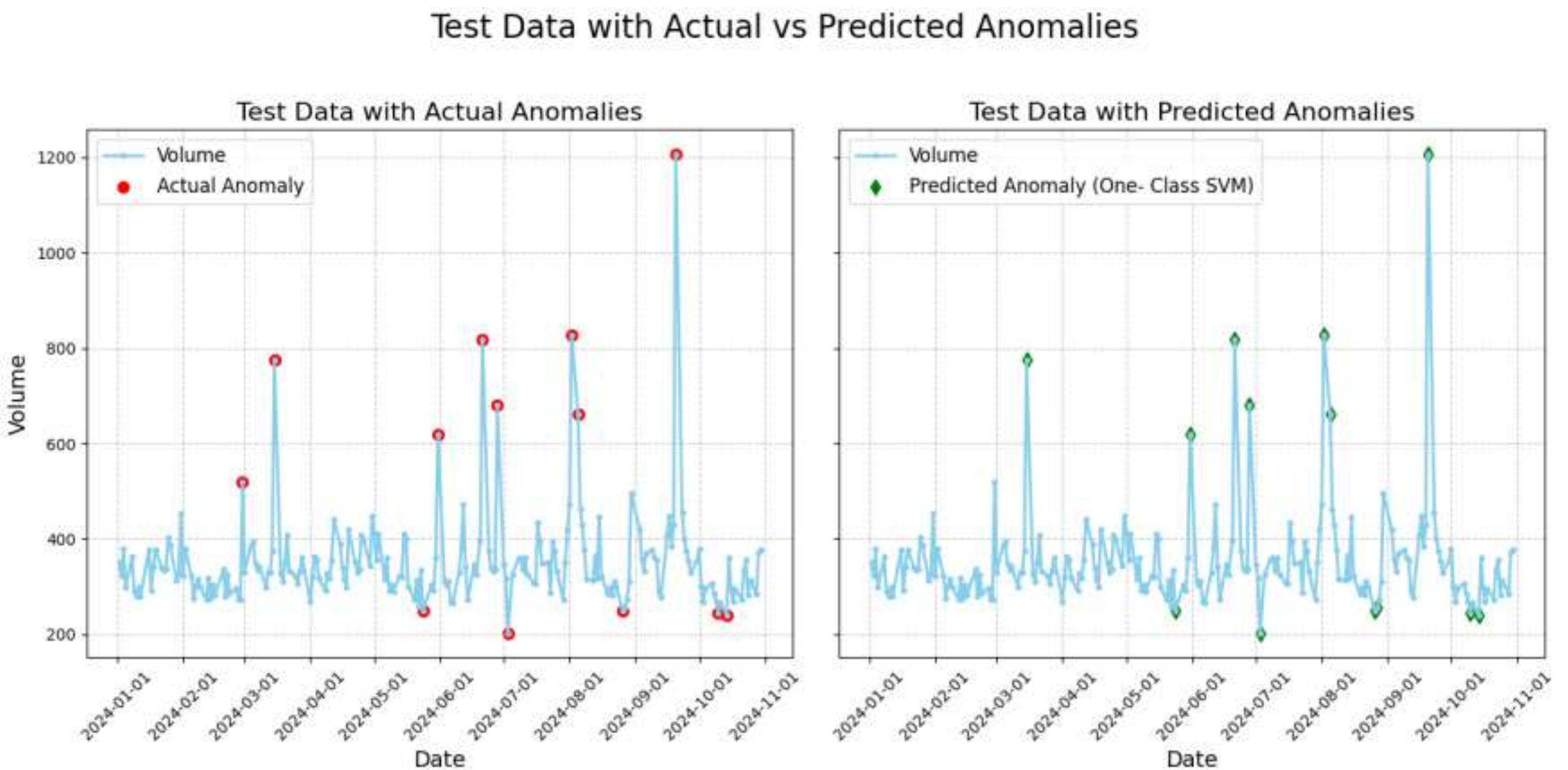
$$\sum_{i=1}^{n} \alpha_i = 1 \quad \text{and} \quad 0 \le \alpha_i \le \frac{1}{\nu n} \quad \text{for all} \quad i = 1,2,\dots,n$$

- Where α_i are the Lagrange multipliers associated with each data point.
- $\kappa(x_i, x_i)$ is the kernel function that computes the similarity between two points in the transformed feature space.

Stony Brook University

ONE CLASS - SUPPORT VECTOR MACHINE

- One class - SVM was trained using a radial basis function kernel and nu 0.15.



CONFORMAL PREDICTION

- Framework that provides a way to create prediction sets based on past data.
- These sets are associated with a confidence level, meaning that with a certain probability, the true outcome will lie within the predicted set.
- **Training Phase** : Train a machine learning model on a subset of the data to produce predictions.
- **Calibration Phase** : Use a calibration dataset (a hold-out set) to compute the nonconformity score for each instance. Compute a threshold such that the proportion of nonconformity scores below it matches the desired confidence level.
- **Prediction Phase** : For new data, compute the nonconformity score for potential outputs. Include outputs in the prediction set if their scores fall below the threshold.

