

# Outlier Detection Methods for Uncovering of Critical Events in Historical Phasor Measurement Records

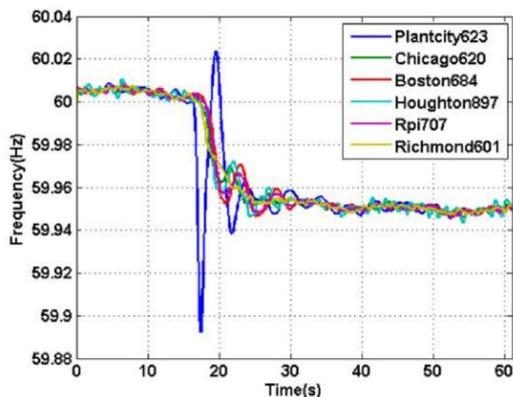
André Kummerow\*, Steffen Nicolai, and Peter Bretschneider

Department of Energy, Advanced System Technology (AST) Branch of Fraunhofer IOSB, Illmenau, Germany

**Abstract.** The scope of this survey is the uncovering of potential critical events from mixed PMU data sets. An unsupervised procedure is introduced with the use of different outlier detection methods. For that, different techniques for signal analysis are used to generate features in time and frequency domain as well as linear and non-linear dimension reduction techniques. That approach enables the exploration of critical grid dynamics in power systems without prior knowledge about existing failure patterns. Furthermore new failure patterns can be extracted for the creation of training data sets used for online detection algorithms.

## 1 Background and main objectives

Phasor measurement units (PMU) are widely used in largescale power grids to enhance static SCADA information with high-resolution (up to 60 f.p.s.) phasor and frequency data. PMUs are interconnected into wide area monitoring systems (WAMS) and increase situational awareness of modern control centres by tracking dynamic events (e.g. abrupt changes, critical oscillations) and perform real-time analytics – see Fig. 1. The detection and classification of grid disturbances (e.g. generator loss, load loss, line outages, system oscillations) from real field measurements is currently a main application field of PMUs in transmission and distribution power systems. [1–5]



**Fig. 1.** Example of a generator loss recorder by PMUs from [4].

In literature [6–9] several concepts have been proposed for the identification of grid disturbances from PMU data. Several methods from multivariate statistics (e.g. Principle Components Analysis or K-

Means clustering) are combined with techniques from frequency transform (e.g. Wavelet Decomposition, Stockwell Transform) and modal analysis (e.g. Prony Analysis, Matrix Pencil Method, Empirical Mode Decomposition). Those studies heavily rely on sufficient disturbance samples for different failure types. Usually, the training samples are generated from dynamic simulations or from the analysis of disturbance record files.

This survey presents a new method for the uncovering of arbitrary grid disturbances from historical PMU data sets when no prior knowledge about grid dynamics or critical events is available. In contrast to previous works this approach is developed for the analysis of mixed and unbalanced PMU data sets, where only a small number of samples is assumed to be critical. The automated analysis procedure deals with data sets from single PMU sensors and uncovers potential failure patterns (e.g. voltage sags, oscillations, frequency drops) from real field measurements. Thereby grid dynamics and disturbances can be explored as well as potential relationships between critical events.

The general concept for the use of outlier detection methods to extract disturbances from mixed PMU data sets is shown in part 2. Additionally, some theoretical background is provided for the most important learning algorithms within this survey. Part 3 presents and discusses the main results from the extraction of time and frequency features as well as the different outlier detection methods.

## 2 General concept and outlier detection methods

The uncovering of critical disturbances from mixed and imbalanced PMU data sets can be addressed to the field of score-based, unsupervised outlier detection.

\* Corresponding author: andre.kummerow@iosb-ast.fraunhofer.de

For this task several algorithms exist to identify data which differs from expectations or majority behaviour. Successful applications include e.g. fraud or thief detection. The available techniques can be distinguished roughly into probabilistic, distance-based, reconstruction-based, domain-based and information-theoretic approaches [10, 11]. The general concept for PMU outlier detection is shown in Fig. 2.

