


Chapter 8

Future Outlier Detection Algorithm for Smarter Industry Application Using ML and AI: Explainable AI and ML for Smart Industry Evolution Using ML/AI Algorithms and Implementations

Kunal Dhibar

Bengal College of Engineering and Technology, India

Prasenjit Maji

 <https://orcid.org/0000-0001-8057-6963>

Dept of CSD, Dr. B.C. Roy Engineering College, Durgapur, India

ABSTRACT

Throughout many real-world investigations, outliers are prevalent. Even a few aberrant data points can cause modeling misspecification, biased parameter estimate, and poor forecasting. Outliers in a time series are typically created at unknown moments in time by dynamic intervention models. As a result, recognizing outliers is the starting point for every statistical investigation. Outlier detection has attracted significant attention in a variety of domains, most notably machine learning and artificial intelligence. Anomalies are classified as strong outliers into point, contextual, and collective outliers. The most significant difficulties in outlier detection include the narrow boundary between remote sites and natural areas, the propensity of fresh data and noise to resemble genuine data, unlabeled datasets, and varying interpretations of outliers in different applications.

DOI: 10.4018/978-1-6684-8785-3.ch008

INTRODUCTION

Most real-world datasets contain data observations that are unlikely to correspond to the particular framework and/or characteristics of such dataset. Outliers are assertions that differ markedly from most of the data points found in the datagram. Outlier identification is a problem that must be addressed in a variety of usage, which include fake prevention and detection (e.g., potentially malicious utilization credit and debit cards or even different kinds of monetary transactions), healthcare information analysis (e.g., capable of recognizing dynamics are changing to therapeutic interventions among patient populations), fault detection in production processes, and detection of network intrusions, among others. Additionally, the presence of outliers effects several data processing tasks, necessitating the limitation or elimination of outlier observations. Recognizing outliers in multidimensional data is a tough task that gets increasingly difficult when dealing with high-dimensional datasets.

Outlier identification is a significant topic in data mining which has been widely investigated for many years by different scholars. The definition of an anomaly is “an occurrence in a dataset that seems to be incongruous with the balance of that the collection of data,” according to. Mining outliers’ methodologies may be utilized in a range of areas, notably identifying credit card fraud, detection of network intrusions, and environmental monitoring. In feature extraction, there are two primary missions. To begin, we should define what data are considered outliers in a given set. Secondly, a productive technique for calculating these outliers is required. The outlier problem was initially investigated by the statistical community. They assume that the presented dataset is normally distributed, and an entity is referred to as an outlier if it differs significantly considerably from these dividends. In contrast, finding a suitable allocation for responsible for considerable is practically difficult. To solve the above noted shortcoming, the information management industry has proposed a number of model-free solutions. Outliers based on distance and outliers based on density are two examples.

Outlier detection is a prominent subject in the data mining field since it intends to find trends that happen seldom when compared to other data mining methodologies. An outlier is a finding that differs substantially from and then contradicting the main section of a datagram, as if though created by a separate methodology. Outlier detection is critical because outliers may provide both raw patterns and actionable insights about something like a dataset. Outlier detection is used in a broad variety of situations, as well as criminal identification, credit card fraud analysis, detection of network intrusions, medical diagnostics, defective detection in essential safety systems, and image processing abnormalities detection. There has been an awful lot of effort recently in outlier detection research, with several approaches described. Current research on outlier detection may be classified into three types based on whether or not label evidence is accessible or can be used to create outlier detection approach: unstructured, supervised, and semi-supervised strategies.

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/future-outlier-detection-algorithm-for-smarter-industry-application-using-ml-and-ai/325995

Related Content

Generative Adversarial Networks for Data Augmentation in Image Recognition: An Exploratory Study

Uriel U. Onye, Sia Charan Lankaand Pujita Kodali (2025). *International Journal of Artificial Intelligence and Machine Learning* (pp. 1-10).

www.irma-international.org/article/generative-adversarial-networks-for-data-augmentation-in-image-recognition/393280

Introduction to Machine Learning and Its Implementation Techniques

Arul Murugan R.and Sathiyamoorthi V. (2022). *Research Anthology on Machine Learning Techniques, Methods, and Applications* (pp. 1-25).

www.irma-international.org/chapter/introduction-to-machine-learning-and-its-implementation-techniques/307441

From Traditional Debugging to Generative Intelligence in the Evolution of Defect Detection With LLMs

Rajeev Ranjan Kumar, Kripa Shankerand Meetu Malhotra (2026). *Automating Software Defect Detection Through Machine Learning and LLMs* (pp. 319-344).

www.irma-international.org/chapter/from-traditional-debugging-to-generative-intelligence-in-the-evolution-of-defect-detection-with-llms/391186

Image Classification Using Deep Neural Networks: Emotion Detection Using Facial Images

Sukanta Ghoshand Amar Singh (2021). *Machine Learning and Data Analytics for Predicting, Managing, and Monitoring Disease* (pp. 75-85).

www.irma-international.org/chapter/image-classification-using-deep-neural-networks/286244

Entropy, Chaos, and Language

Daniela Lopez De Luise (2022). *Handbook of Research on New Investigations in Artificial Life, AI, and Machine Learning* (pp. 331-370).

www.irma-international.org/chapter/entropy-chaos-and-language/296810