Stony Brook University

CAD : CONFORMAL ANOMALY DETECTION

- CAD is extended to handle unsupervised anomaly detection, allowing us to identify data points that do not conform to the normal distribution of a dataset.

- The goal is to assign a statistical guarantee to the base model, ensuring control over the false positive rate.

- Assigns an anomaly score $S(x_i)$ to each data point X_i. Higher scores indicate a higher likelihood of being an outlier.

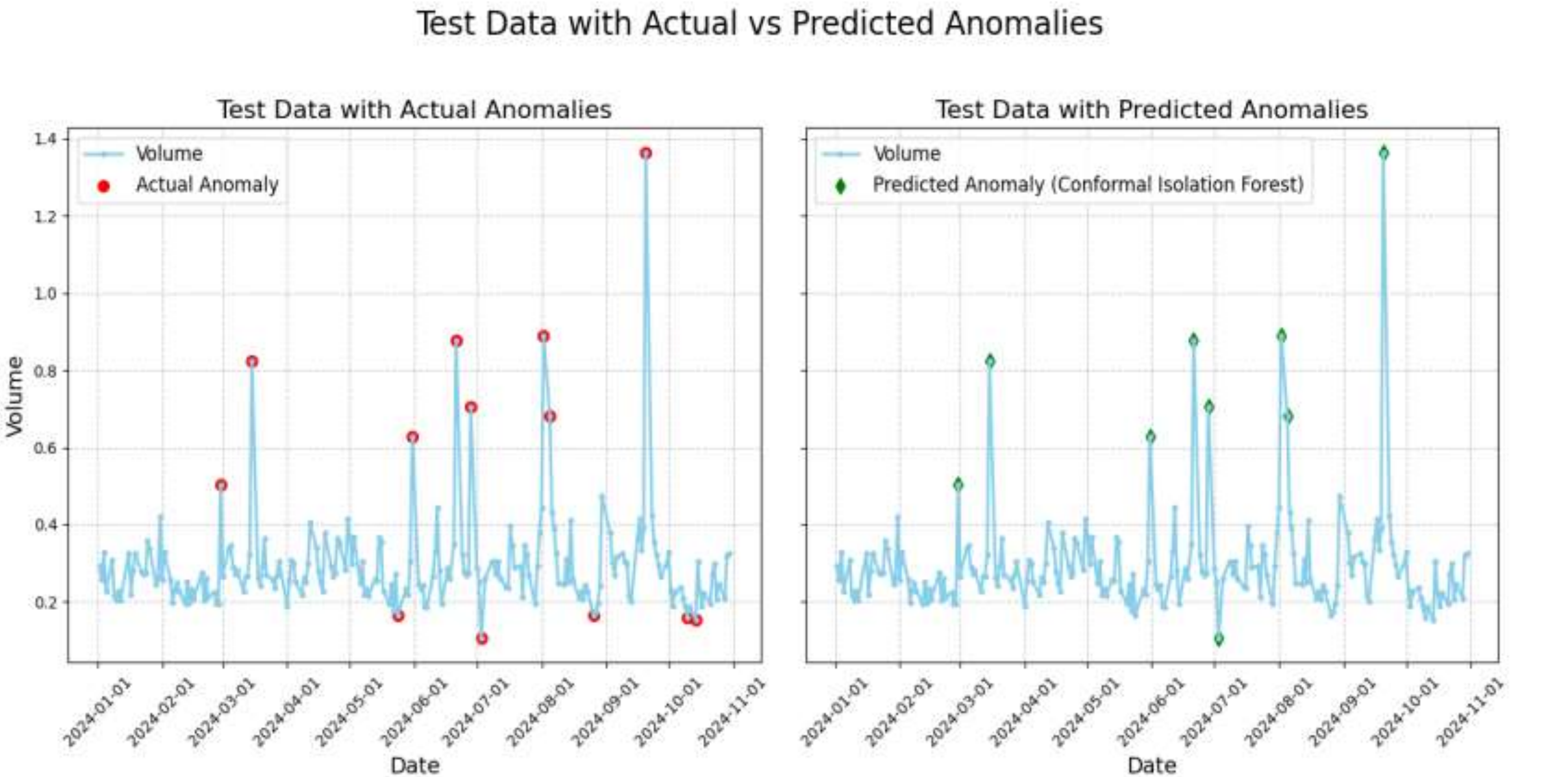
- For each example, X_i in the calibration dataset, compute the nonconformity score as $R_i = S(x_i)$, and store all nonconformity scores in a vector R.

