

Dr hab. inż. Katarzyna Haręźlak, prof. PŚ
Katedra Informatyki Stosowanej
Wydział Automatyki, Elektroniki i Informatyki
Politechnika Śląska w Gliwicach

Recenzja rozprawy doktorskiej

mgr. inż. Czesława Horynia

pt. „Identyfikacja anomalii w dziedzinowych zbiorach danych złożonych”

przygotowanej pod kierunkiem

dr hab. Agnieszki Nowak - Brzezińskiej, prof. UŚ

Niniejsza recenzja została przygotowana na podstawie uchwały Rady Naukowej Instytutu Informatyki Uniwersytetu Śląskiego w Katowicach z dnia 5.11.2024 w przedmiocie wyznaczenia recenzentów niniejszej pracy doktorskiej (Uchwała nr 55/2024).

1. Przedmiot rozprawy, aktualność i ważność tematyki rozprawy

Eksploracja oraz proces odkrywania wzorców i wiedzy z dużych ilości danych są kluczowymi umiejętnościami we współczesnym świecie opartym na informacjach. Jednak anomalie lub wartości odstające mogą znacząco zakłócić ten proces, prowadząc do niedokładnych wyników i błędnych decyzji. Punkty danych, które odbiegają od oczekiwanego wzorca, mogą być spowodowane różnymi zdarzeniami, do których należą między innymi błędy pomiaru, uszkodzenia danych lub nowe informacje. Dlatego zapewnienie integralności procesu eksploracji danych z uwzględnieniem danych odstających ma istotne znaczenie dla wiarygodności wyników.

Wykrywanie anomalii ma swoją historię w statystyce, jednakże wraz ze wzrostem wolumenów i tempem gromadzenia danych, potrzeba zautomatyzowanego wykrywania punktów odstających w czasie rzeczywistym stała się dużym wyzwaniem. W ostatnich dekadach podjęto się tego zadania z wykorzystaniem technik uczenia maszynowego, a badania naukowe koncentrują się nad znalezieniem wydajnych i efektywnych sposobów wykrywania różnych typów anomalii.

W procesie tym należy zmierzyć się z kilkoma problemami. Jednym z nich jest rozmiar danych, który z jednej strony odgrywa ważną rolę w zapewnieniu wystarczającego kontekstu dla rozróżniania próbek wzorcowych i odstających, z drugiej strony wprowadza wymagania dotyczące zasobów i czasu przetwarzania. Ponadto, zwykle natrafiamy tu na problem niezbalansowania danych, ze względu na przytłaczającą większość próbek typowych w stosunku do anomalii. Stanowi to dość duże ograniczenie dla metod uczenia maszynowego, które dodatkowo wykazują różną skuteczność w zależności od rozwiązywanego problemu. Nie bez znaczenia jest również wielowymiarowość oraz obecność kategoriycznych typów danych, które — w kontekście uczenia maszynowego i wyznaczania stopnia podobieństwa — wymagają stosowania odpowiednich metod kodowania.

Wszystko to sprawia, że zadanie wykrywania anomalii w danych jest procesem złożonym, a brak jednoznacznych rozwiązań w tej dziedzinie wymaga prowadzenia intensywnych badań. W tym

kontekście tematykę podjętą w rozprawie doktorskiej Pana magistra inżyniera Czesława Horynia, dotyczącą rozwoju metod wykrywania anomalii, należy uznać za ważną, celową i aktualną.

Główny cel, jaki postawiono w rozprawie dotyczy:

opracowania metod i narzędzi, które zwiększą czułość i efektywność wykrywania anomalii w złożonych zbiorach danych, z naciskiem na praktyczne zastosowania i optymalizację algorytmu LOF.

Cel ten podzielony został na cele podrzędne:

- zdefiniowanie pojęcia anomalii oraz identyfikacja wyzwań związanych z ich wykrywaniem w złożonych, kategoriowych i wielowymiarowych danych,
- zastosowanie i ocena nowoczesnych technologii, takich jak uczenie maszynowe i głębokie uczenie, w kontekście identyfikacji anomalii, z naciskiem na zwiększenie czułości,
- rozwój i optymalizacja zaawansowanych technik zespołowych, w celu zwiększenia skuteczności i wydajności procesu wykrywania anomalii,
- implementacja i testowanie proponowanych metod na rzeczywistych zbiorach danych, aby ocenić ich efektywność, czułość oraz użyteczność w praktycznych zastosowaniach.