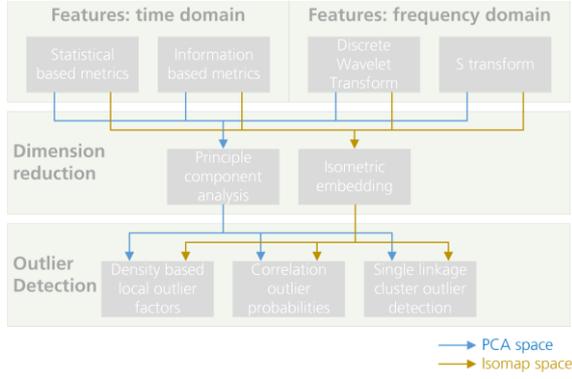


Fig. 2. Three-phase concept for PMU outlier detection.

Table 1. Comparison of outlier detection methods.

Method	Main principle	Metric used as outlier score
LOF	local densities of data points and its neighbourhood	local outlier factor
COP	deviation within local correlation using robust PCA	correlation outlier score
SiLiOd	hierarchical clustering using shortest distances	path lengths to final cluster

As a preprocessing step the historical data set is split into PMU samples of equal length. For each sample features are extracted in time and frequency domain using statistical and information based metrics as well as coefficients from Discrete Wavelet Transformation (DWT) and Stockwell transformation or S transform [12–14]. The features have to be designed carefully to capture the signal behaviour in presence of critical events. In a second step the feature space is reduced with linear and nonlinear dimension reduction techniques. This increases the performance of the outlier detection algorithms and redundant features can be eliminated. To preserve global data structures Principle Component Analysis (PCA) and Isometric Embedding (Isomap) are used. In the last step the outlier detection algorithms compute an outlier score for each PMU sample. Within this survey three methods are used: Local Outlier Factors (LOF), Correlation Outlier Probabilities (COP) and Single Linkage Outlier Detection (SiLiOd). All of those methods generate different metrics as outlier scores. A short summary is given in Table I. As an unsupervised approach the number of outliers to be extracted has to be defined in advance. This can be seen as the amount of expected contamination of the PMU dataset and has to be selected by trial and error.

The first method is a well-established technique for outlier detection and calculates an outlier factor for each data point [15, 16]. The local densities of the data points are compared with the densities of its neighbours.

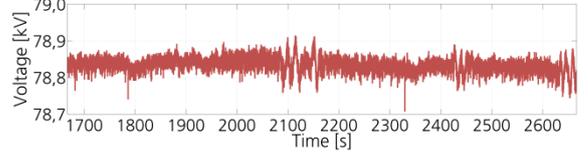


Fig. 3. Voltage magnitude raw signal.

Using the LOF method, outliers are treated as data points with high differences between the density of that point and its surroundings, so that  $LOF \gg 1$ . In contrast to the previous approach, COP (correlation outlier probabilities) is a subspace-based outlier detection technique using PCA as local correlation model [17]. Outliers are defined as data points that significantly differ from the local correlation within a defined neighbourhood. In addition to the two above mentioned techniques an own implementation of a cluster-based outlier detection algorithms is applied using hierarchical agglomerative clustering (HAC). At each iteration step the two clusters are merged which are closest to each other. This procedure is repeated until one final cluster is left. In single linkage clustering the similarity between two clusters  $c_i$  and  $c_j$  is defined as the distance  $D$  between the two closest cluster members  $x_i$  and  $x_j$  – see (1):

$$D(c_i, c_j) = \min(|x_i - x_j|) \quad (1)$$

The generated cluster tree can be seen as a bipartite graph  $G$  with vertices  $V$  and edges  $E$ . From that, the corresponding adjacency matrix  $A$  is computed according to (2):

$$A = [a_{i,j}] \text{ with } a_{i,j} = \begin{cases} 1 & \text{when } (i,j) \in E \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

That adjacency matrix can be used to detect outliers by calculating the shortest paths or minimum number of edges between the data points and the final cluster group. Those data points with short path lengths are treated as outliers. This method is called Single-Linkage Based Outlier Detection (SiLiOd). Further literature for agglomerative clustering methods and single linkage clustering can be found in [18].