- Next, compute the anomaly score threshold δ_α as the $(1-\alpha)(1+1/n_{calib})$-th empirical quantile of R.
- For a new test point X_{new}, the conformalized anomaly detector classifies it as:

$$C_\alpha = \{ \text{Normal, if } S(X_{new}) \le \delta_\alpha \text{, Anomaly, otherwise.}$$

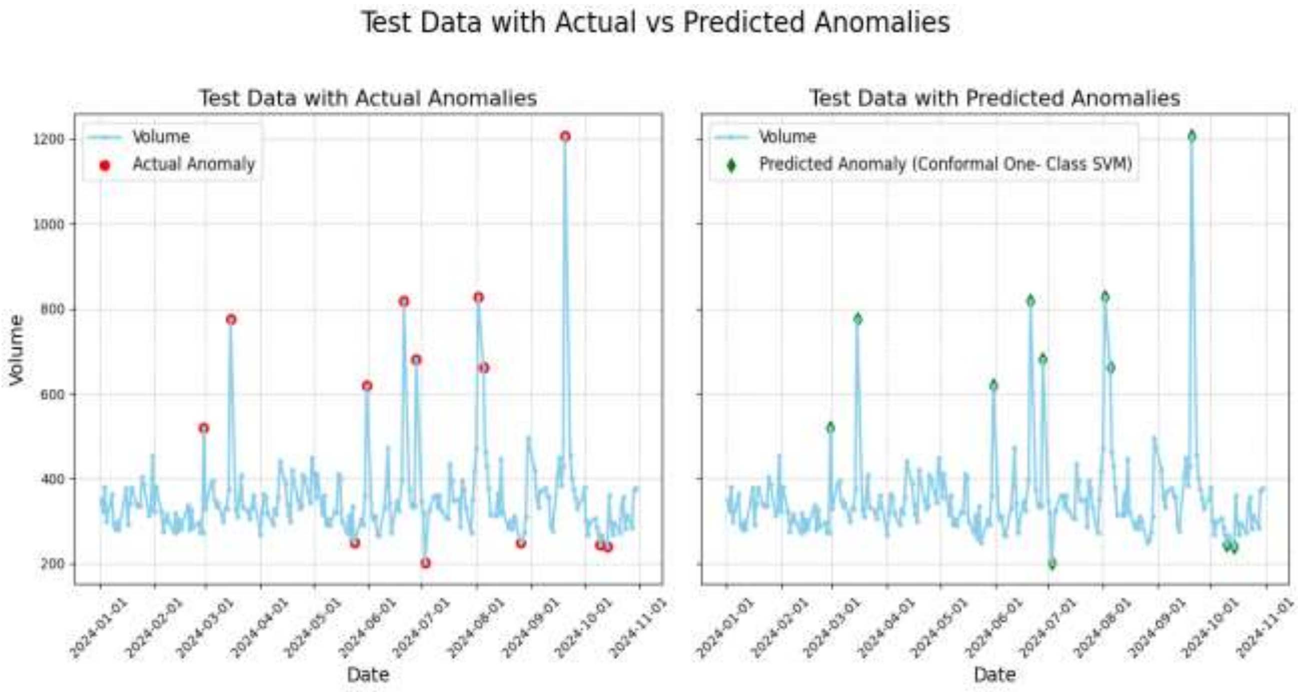
CAD : CONFORMAL ANOMALY DETECTION

- The same parameters were used in CAD as in the base model, except for the confidence level.



CAD : CONFORMAL ANOMALY DETECTION

- The same parameters were used in CAD as in the base model, except for the confidence level.