Teza rozprawy została sformułowana następująco:

Zastosowanie zaawansowanych technik, takich jak SOM, LOF z podziałem na bloki oraz autoenkodery w ramach zespołu algorytmów, umożliwia zwiększenie czułości i wydajności wykrywania anomalii w rzeczywistych, złożonych zbiorach danych, przy jednoczesnym zachowaniu wydajności procesu analizy, szczególnie dzięki optymalizacji LOF.

2. Zawartość rozprawy

Rozprawa doktorska podzielona została na dziewięć rozdziałów oraz trzy dodatki. Łącznie z bibliografią, spisami tabel i rysunków składa się z **393** stron. Pierwszych **230** stron zawiera opis zagadnień teoretycznych, natomiast opisowi badań zrealizowanych w ramach rozprawy doktorskiej poświęcono ostatnie trzy rozdziały, liczące łącznie **90** stron. Taka proporcja wydaje się niewłaściwa dla rozprawy doktorskiej, szczególnie, że Doktorant wielokrotnie powtarzał te same treści, czasem wykorzystując taki sam zestaw słów, co niekorzystanie wpłynęło na rozmiar pracy.

W rozdziale pierwszym zawarto wprowadzenie w tematykę anomalii, przedstawiono motywację prowadzonych badań, oraz cele i tezę rozprawy. Rozdział ten zawiera również opis układu pracy. W rozdziale drugim Doktorant omówił problem identyfikacji anomalii w danych złożonych. Przedstawił rys historyczny oraz przytoczył kilka definicji i terminologii anomalii, wzbogacając opis rysunkami. W rozdziale tym Autor skupił uwagę również na typach i strukturach danych. Dużo miejsca poświęcił matematycznym podstawom przetwarzania danych w uczeniu maszynowym. W mojej opinii ta część, jak również wyjaśnianie, takich pojęć jak nauka o danych czy eksploracja, stanowią nadmiarowe elementy rozprawy. Jest to wiedza podstawowa, którą można publikować w skryptach lub podręcznikach naukowych. Interesujące i powiązane z tematyką pracy były rozdziały dotyczące miar odległości i metod wykrywania anomalii. Chociaż opisy, prezentowane w drugim z wymienionych podrozdziałów, cechują się zbyt dużą szczegółowością, jak na podrozdział rozprawy doktorskiej. W podsumowaniu rozdziału drugiego zabrakło natomiast informacji, które z wymienionych metod i w jaki sposób wpłynęły na realizację postawionego w rozprawie celu. Trzeci rozdział, w pierwszej

swojej części, poświęcony został omówieniu problemów związanych z przetwarzaniem danych kategorycznych. Przedstawione zostały metody wykrywania anomalii w danych kategorycznych oraz mieszanych, a także miary podobieństwa zdefiniowane dla tego typu danych. Druga część tego rozdziału skupia się na metodach przetwarzania danych wielowymiarowych. Podobnie, jak w przypadku rozdziału drugiego, rozdział trzeci zawiera zbyt obszerne treści, dla których lepszym miejscem prezentacji byłby podręcznik naukowy. W podsumowaniu rozdziału trzeciego również zabrakło implikacji dla zadań prowadzonych w ramach rozprawy doktorskiej. Rozdział czwarty stanowi omówienie teoretycznych podstaw uczenia głębokiego oraz zastosowanie tego typu metod do wykrywania anomalii. Niestety, tylko część tego rozdziału zawiera interesujące, z punktu widzenia rozprawy doktorskiej, informacje. Reszta stanowi pewien szum informacyjny, który dodatkowo zawiera powtarzające się treści. Dla przykładu:

- Strona 130: *Tradycyjne metody często wymagają ręcznego projektowania cech i predefiniowania struktury modelu, podczas gdy sieci neuronowe wykorzystują optymalizację, co pozwala na bardziej efektywne uczenie się i adaptację do skomplikowanych wzorców w zbiorach informacji.*
- Strona 131: *W tradycyjnych modelach uczenia maszynowego cechy (ang. features) są zazwyczaj ręcznie projektowane przez ekspertów dziedzinowych. W głębokim uczeniu proces ten jest zautomatyzowany - sieci neuronowe same uczą się odpowiednich cech podczas procesu treningu.*
- Strona 132: *Tradycyjne metody skupiają się na ręcznym definiowaniu cech i prostych modelach, które działają na pojedynczych warstwach danych, co ogranicza ich zdolność do automatycznego wykrywania złożonych wzorców.*