## 3 Analysis results and discussion

### 3.1 Experimental setup

This study uses a record from real measured PMU data which includes voltage magnitudes from different PMU sensors [19]. Fig. 3 shows an excerpt from the measurements of one PMU sensor. The reporting rate

is 30 frames per second. In a pre-processing step the PMU raw data is filtered for high-frequency noise and normalized. Table II gives an overview of the main parameters and specifications for the used methods. The general calculations and discrete Wavelet decompositions are done in [20]. For the S Transform an open-source code is used [13]. The Isometric embedding (Isomap) is taken from an open-source Java based machine learning library [21].

TABLE I. CHOSEN PARAMETERS FOR OUTLIER DETECTION

Parameter	Value
Number of outliers	15
S Transform max. frequency	1 Hz
DWT decomposition level	12
DWT wavelet function	db5
Isomap number of neighbours	50
SiLiOd distance function	Mahalanobis
LOF / COP distance function	Euclidean
LOF / COP number of neighbours	50

The Local Outlier Factor (LOF) and Correlation Outlier Probabilities (COP) method is taken from the Java based ELKI (Environment for Developing KDD-Applications Supported by Index-Structures) data mining framework [22].

### 3.2 Feature extraction in time and frequency domain

According to Fig. 2 in the first stage features are extracted from the PMU samples in time and frequency domain. Within this study the total feature space consists of 16 different features F1 to F16. Typical disturbance patterns (e.g. oscillations, line or generator trips, voltage sags) are characterized by high signal variations and abrupt signal changes.

Those signal traits can be captured by statistical moments like variance (F1), skewness (F3) and kurtosis (F2), model-based metrics like linear slope (F4) or information based metrics like sampled entropy (F5). Additional information are gained by time-frequency representation of the PMU samples. The features are derived from the energy values of the absolute Stockwell Transform (S-Transform) coefficients and the energy values of the detailed coefficients from multi-level Wavelet decomposition. Fig. 4 shows an example for the S-Transform of a PMU sample. The S-Transform derived features include mean energy (F6), maximum energy along time axis (F7) and frequency axis (F8) as well as energy variance along time axis (F9) and frequency axis (F10).

The DWT derived features include mean energy (F11), energy variance (F12) and maximum energy (F13) over all decomposition levels. For a better understanding Fig. 5 and Fig. 6 show some exemplary time and frequency domain features.

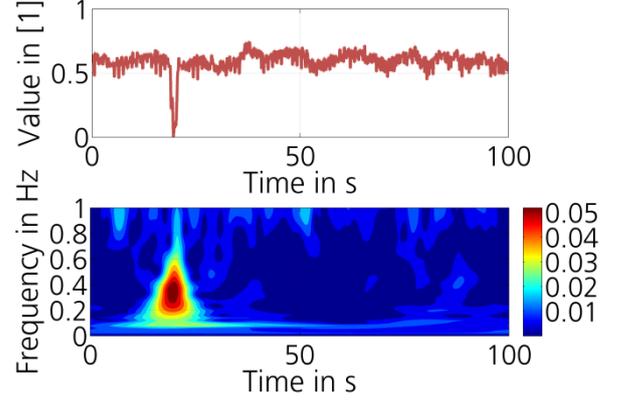


Fig. 4. S-Transform absolute coefficients for a PMU sample

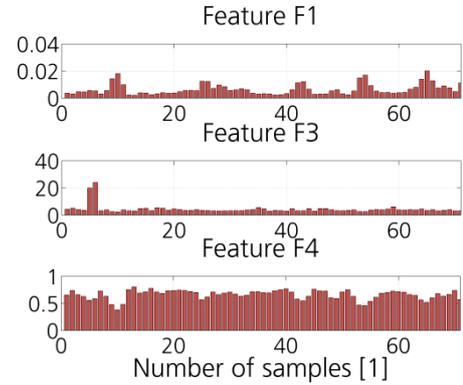


Fig. 5. Results for time domain features F1, F3 and F4

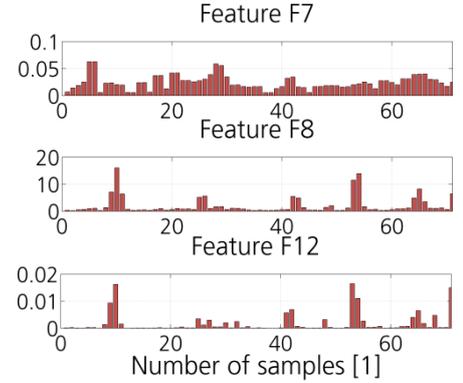


Fig. 6. Results for frequency domain features F7, F8 and F12.