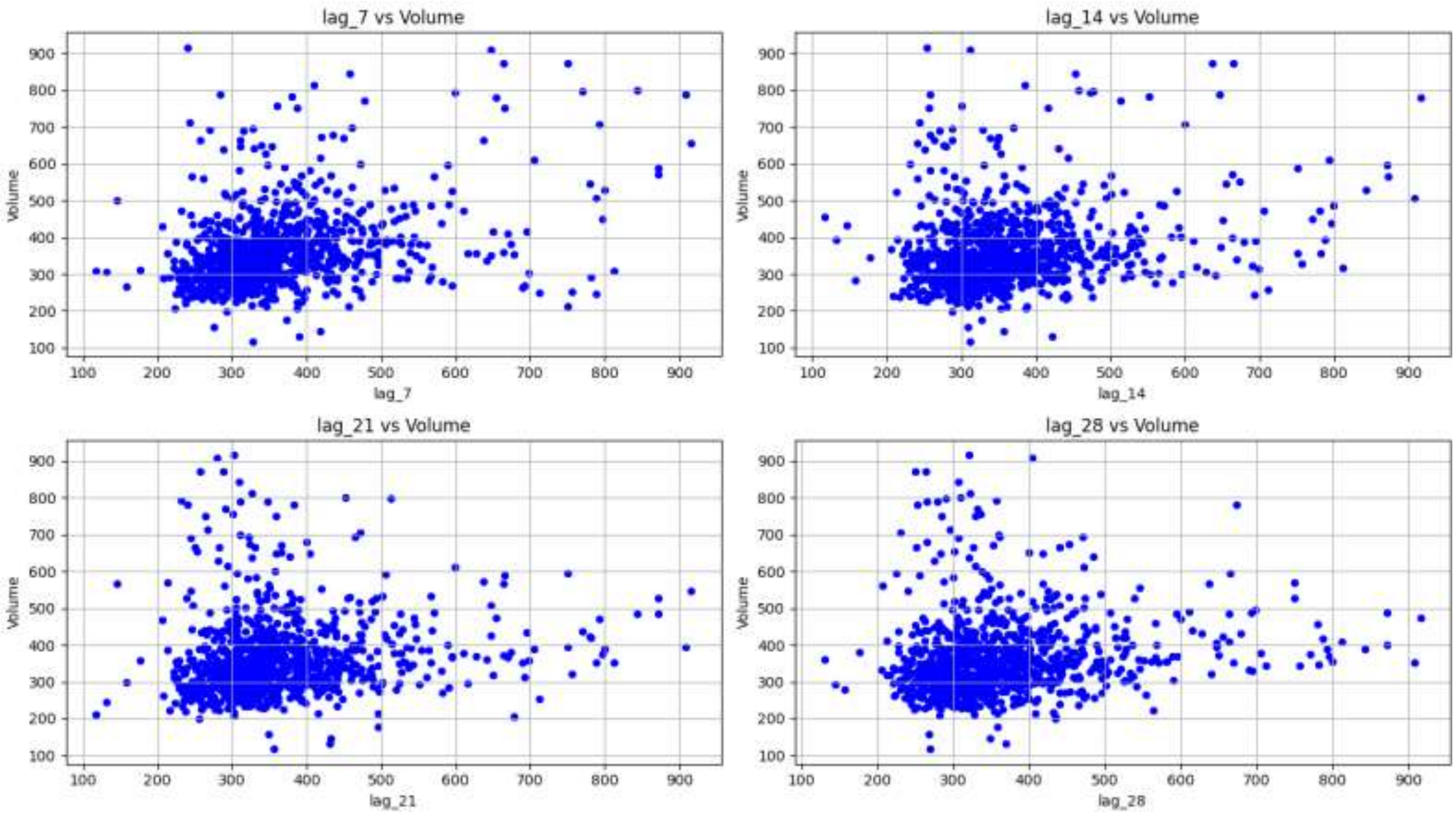


REGRESSION : QUANTILE REGRESSION

- Standard regression works with the mean of the distribution and works effectively with homoscedasticity data.
- When data is heteroscedasticity standard regression fails and the alternate model is quantile regression.
- Implement quantile regression where upper and lower quantile.
- **Quantile Regression** : Conditional quantiles of the target variable as a linear function of the feature variables.
- Data points above and below these quantile lines will be flagged as anomalies.

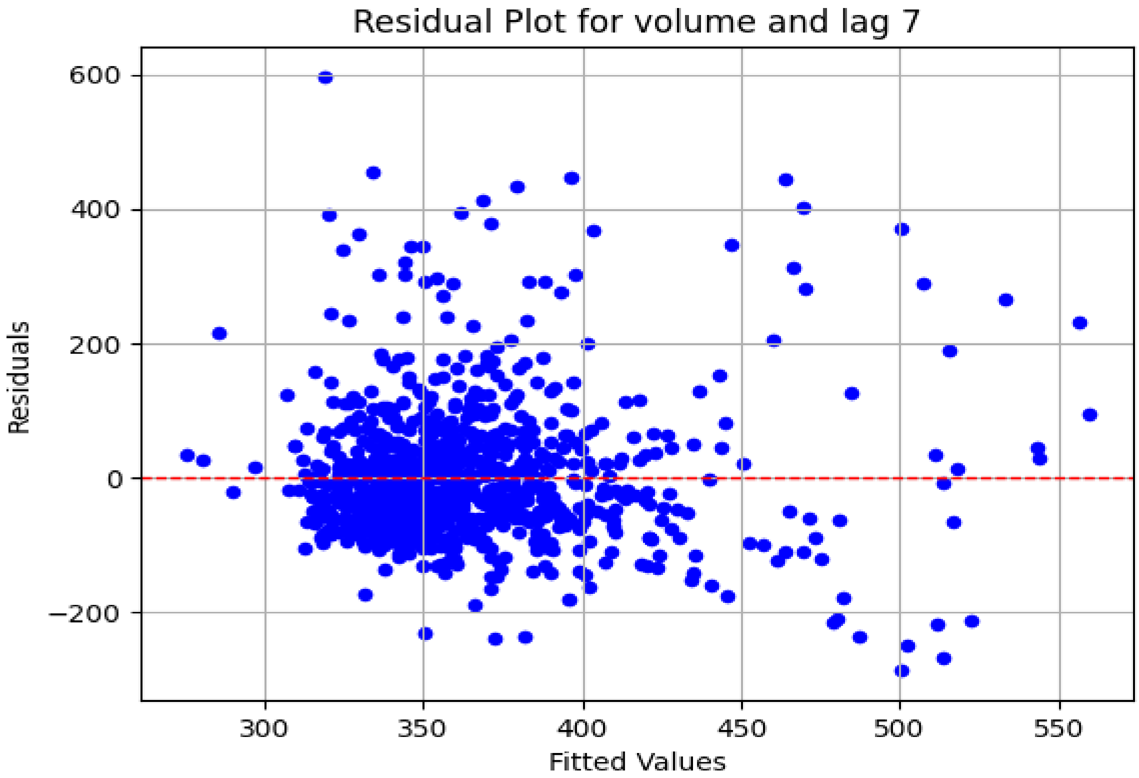
QUANTILE REGRESSION

- Non-linear relationship in different lags and volume and we've chosen lag 7.



QUANTILE REGRESSION

- Residuals are Heteroscedastic.