Tego typu powtórzeń jest w pracy niestety znacznie więcej. Pewne zdziwienie budzi fakt, że Doktorant zdecydował się omówić w pracy, na przykład, genezę i rozwój sieci neuronowych, a prace badawcze [85, 101, 264, 286, 304, 305] prowadzące Doktoranta najprawdopodobniej do końcowego wyniku, jakim jest rozprawa doktorska, potraktował bardzo skrótowo, poświęcając im sumarycznie zaledwie kilka akapitów. Należało odwrócić rozkład uwagi i własne, współautorskie prace badawcze, potraktować bardziej szczegółowo, omawiając uzyskane tam wyniki.

Kolejne dwa rozdziały, to kolejne teoretyczne rozważania dotyczące technik zespołowych w analizie anomalii oraz wskaźników skuteczności identyfikacji anomalii. Do obu rozdziałów mam podobne zastrzeżenia, jak w przypadku poprzednich. Choć są to ważne zagadnienia z punktu widzenia rozprawy, szczegółowość ich omówienia wykracza poza zakres takich opracowań.

W rozdziale siódmym opisano szczegóły implementacyjne aplikacji wykorzystującej opracowany system Trinity SALT oraz scenariusze jej używania. Opis ten wzbogacono zrzutami z ekranu zawierającymi poszczególne okna aplikacji. Weryfikacja opracowanej metody zespołowej została zrealizowana za pomocą eksperymentów z wykorzystaniem zbiorów danych o różnej charakterystyce. Opis zbiorów danych, uzyskane wyniki wraz z ich analizą zamieszczono w rozdziale ósmym. W ostatnim rozdziale Autor przedstawia krótkie podsumowanie osiągnięć i proponuje dalsze kierunki prac.

Analizując zawartość rozprawy, można stwierdzić, że większość oczekiwanych treści została w niej przedstawiona, czasem w nadmiarze. W moim odczuciu, o czym wspomniałam wcześniej, rozprawa zawiera wiele informacji, o niepotrzebnym poziomie szczegółowości. Trudno się zgodzić ze zdaniem Doktoranta: *Obszerność pracy jest uzasadniona kompleksowym podejściem do problematyki detekcji anomalii w złożonych zbiorach danych.* Nadmiar informacji przekazany w rozprawie utrudnia skupienie się na rozwiązaniach będących jej podstawą. Natomiast duży niedosyt pozostaje w zakresie opisu

badani prowadzonych przez Doktoranta. W rozprawie zabrakło również analizy odnoszącej się do wyników osiąganych przez zespołowe metody wykrywania anomalii prezentowanych w literaturze.

Rozprawa została napisana poprawnym językiem, z niewielką liczbą błędów interpunkcyjnych, głównie w zwrotach „taki jak”. Pojawiają się również stwierdzenia: *tabela/rysunek przedstawia*, zamiast *w tabeli/na rysunku przedstawiono*, czy słowo *biasów*, które należy potraktować jako żargon. Również określenie *metodologia* odnosi się raczej do nauki o metodach; w rozprawie lepszym byłoby użycie *metoda* czy *metodyka*.

3. Oryginalne osiągnięcia

Do najważniejszych osiągnięć rozprawy, decydujących o jej wartości teoretycznej i praktycznej, a które wyróżniają ją spośród dostępnych w literaturze przedmiotu, zaliczyć można:

- a) Zaproponowanie autorskiej definicji anomalii.
- b) Modyfikację algorytmu LOF polegającą na wprowadzeniu równoległego przetwarzania zbioru danych w podziale na bloki. Dzięki zastosowanemu podejściu uzyskano poprawę skuteczności metody LOF w wykrywaniu anomalii i skrócenie czasu przetwarzania danych.
- c) Opracowanie metody automatycznego dopasowania rozmiaru bloku danych do określonych warunków realizacji eksperymentu, a w tym rozmiaru zbioru danych, jego struktury i dostępnej pamięci RAM.
- d) Opracowanie pakietu SOaCRaport w języku R umożliwiającego wykrywanie odchyłeń w dowolnym zbiorze danych przy użyciu metod LOF i COF, a następnie hierarchiczne grupowanie danych z odchyleniami i bez odchyłeń oraz porównanie jakości utworzonych grup.
- e) Opracowanie zespołowego podejścia do wykrywania anomalii, bazującego na trzech metodach różnego typu. Osiągnięto w ten sposób osłabienie problemu, jakim jest różna skuteczność metod w rozwiązywaniu różnych problemów.
- f) Zaproponowanie metryki dla integracji rezultatów uzyskanych w zespołowym wykrywaniu anomalii w postaci jednego, skonsolidowanego wyniku.
- g) Projekt i implementację aplikacji tworzącej środowisko dla wykorzystania nowej zespołowej metody wykrywania anomalii i dla uruchomienia każdej z jej metod składowych.
- h) Eksperymenty zaprojektowane dla oceny skuteczności nowego podejścia Trinity SALT oraz ich realizację z użyciem różnych typów zbiorów danych o różnej charakterystyce. Należy podkreślić liczbę i różnorodność zastosowanych zbiorów danych, co pozwoliło na uzyskanie przekrojowych analiz i generalizację wyników.

Silną stroną rozprawy jest niewątpliwie całościowe podejście do problemu, czyli opracowanie nowej metody, wstępne zweryfikowanie jej na różnorodnych zbiorach danych, a następnie przygotowanie aplikacji do szerokiego wykorzystywania, którą każdy badacz może uruchomić poprzez stronę internetową i użyć do swoich analiz.

Potwierdzeniem osiągnięć Doktoranta jest dorobek publikacyjny, na który składa się 8 publikacji, z czego dwie z wyznaczonym współczynnikiem wpływu (IF) o sumarycznej wartości 9,3, znajdujących się na liście ministerialnej (100 i 200 pkt. MNiSW) oraz cztery publikacje konferencyjne z listy ministerialnej: trzy z nich to 70 pkt. MNiSW a jedna 20 pkt. MNiSW. Siedem publikacji indeksowanych jest w bazie Scopus, z czego jedna, najnowsza, nie znajduje się w spisie bibliograficznym rozprawy. Trzy publikacje widnieją w bazie Web of Science, wszystkie zostały ujęte w spisie bibliograficznym rozprawy.

W bazie Scopus Indeks Hirscha Doktoranta wynosi 4, a w Web of Science 2. W trzech pracach Doktorant jest pierwszym autorem, włączając to publikację za 200 pkt ze współczynnikiem wpływu 7,2.

4. Analiza źródeł

Autor rozprawy wykazał się dobrą znajomością dorobku literaturowego dotyczącego zagadnień, którym rozprawa jest poświęcona. Podstawą oceny literatury jest spis liczący 436 pozycji, w tym pozycje stanowiące dokumentację używanych narzędzi. Doktorant uwzględnił prace odnoszące się zarówno do najnowszych badań w tematyce rozprawy, jak i prace dokumentujące rozwój metod sztucznej inteligencji i eksploracji danych na przestrzeni ostatnich czterdziestu lat. Pojawiły się pozycje nawet sprzed 100 lat. Taki dobór literatury potwierdza dostateczną wiedzę Doktoranta w obszarze badawczym rozprawy doktorskiej i należy ocenić go pozytywnie.

5. Wady i słabe strony rozprawy, uwagi dyskusyjne

Chociaż cele prace zostały zrealizowane, a postawiona teza została udowodniona, rozprawa posiada również swoje słabe strony. Do najpoważniejszych z nich należy brak bezpośredniego udokumentowania oryginalności osiągnięć i odniesienia zaproponowanego rozwiązania oraz uzyskanych wyników do innych prac naukowych. Brakuje wskazań, jakie luki w wykrywaniu anomalii zostały wypełnione.