### 3.3 Outlier scores and disturbance patterns

The following investigations refer to the extracted features in time and frequency domain for one PMU sensor. The results are mapped into a two-dimensional space using PCA and Isomap – see again Fig. 2. Afterwards the three outlier detection methods LOF, COP and SiLiOd calculate an outlier score for each PMU sample. Given that, for each feature space there are 6 different outlier detectors:

- Detector 1,2: PCA / Isomap + LOF
- Detector 3,4: PCA / Isomap + COP
- Detector 5,6: PCA / Isomap + SiLiOd

As noted in Table II the number of outliers to be extracted is set to 15 given a total number of 71 PMU samples. In case of LOF and COP the number of nearest neighbours substantially determines the results. Fig. 7 show the outlier scores from detector 1 for different numbers of nearest neighbours using the time-domain features.

The reduced two-dimensional space is spanned by the first and second principle component (PC) of the PCA. The detected outliers are highlighted in red. High outlier scores with  $LOF \gg 1$  indicate a high distance between the data point and its surroundings. Far-away outliers are detected successfully by LOF. When increasing the number of nearest neighbours the outlier scores are changing especially for data points with a low distance to its surrounding. Within this survey the number of nearest neighbours is set to 50 which corresponds to more than half of the total PMU samples.

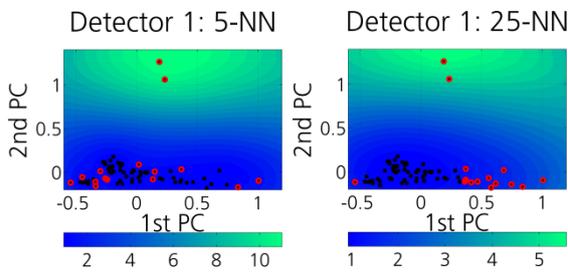


Fig. 7. Outlier scores for detector 1 (time domain features).

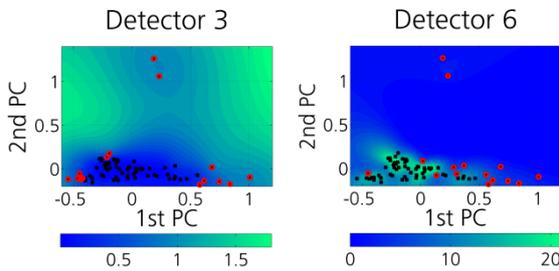


Fig. 8. Outlier scores for detector 3 and 6 (time domain features).

Fig. 8 presents the COP and SiLiOd outlier scores for the same data set based on the PCA mapping (detector 3 and 6). The COP outlier scores show similar results compared to the LOF outlier scores using a number of nearest neighbours of 50. This parameter is fixed for all subsequent investigations. In case of SiLiOd low outlier scores or small path lengths correspond to a high outlier degree. The metric is inversely proportional to the metrics from LOF or COP. Within this survey the Mahalanobis distance is used to create the cluster tree – see again Table II. In Fig. 9 the aggregated results from the 6 outlier detectors are presented within ranking matrices. A low rank indicates a high outlier degree and vice versa. The ranking matrices show the top 15 outliers over all PMU samples. For far-away outliers with ranks from 1 to 4 low variations between the results can be observed which indicates high certainty. This accounts

for the sample numbers 6, 7 and 65, 66. In contrast to that, near-by outliers with high ranks are widely spread among the PMU samples indicating low certainty of the results. Some high-rank examples can be detected for the sample number 1 and 60. Different results are obtained using features from frequency domain analysis – see Fig. 10. In this case high outlier ranks can be observed inter alia for the sample numbers 28 and 29.