Stony Brook University

QUANTILE REGRESSION

- Standard Regression :

$$L(x,\theta) = \min_{\beta_0,\beta_1} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

- Quantile regression :

$$\text{Min } E\,[\alpha\,{*}\max\{0,L(x,\theta)\} + (1-\alpha){*}\max\{0,-L(x,\theta)\}]$$

QUANTILE REGRESSION

- Fit the 25th and 75th Quantiles:

$$Q_{25} = 25^{th} \text{ percentile}, \ Q_{75} = 75^{th} \text{ percentile}$$

- Compute the Interquartile Range (IQR):

$$IQR = Q_{75} - Q_{25}$$

- Set the threshold parameter k: We define k as ½ constant that controls the threshold range.

- Calculate the Lower Threshold F1:

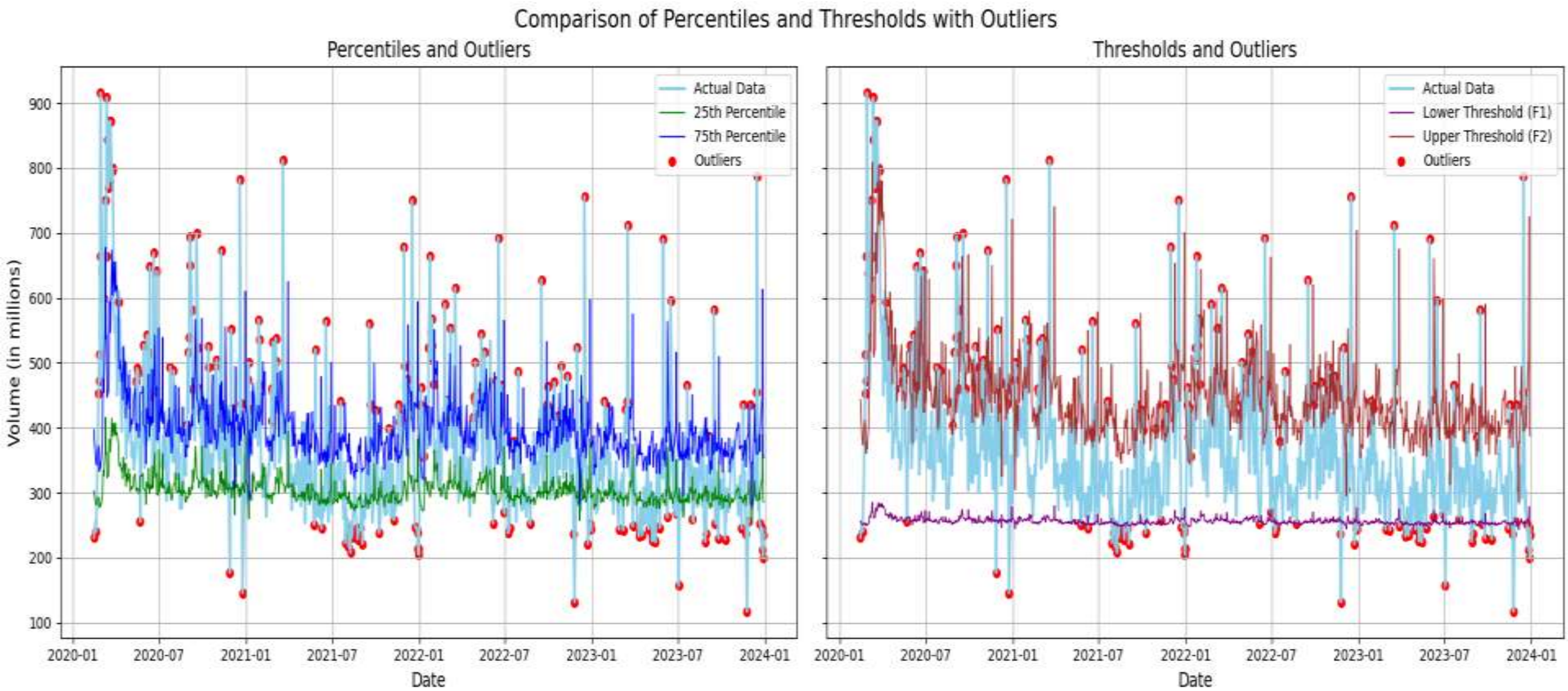
$$F1 = Q_{25} - k \times IQR$$

- Calculate the Upper Threshold F2:

