Anomaly Detection

Definition

A rare occurrence or event that doesn't fit into the pattern, and, therefore, seems suspicious.

Types of Anomaly

- Global Outliers: When a change occurs far outside the overall historical value ranges.
- Contextual Outliers: When a change occurs far outside the historical changes with the same context.
- Collective Outliers: When a change occurs far outside and contrary to the normal behavior.

Anomaly Detection ML Methods

- Local Outlier Factor (LOF): Calculate local density of an entry. Lower than usual density means an outlier.
- K-nearest Neighbors (kNN): Determine normal and abnormal values manually, then calculate if the new entry has more nearest neighbors in normal or abnormal values.
- Support Vector Machine (SVM): Determine a divider between data points, then classify the new entry based on its location relative to the divider.
- DBSCAN: Determine local large density groups of your data, then classify the entries that don't belong to any of these groups as abnormal.
- Autoencoders: Encode the data to reduce dimensionality, then decode to find outliers.
- Bayesian networks: Take an event that occurred and predict the likelihood that any one of several possible known causes was the contributing factor.

Our Current Anomaly Detection Algorithm:

Market Confidence Level:

- Read historical data for the daily/hourly changes in the Bitcoin price.
- Given k number of maximum changes, calculate the mean and standard deviation.
- Use the mean and standard deviation values to determine the expected minimum and maximum ranges.
- Given a new entry, calculate the change from the previous entry.
- Calculate the Market Confidence Level based on the latest change and the minimum and maximum ranges defined earlier.

- Lower than minimum is equated to %100 confidence in the market.
- More than maximum is equated to %0 confidence in the market.
- Any value in between is calculated using abs(current change - minimum)/(maximum - minimum)*100

Model Confidence Level:

- Predict the next value using our model.
- When the actual next entry becomes available, calculate the difference between what was predicted and what the actual value is.
- abs(predicted - actual)/actual*100

How to Advance Our Anomaly Detection Algorithm:

- Investigate further the Anomaly Detection ML methods introduced earlier.
- Determine which ones are the best fit for our purpose.
- Integrate the possible candidates to our code for Anomaly Detection.
- If several methods will be used, determine coefficients for weighing up their effects on the final decision.