

Rozprawa doktorska pt. „*Identyfikacja anomalii w dziedzinowych zbiorach danych złożonych*” koncentruje się na opracowaniu metod detekcji anomalii, mających na celu poprawę skuteczności ich identyfikacji w złożonych zbiorach danych. Celem pracy jest rozwój technik, które zwiększą skuteczność wykrywania anomalii w porównaniu do klasycznych podejść, zwłaszcza w danych wielowymiarowych i kategoriowych. Istotnym wyzwaniem, które podejmuje praca, jest problem nierównowagi klas, kluczowy w takich dziedzinach jak finanse, medycyna czy systemy monitoringu. Wczesne i skuteczne wykrywanie anomalii może zapobiec krytycznym zdarzeniom, takim jak oszustwa, awarie techniczne czy zagrożenia bezpieczeństwa. Rozprawa wprowadza innowacyjne podejścia, łącząc metody głębokiego uczenia i tradycyjne techniki wykrywania anomalii w danych o nierównomiernie dystrybuacji.

Praca koncentruje się na opracowaniu metod zwiększających czułość i wydajność wykrywania anomalii, z naciskiem na praktyczne zastosowania oraz optymalizację algorytmu Local Outlier Factor (LOF). Autorska optymalizacja rozmiaru bloków danych pozwala skrócić czas przetwarzania przy zachowaniu wysokiej skuteczności. W pracy definiuje się anomalie jako wzorzec znacząco odbiegający od oczekiwanego zachowania oraz omawia wyzwania związane z analizą danych złożonych. Istotną rolę odgrywa ocena nowoczesnych technologii, takich jak uczenie maszynowe i głębokie uczenie, w kontekście poprawy czułości detekcji. Celem jest także rozwój technik zespołowych (outlier ensemble), które zwiększają skuteczność i wydajność procesów. Praca kończy się walidacją proponowanych metod na rzeczywistych zbiorach danych, co umożliwia ocenę ich skuteczności w praktyce.

Teza niniejszej rozprawy głosi, że zastosowanie zaawansowanych technik, takich jak Self-Organizing Maps (SOM), LOF z podziałem na bloki oraz Autoenkodery (AE) w ramach zintegrowanego zespołu algorytmów, pozwala znacząco zwiększyć czułość i wydajność wykrywania anomalii w złożonych, rzeczywistych zbiorach danych. Jednocześnie zapewnia optymalną wydajność procesu analizy, co jest szczególnie widoczne dzięki zastosowaniu metody optymalizacji LOF. Potwierdzenie tej tezy znajduje się w wynikach badań, omówionych w pracy.

Rozprawa składa się z dziewięciu rozdziałów, omawiających ważne aspekty wykrywania anomalii. Rozdział drugi definiuje na nowo anomalie i analizuje wyzwania ich detekcji w złożonych zbiorach danych oraz omawia współczesne metody wykrywania anomalii, podkreślając znaczenie w medycynie, finansach i monitorowaniu infrastruktury. Rozdział trzeci omawia trudności i pokazuje techniki przetwarzania danych kategoriowych i wielowymiarowych. Rozdział czwarty przedstawia system Trinity SALT, integrujący techniki uczenia maszynowego i głębokiego (SOM, AE i LOF). Autorska technika maksymalnej znormalizowanej agregacji (MNA) stanowi innowacyjne podejście, które wyróżnia się na tle tradycyjnych metod dzięki unikalnemu premiowaniu konsensusu modeli. Rozdział piąty analizuje zespołowe metody detekcji anomalii, które zwiększają skuteczność ich wykrywania przez łączenie wyników. Choć to nowe podejście, rozdział wykazuje jego potencjał w analizie anomalii, bazując na sukcesach w klasyfikacji i systemach rekomendacji. W rozdziale szóstym omówiono miary skuteczności oraz wpływ progów decyzyjnych na balans między wykrywaniem anomalii a fałszywymi alarmami, wskazując również na typowe błędy przy

ocenie wydajności algorytmów. Rozdział siódmy opisuje implementację Trinity SALT i aplikację webową do wizualizacji wyników. Rozdział ósmy przedstawia wyniki eksperymentów, które potwierdzają wyższość Trinity SALT nad pojedynczymi algorytmami, wskazując na wyższe miary czułości i F1 oraz większą stabilność. Optymalizacja bloków dodatkowo przyspiesza przetwarzanie, co zwiększa wydajność systemu. Trinity SALT może pomóc w medycynie i cyberbezpieczeństwie, gdzie liczy się wysoka skuteczność i szybkość. Rozdział dziewiąty podsumowuje wyniki i proponuje przyszłe kierunki badań.