In a last step the results of the different outlier detectors can be aggregated via rank transformation. With that the PMU samples with the highest outlier ranks are selected among all detectors. The results are presented in Fig. 11 for the time excerpt from 200 to 1800 s for different outlier numbers. The whole voltage signal (blue) and the identified outlier samples (red) are shown. As it can be seen in both cases a voltage sag between 200 and 300 s and a voltage oscillation between 450 and 580 s can be detected successfully. Additionally some other outliers are identified in the lower figure between 1200 and 1500 s including other voltage oscillations. As it can be seen, when choosing a low number of outliers not all disturbances or unusual pattern might be captured. When choosing a high number of outliers more disturbances can be identified but also a certain amount of non-relevant patterns. For a better illustration Fig. 12 compares some outlier (left picture) and non-outlier (right picture) PMU patterns. Sample 6 corresponds to a voltage sag and sample 10 corresponds to a voltage oscillation. Both samples have a high outlier rank whereas samples 20 and 60 have the lowest outlier ranks – see again in Fig. 9 and Fig. 10.

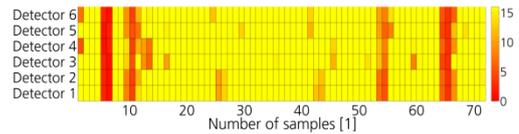


Fig. 9. Summary of outlier ranks (time domain features).

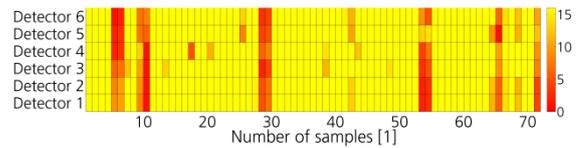


Fig. 10. Summary of outlier ranks (frequency domain features).

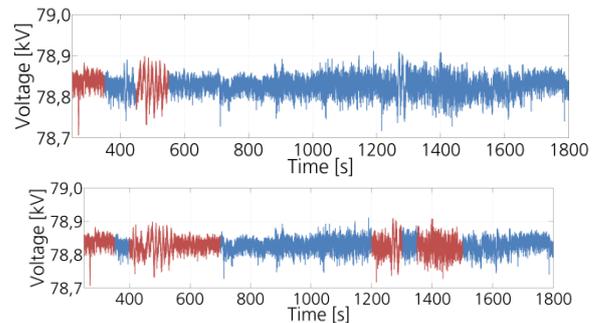
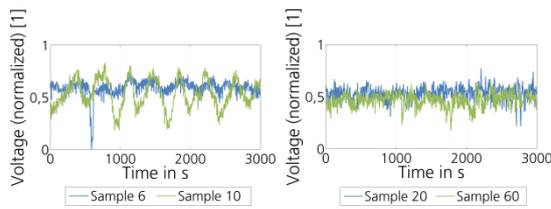


Fig. 11. Voltage signal (blue) and detected outlier samples (red) for outlier number = 5 (top) and outlier number = 15 (bottom).



**Fig. 12.** Outlier samples (left) and non-outlier samples (right).

## 4 Summary and outlook

This study introduces a new method for the uncovering of critical disturbances from historical phasor measurements. Within a three-phase concept outlier scores are computed for each PMU sample to detect critical disturbances. The relevant features are generated in time and frequency domain including e.g. statistical moments and multi-level Wavelet decomposition. Different outlier detection methods are combined using voltage magnitudes from a real phasor measurements record. The algorithms include established techniques like Local Outlier Factors (LOF) and Correlation Outlier Probabilities (COP) as well as an own implementation of a single linkage based outlier detection method (SiLiOd). The results of 6 different outlier detectors are compared which each other. It is shown that typical disturbance patterns like voltage sags and voltage oscillations can be extracted with the use of outlier detection methods. With this concept mixed PMU data sets can be explored and analyzed by uncovering unusual or potential critical events. This can support further decision making and serve as training data for online detection and classification algorithms.

For further investigations additional test data is required to validate the results and to extract other disturbance types like line or generator trips as well as analyzing other measurements like frequencies or voltage angles. Also, the extracted potential disturbances have to be clustered or grouped in a post-processing step to better select relevant and non-relevant signal patterns. Apart from that new techniques for feature extraction and outlier detection (e.g. Isolation Forests or Deep Autoencoder) have to be tested and compared to the existing approach. Also an procedure has to be developed for the automated extraction of critical disturbances or events from mixed and imbalanced PMU data sets. With such an approach large PMU data sets can be explored and filtered for subsequent event analysis or classification tasks.

