Abstract

Outlier detection is a critical task in various domains, such as finance and cybersecurity, as it helps identify anomalies that can provide valuable insights for data cleansing and decision-making. The increasing availability of large and complex datasets has led to a growing demand for effective outlier detection models. While numerous approaches exist, there is a need for comprehensive research that compares and evaluates these models to understand their performance and suitability for different datasets and outlier scenarios. This thesis aims to conduct a comparative analysis of outlier detection models and apply them to data used in transaction monitoring, to gain insights into their strengths, weaknesses, and real-world applicability in this field. The models examined include Isolation Forest, cluster-based analysis, and copulas, each suitable for different sets of use cases. Given the challenges of evaluating transaction monitoring data due to missing or unreliable data labels, this comparative analysis seeks to provide a clear understanding of how these models perform under such conditions and how can they be evaluated based on the expert-based knowledge.