Drugą słabością rozprawy, o której wspomniano wcześniej, jest nadmiar zbędnych treści, częste powtórzenia, które powodują wydłużenie czasu na zrozumienie i ocenę rozwiązań. Dla przykładu podanych zostanie tylko kilka z wielu występujących w pracy:

- a) Str. 70, „test Kołmogorowa-Smirnowa (K-S) [133, 134] – używany do porównywania rozkładu próbki z rozkładem teoretycznym lub do porównywania dwóch próbek między sobą” i zdanie dalej „Służy do oceny zgodności próbki z określonym rozkładem teoretycznym lub do porównania, czy dwie próby pochodzą z tego samego rozkładu”.
- b) Str. 150, „SOM należy do tradycyjnych metod uczenia maszynowego i nie stosuje wielowarstwowej architektury do ekstrakcji cech.” I kilka zdań dalej: „SOM to osobna grupa struktur sieci neuronowych i metod uczenia, w której nie ma warstw neuronów.”
- c) Str. 155, „Tematyka algorytmu SOM była omawiana wraz ze współautorem w artykułach [264, 286]. W pierwszym artykule skupiono się na porównaniu dwóch algorytmów wykrywania anomalii: LOF i SOM” A na str. 170, „Artykuł [264] zajmuje się problemem wykrywania anomalii przy użyciu algorytmu samoorganizujących się map SOM. W badaniu porównano skuteczność wykrywania anomalii przy użyciu algorytmu LOF z wynikami uzyskanymi za pomocą algorytmu SOM, ...”
- d) Wielokrotne powtarzanie: zalet zespołowego wykrywania anomalii w tym Trinity SALT (str. 178, 183, 303, 304), opisu architektury systemu Trinity SALT (str. 234, 238, 251, 253), wyjaśnienie jego akronimu (147, 238, 288, 289), wyjaśnienie pojęcia stackingu (179, 254, 288, 290). Takich przykładów można mnożyć znacznie więcej.

Ponadto w rozprawie są pewne zagadnienia wymagają dodatkowego wyjaśnienia:

- a) Str. 94 – w opisie algorytmu FindFPOF pojawia się pojęcie transakcji w bazach danych – czy Autor miał na myśli zestaw operacji na danych zawartych w bazie danych, np. z wykorzystaniem języka SQL?
- b) Str. 95 – we wzorze (3.6) w mianowniku znajduje się indeks k – czy jest to liczba komponentów?
- c) Str. 153 – *Zbiór ten zawiera n obiektów z których każda jest opisana przez p cech* – chodziło o „z których każdy jest opisany ...”?
- d) Str. 164, 165, 166 – w podpisach pod rysunkami 4.12, 4.13, 4.14 znajduje się komentarz: *Źródło: Opracowanie własne*. Tymczasem w Internecie można znaleźć identyczne ilustracje. Czy zatem komentarz nie powinien brzmieć „*Źródło: Opracowanie własne, na podstawie ...*”?
- e) Str. 165 – wzór (4.21) zawiera formułę $\{ \max\{d_k(q), d(X_i, q)\} \}$, natomiast w opisie, w linii poniżej (który nazwami zmiennych nie odpowiada wzorowi) oraz z podpisu pod rysunkiem 4.14 wynika, że w pierwszym członie funkcji \max powinno chyba być X_i .
- f) Str. 167 – *„Natomiast punkt X z wartością LOF znacznie większą od 1 ma mniejszą gęstość sąsiedztwa niż jego k -najbliżsi sąsiedzi, co może wskazywać, że jest anomalią”* – co oznacza znacznie wyższą – dwa razy, czy rząd wielkości?
- g) Str. 235 – zawiera następujące zdanie *„Jednakże, użycie różnych metryk odległości stanowi rozszerzenie oryginalnej koncepcji, które nie zostało opisane w pracy.”* Dlaczego pomięto ten fakt w rozprawie, skoro w optymalizacji rozmiaru bloku takie opcje są możliwe do wybrania?
- h) Str. 241 – rysunek 7.4 błędnie został nazwany diagramem przypadków użycia. Jest to raczej diagram przepływu.
- i) Str. 257 – w różnych zdaniach podawany jest rozmiar bloku – natomiast na tym etapie czytelnik, nie wie, w jakich jednostkach mierzymy ten rozmiar. Wzór (8.3) podaje nowe podejście, ale poprzednie nadal jest nieznanne.
- j) Str. 262 – w tabeli 8.1 zaprezentowano wyniki eksperymentów dla różnych zbiorów danych. W jaki sposób dobierany był blok danych do eksperymentów?
- k) Str. 266 – w opisie wzoru (8.2) pojawia się informacja, że: *Wagi poszczególnych składowych funkcji celu zostały starannie dobrane, aby dokładnie odzwierciedlać istotę analizy i znaczenie różnych czynników.* W jaki sposób dokonano tego wyboru?
- l) Str. 269 – we wzorze (8.4) liczba n ogranicza nam rozmiar bloku do liczby obiektów, czy drugi człon funkcji \min jest wielokrotnością wielkości obiektu?
- m) Str. 272 – dla zbioru *Labeled Vehicle Claims* napisano takie zdanie: *Zbiór został przefiltrowany, by skupić się na istotnych anomaliach.* W jaki sposób zdefiniowano **istotne** anomalie?
- n) Str. 274 – dla zbioru *w7a libsvm* stwierdzono, brak potrzeby *imputacji* – skoro w przypadku tego zbioru wyraźnie to podkreślono, czy w innych zbiorach była potrzebna?
- o) Str. 280 – wyrażenia: *szybszy czas przetwarzania* oraz *zbiór skrócony w tabelach*, powinny raczej brzmieć: *krótszy czas ...* oraz *nazwa zbioru skrócona...*
- p) Str. 287 – danymi wyjściowymi algorytmu 5 są wyniki wykrytych anomalii w zbiorze treningowym i testowym, czy te wyniki wyprowadzane są z podziałem na algorytm, czy jest to jedna wspólna lista, czy jedna liczba?
- q) Str. 289 – w algorytmie 6 zastosowano 301 iteracji – brakuje wyjaśnienia, w jaki sposób ta liczba została wybrana i czy jest ona kluczowa dla uzyskanych wyników?