## References

1. J. F. Hauer, D. J. Trudnowski, and J. G. DeSteese, "A Perspective on WAMS Analysis Tools for Tracking of Oscillatory Dynamics," in *2007 IEEE Power Engineering Society General Meeting*, Tampa, FL, USA, pp. 1–10.
2. V. Terzija *et al.*, "Wide-Area Monitoring, Protection, and Control of Future Electric Power Networks," *Proc. IEEE*, vol. 99, no. 1, pp. 80–93, 2011.
3. L. F. d. Santos, G. Antonova, M. Larsson, and S. Fujii, "The Use of Synchrophasors for Wide Area Monitoring of Electrical Power Grids," 2013.
4. J. Guo *et al.*, "Design and Implementation of a Real-Time Off-Grid Operation Detection Tool from a Wide-Area Measurements Perspective," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2080–2087, 2015.
5. S. Balischewski *et al.*, "A new development platform for the next generation of power system control center functionalities for hybrid AC-HVDC transmission systems," in *IEEE PES General Meeting Chicago July 2017*.
6. S. Brahma *et al.*, "Real-Time Identification of Dynamic Events in Power Systems Using PMU Data, and Potential Applications—Models, Promises, and Challenges," *IEEE Trans. Power Delivery*, vol. 32, no. 1, pp. 294–301, 2017.
7. Q. H. Alsafasfeh, "PATTERN RECOGNITION FOR FAULT DETECTION, CLASSIFICATION, AND LOCALIZATION IN ELECTRICAL POWER SYSTEMS," Dissertation, Department of Electrical and Computer Engine, Western Michigan Univ, Kalamazoo, Michigan, 2010.
8. M. Biswal and P. K. Dash, "Detection and characterization of multiple power quality disturbances with a fast S-transform and decision tree based classifier," *Digital Signal Processing*, vol. 23, no. 4, pp. 1071–1083, 2013.
9. Y. Ge *et al.*, "Power System Real-Time Event Detection and Associated Data Archival Reduction Based on Synchrophasors," *IEEE Trans. Smart Grid*, vol. 6, no. 4, pp. 2088–2097, 2015.
10. A. Chiang, E. David, Y.-J. Lee, G. Leshem, and Y.-R. Yeh, "A study on anomaly detection ensembles," *Journal of Applied Logic*, vol. 21, pp. 1–13, 2017.
11. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.
12. C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to wavelets and wavelet transforms: A primer*. Upper Saddle River, NJ: Prentice Hall, 1998.
13. R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: The S transform," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 998–1001, 1996.
14. J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," (eng), *American journal of physiology. Heart and circulatory physiology*, vol. 278, no. 6, H2039–49, 2000.

15. J. H. Janssens, "Outlier Selection and One-Class Classification," Dissertation, Tilburg University, Tilburg, 2013.
16. M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-Based Local Outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, Dallas, Texas, United States, 2000, pp. 93–104.
17. H.-P. Kriegel, P. Kroger, E. Schubert, and A. Zimek, "Outlier Detection in Arbitrarily Oriented Subspaces," in *IEEE 12th International Conference on Data Mining (ICDM), 2012: 10 - 13 Dec. 2012, Brussels, Belgium ; proceedings*, Brussels, Belgium, 2012, pp. 379–388.
18. E. Alpaydin, *Introduction to machine learning*, 2nd ed. Cambridge, Mass: MIT Press, 2010.
19. A. Allen, M. Singh, and E. Muljadi: NREL and S. Santoso: University of Texas at Austin, "PMU Data Event Detection: A User Guide for Power Engineers," Technical Report, NREL National Renewable Energy Laboratory, The University of Texas at Austin NREL/TP-5D00-61664, 2014.
20. I. The MathWorks, Ed., "MATLAB 7.5 and Signal Processing Toolbox 6.1, Statistics Toolbox 6.1, Wavelet Toolbox 4.1," Natick, Massachusetts, United States, 2007.
21. E. Raff, "JSAT: Java Statistical Analysis Tool, a Library for Machine Learning," *Journal of Machine Learning Research*, vol. 18, no. 23, pp. 1–5, <http://jmlr.org/papers/v18/16-131.html>, 2017.
22. E. Achtert, S. Goldhofer, H.-P. Kriegel, E. Schubert, and A. Zimek, "Evaluation of Clusterings - Metrics and Visual Support Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012," pp. 1285–1288.