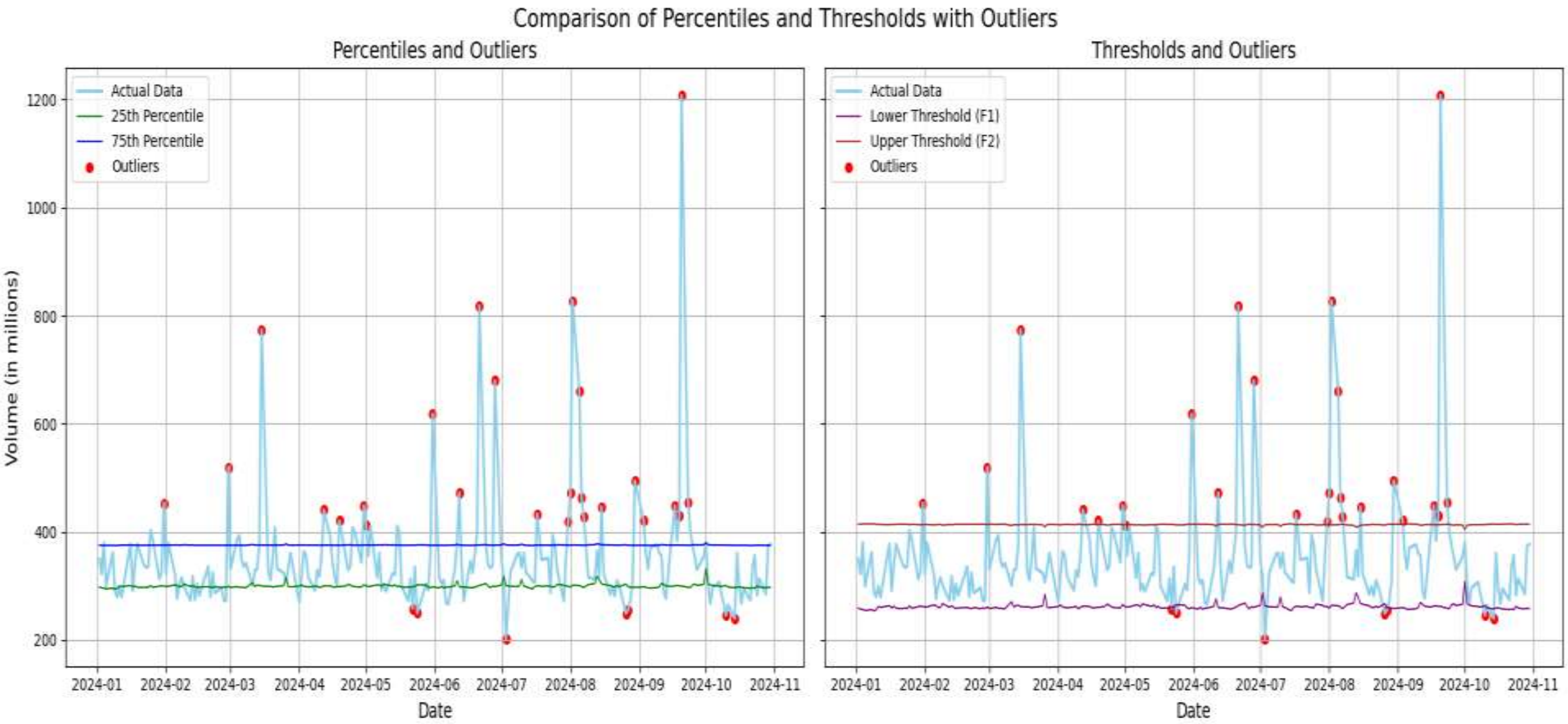
$$F2 = Q_{75} + k \times IQR$$

Stony Brook University

QUANTILE REGRESSION



QUANTILE REGRESSION

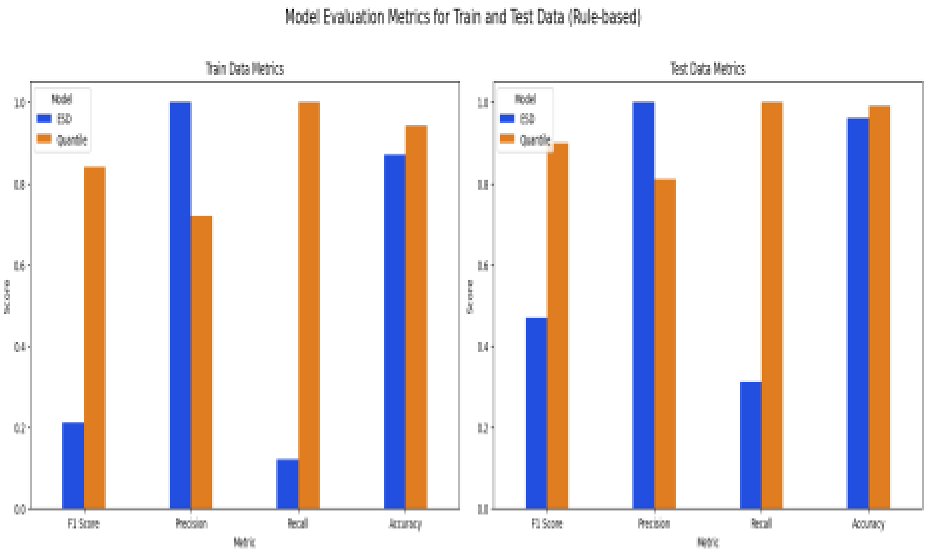


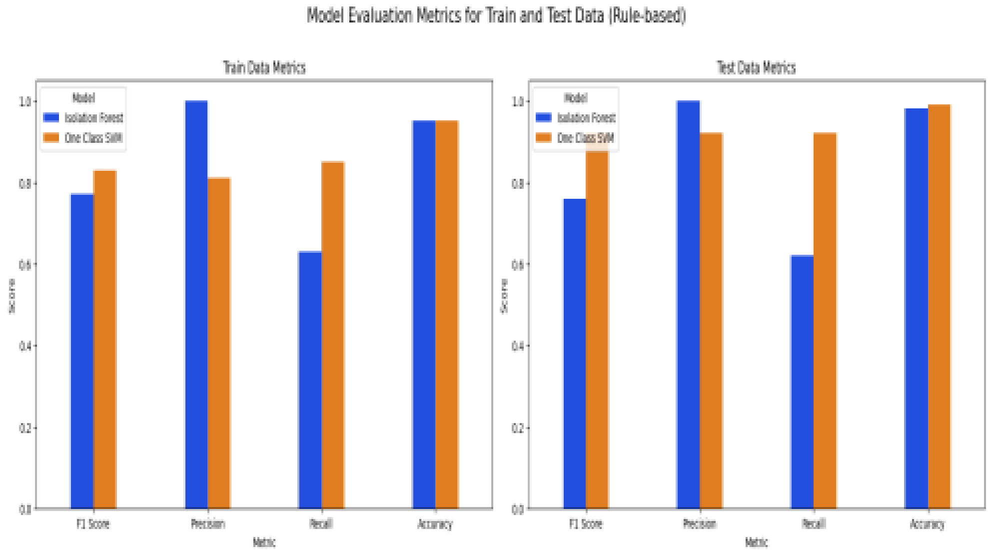
METRICS AND PERFORMANCE

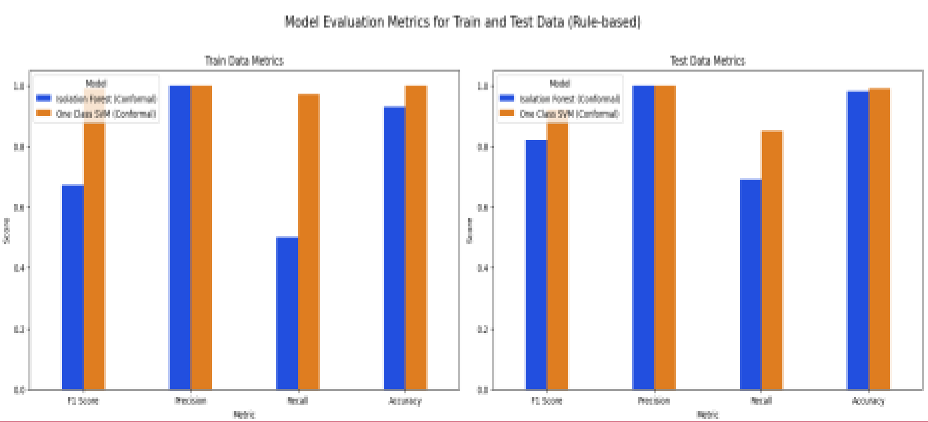
- One class - SVM outperform all other models with F1 score 0.99 on training data and 0.92 on test data.

Model	Type	Data	F1 Score	Precision	Recall	Accuracy
ESD	Rule-based	Test	0.47	1.00	0.31	0.96
ESD	Rule-based	Train	0.21	1.00	0.12	0.87
Quantile	Rule-based	Test	0.90	0.81	1.00	0.99
Quantile	Rule-based	Train	0.84	0.72	1.00	0.94
Isolation Forest	Unsupervised	Test	0.76	1.00	0.62	0.98
Isolation Forest	Unsupervised	Train	0.77	1.00	0.63	0.95
Isolation Forest (Conformal)	Conformal	Test	0.82	1.00	0.69	0.98
Isolation Forest (Conformal)	Conformal	Train	0.67	1.00	0.50	0.93
One Class SVM	Unsupervised	Test	0.92	0.92	0.92	0.99
One Class SVM	Unsupervised	Train	0.83	0.81	0.85	0.95
One Class SVM (Conformal)	Conformal	Test	0.92	1.00	0.85	0.99
One Class SVM (Conformal)	Conformal	Train	0.99	1.00	0.97	1.00
Quantile Regression	Regression	Test	0.58	0.41	1.00	0.91
Quantile Regression	Regression	Train	0.72	0.60	0.91	0.90

