УДК 004.7

Mykola Varanytsia

STATISTICAL METHODS FOR ANOMALY DETECTION IN DATA USING THE PYTHON PROGRAMMING LANGUAGE

Vinnytsia National Technical University

Abstract

The paper examines the development of a software implementation of statistical methods for anomaly detection in data using the Python programming language. It explores approaches to data processing, analysis of statistical deviations, and the construction of algorithms for automated anomaly detection.

Keywords: Anomalies, statistical methods, Python, machine learning, data analysis, anomaly detection, algorithms, automation

Анотація

У роботі розглядається розробка програмної реалізації статистичних методів для виявлення аномалій у даних із використанням мови програмування Python. Досліджуються підходи до обробки даних, аналізу статистичних відхилень і побудови алгоритмів для автоматизованого виявлення аномалій.

Ключові слова: Аномалії, статистичні методи, Python, машинне навчання, аналіз даних, виявлення аномалій, алгоритми, автоматизація

Introduction

Anomaly detection in data is a crucial task in data analysis, machine learning, and statistics. Anomalies may indicate errors, fraudulent activities, technical failures, or other unusual situations. Therefore, developing effective methods for detecting such deviations is a relevant challenge.

Statistical methods play a key role in the anomaly detection process, as they allow for assessing deviations from the normal distribution of data. The main approaches include using the mean and standard deviation, threshold-based methods, interquartile range, and other statistical analysis metrics. Combined with the capabilities of the Python programming language, which offers a wide range of data analysis libraries (numpy, pandas, scipy, scikit-learn, etc.), these methods can be effectively implemented and applied in various fields.

This work is dedicated to researching statistical methods for anomaly detection and their practical application in the Python environment. Special attention is given to algorithm efficiency, performance, and automation possibilities in data processing.

Prospects for development

The development prospects of anomaly detection methods are associated with improving algorithms and adapting them to different types of data. Modern approaches include combining statistical methods with machine learning and artificial intelligence to enhance prediction accuracy and reduce the number of false-positive results[1].

One of the promising directions is the use of neural network approaches for anomaly detection, which can effectively identify complex patterns in data[2]. Autoencoders, Generative Adversarial Networks (GANs), and Recurrent Neural Networks (RNNs) demonstrate high potential in this field. Combining statistical and neural network methods allows for greater adaptability of the system to data changes.

Additionally, an important aspect is the development of tools for anomaly visualization and interactive analysis. The use of libraries such as Matplotlib, Seaborn, and Plotly enables the creation of clear graphs and diagrams for better understanding of data structure and anomaly localization[3].

Another direction is the development of automated monitoring and alert systems that integrate with existing corporate solutions. Such systems can analyze data streams in real-time and notify users about potential anomalous situations[4].

Particular attention should be given to improving data preprocessing methods. Data noise removal, normalization, and transformation can significantly impact algorithm performance. The use of modern methods,

such as Principal Component Analysis (PCA) or t-SNE, allows for dimensionality reduction and identification of key features influencing anomaly detection[5].

Another crucial area is the implementation of continuously updating anomaly detection models that can adapt to changes in incoming data without manual configuration. This approach enhances model robustness and effectiveness in the long run.

Thus, further research in anomaly detection can significantly improve analysis accuracy, increase the efficiency of data processing systems, and find applications across various industries—from financial institutions to medical diagnostic systems.

Conclusions

The development of software methods for anomaly detection in data using Python is a relevant area in data analysis. Statistical methods provide reliable approaches for assessing deviations, and their combination with modern machine learning algorithms enhances the effectiveness of anomaly detection.

The application of these methods in various fields, such as finance, security, and quality control, confirms their practical value. Further advancements in this field involve the implementation of deep neural networks, improvements in data preprocessing algorithms, and the creation of integrated solutions for real-time anomaly monitoring.

Thus, the use of statistical methods for anomaly detection in Python opens up vast opportunities for improving data analysis and developing automated anomaly detection systems.

REFERENCES

1.Aggarwal, C. C. (2017). Outlier Analysis. Springer.

2.Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly Detection: A Survey. ACM Computing Surveys.

3.Raschka, S., & Mirjalili, V. (2019). Python Machine Learning. Packt Publishing.

4.Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research.

5. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

Mykola Varanytsia **–** student of group 4PE-21b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: pro100spaderket@gmail.com

Scientific supervisor: ***Liudmyla Magas*** – a senior teacher of English, Foreign Languages Department, Vinnytsia National Technical University, Vinnytsia, e-mail:magas@vntu.edu.ua

Вараниця Микола Сергійович – студент групи 4ПІ-21б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: pro100spaderket@gmail.com

Науковий керівник: ***Магас Людмила Миколаївна*** – старший викладач англійської мови, кафедра іноземних мов, Вінницький національний технічний університет, м. Вінниця, e-mail: magas@vntu.edu.ua