The doctoral dissertation titled "Anomaly Identification in Domain-Specific Complex Data Sets" focuses on the development of anomaly detection methods aimed at improving the effectiveness of identifying anomalies in complex data sets. The primary objective of the research is to develop techniques that enhance anomaly detection accuracy compared to traditional approaches, particularly in multidimensional and categorical data. A key challenge addressed in the dissertation is the issue of class imbalance, which is critical in fields such as finance, medicine, and monitoring systems. Early and effective anomaly detection can help prevent critical events such as fraud, technical failures, or security threats. The dissertation introduces innovative approaches by combining deep learning methods with traditional anomaly detection techniques in data with uneven distributions.

The work emphasizes the development of methods that increase both the sensitivity and efficiency of anomaly detection, with a focus on practical applications and the optimization of the Local Outlier Factor (LOF) algorithm. The author's optimization of data block sizes reduces processing time while maintaining high effectiveness. Anomalies are defined in the dissertation as patterns that significantly deviate from expected behaviour, and the work discusses the challenges associated with the analysis of complex data. A significant aspect of the research is the evaluation of modern technologies, such as machine learning and deep learning, in the context of improving detection sensitivity. The research also aims to develop ensemble techniques (outlier ensemble) that enhance both the accuracy and efficiency of the detection process. The dissertation concludes with the validation of the proposed methods on real-world data sets, allowing for an assessment of their practical effectiveness.

The thesis of the dissertation asserts that the use of advanced techniques, such as Self-Organizing Maps (SOM), block-divided LOF, and Autoencoders (AE) within an integrated ensemble of algorithms, significantly increases the sensitivity and efficiency of anomaly detection in complex, real-world data sets. At the same time, it ensures optimal processing performance, especially through the application of the LOF optimization method. The validation of this thesis is demonstrated in the results discussed in the dissertation.

The dissertation consists of nine chapters covering key aspects of anomaly detection. Chapter two redefines anomalies, analyses the challenges of detecting them in complex data sets, and discusses contemporary anomaly detection methods, emphasizing their significance in fields such as medicine, finance, and infrastructure monitoring. Chapter three discusses the difficulties and techniques for processing categorical and multidimensional data. Chapter four presents the Trinity SALT system, which integrates machine learning and deep learning techniques (SOM, AE, and LOF). The author's technique of maximum normalized aggregation (MNA) is an innovative approach that stands out from traditional methods due to its unique model consensus weighting. Chapter five analyses ensemble anomaly detection methods, which improve detection accuracy by combining results. While this is a relatively new approach, the chapter demonstrates its potential in anomaly analysis, drawing on

successes in classification and recommendation systems. Chapter six discusses performance metrics and the impact of decision thresholds on balancing anomaly detection with false alarms, also highlighting common mistakes in algorithm performance evaluation. Chapter seven describes the implementation of Trinity SALT and a web application for visualizing the results. Chapter eight presents experimental results that confirm Trinity SALT's superiority over individual algorithms, showing higher sensitivity and F1 scores, as well as greater stability. Block optimization further speeds up processing, enhancing system efficiency. Trinity SALT has potential applications in medicine and cybersecurity, where high accuracy and speed are crucial. Chapter nine summarizes the findings and proposes future research directions.