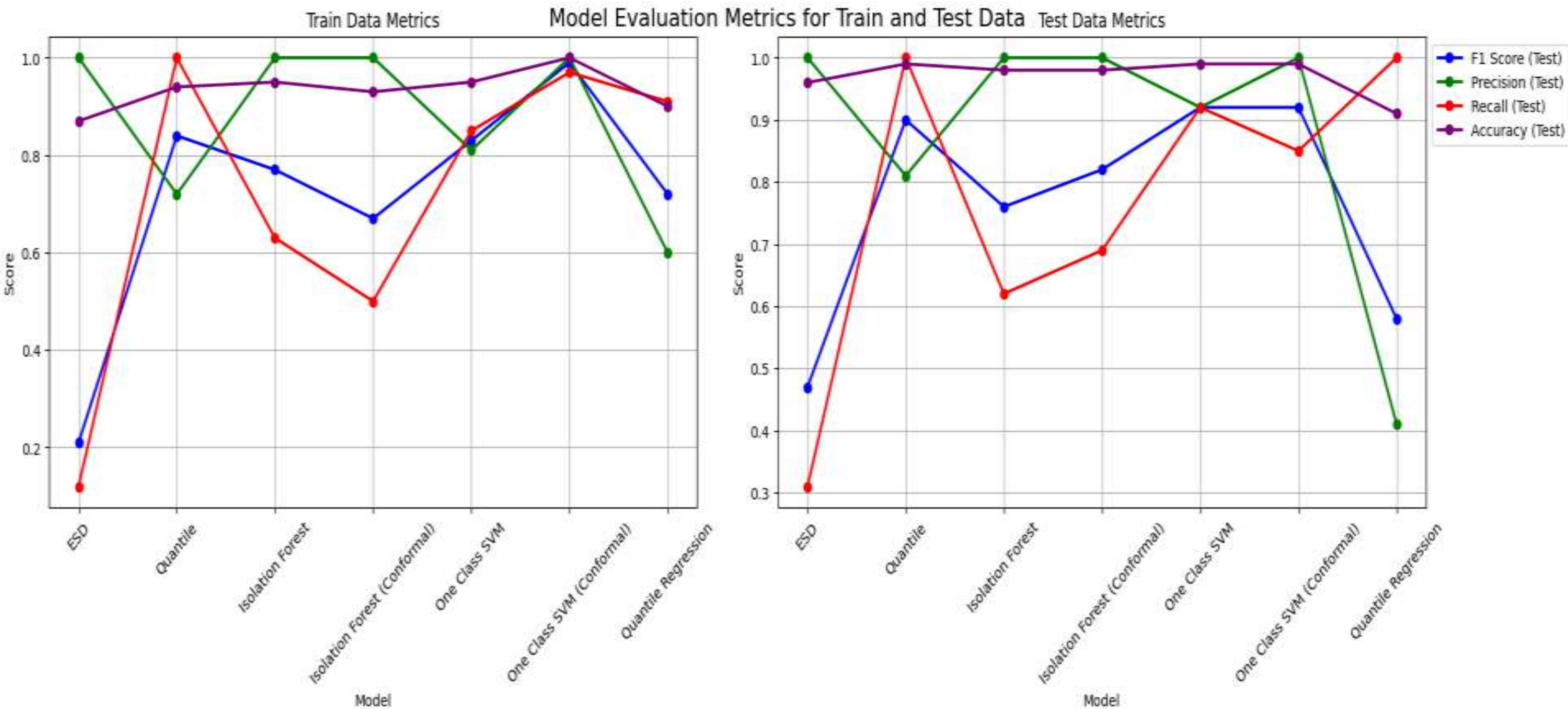
METRICS AND PERFORMANCE







METRICS AND PERFORMANCE



CONCLUSION

- The generalized extreme studentized deviation (ESD) test performs poorly.
- The quantile method exhibits strong performance, as it effectively identifies anomalies.
- Isolation forest and its conformal variants do not perform well in this context.
- One-class SVM and its conformal variants outperform all other models.
- Quantile regression also performs well.

Stony Brook University

REFERENCES

- Chandola, Varun Banerjee, Arindam Kumar, Vipin. (2009). Anomaly Detection: A Survey. ACM Comput. Surv.. 41. 10.1145/1541880.1541882.
- Barnett, V., Lewis, T. (1978). Outliers in statistical data. Wiley. https://archive.org/details/outliersinstatis0000barn.
- F. T. Liu, K. M. Ting and Z. -H. Zhou, ”Isolation Forest,” 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 2008, pp. 413-422, doi: 10.1109/ICDM.2008.17.
- Tax, DMJ. (2001). One-class classification; concept-learning in the absence of counter-examples. [Dissertation (TU Delft), Delft University of Technology].
- Angelopoulos, Anastasios; Bates, Stephen (2021). ”A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification”. arXiv:2107.07511.
- @inproceedingsmendil2023puncc,title=PUNCC: a Python Library for Predictive Uncertainty Calibration and Conformalization,author=Mendil, Mouhcine and Mossina, Luca and Vigouroux, David,booktitle=Conformal and Probabilistic Prediction with Applications, pages=582–601,year=2023,organization=PMLR
- @inproceedingsmendil2022robust,title=Robust Gas Demand Forecasting With Conformal Prediction,author=Mendil, Mouhcine and Mossina, Luca and Nabhan, Marc and Pasini, Kevin, booktitle=Conformal and Probabilistic Prediction with Applications,pages=169–187, year=2022,organization=PMLR

Stony Brook University

REFERENCES

- Koenker, Roger (2005). Quantile Regression. Cambridge University Press. pp. 146–7. ISBN 978-0-521-60827-5.
- https://adtk.readthedocs.io/en/stable/quickstart.html
- Powers, David M. W. (2011). ”Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness Correlation”. Journal of Machine Learning Technologies. 2 (1): 37–63. S2CID 55767944.
- https://github.com/valeman/awesome-conformal-prediction
- https://www.mathworks.com/help/stats/outlier-detection-using-quantileregression. html
- https://medium.com/aimonks/univariate-time-series-anomaly-detection-capturingthe-unusual-a5cc5ee24462.
- https://www.analyticsvidhya.com/blog/2022/05/an-end-to-end-guide-onanomaly detection/
- https://medium.com/@ngiengkianyew/quantile-regression-62a398186f7d

THANK YOU !