- r) Str. 289 – w algorytmie 6 w krokach 10 i 11 obliczane są wyniki dla zbioru treningowego. Dane te nie znalazły się w parametrach wyjściowych algorytmu. Czy są do czegoś wykorzystywane?
- s) Str. 293 – w stosunku do różnych zbiorów danych zaznaczono różny podział na dane treningowe i testowe: 70%-30% oraz 50%-50%. Nie wyjaśniono przyczyny tak zróżnicowanego podziału.
- t) Str. 293 – zbiór danych *Thyroid Disease* został okrojony i do badań wybrano tylko jego podzbiór. Jakie było kryterium tego wyboru i czy inny wybór mógł mieć wpływ na uzyskiwane rezultaty?
- u) Str. 296 – w opisie eksperymentów pojawia się zdanie „*Na początku wczytywany jest badany plik CSV, co pokazano na rysunku 8.17. Na tym etapie ocenia się skuteczność detekcji anomalii w zbiorach szkoleniowych i testowych, zarówno dla pojedynczych algorytmów, jak i zespołu*” Czy plik ten zawiera już wyniki wykrywania anomalii, czy tylko dane, które posłużą do tej detekcji? Nie jest to oczywiste, porównując ten opis i ten z rozdziału 7.

Uwagi o charakterze dyskusyjnym

- a) W rozprawie zabrakło analizy i dyskusji kluczowych ograniczeń zaproponowanych metod oraz odniesienia uzyskanych wyników, nie tylko do własnych badań lecz również do wcześniejszych prac prowadzonych w tym obszarze.
- b) Metryka MNA jest jednym z kluczowych elementów mających wpływ na wyniki uzyskiwane przy wykrywaniu anomalii. Jednak w pracy nie podjęto dyskusji, dlaczego taką metrykę zastosowano spośród dużej liczby różnych podejść, opisanych w rozdziale 5.5. W podsumowaniu rozdziału 5. nie znajdujemy uzasadnienia wyboru metod do zdefiniowania metryki oraz argumentów dla odrzucenia innych. Nie wskazano również metody doboru bonusów przyznawanych za zgodność wyników dla poszczególnych składowych metody zespołowej.
- c) Podobnie rzecz się ma w zakresie wskaźników skuteczności wykrywania anomalii. Wskazane byłoby uzasadnienie, dlaczego z bardzo dużej liczby opisanych w rozprawie metryk, wybrano czułość i F1, a inne odrzucono.
- d) W związku z tym, że dla obliczenia metryki MNA wybiera się maksymalny, z uzyskanych przez zastosowane algorytmy, wyników, rodzi się pytanie, czy na podstawie przeprowadzonych badań można wskazać rekomendacje, co do skuteczności poszczególnych metod systemu Trinity SALT dla zbiorów danych o określonej charakterystyce.

6. Wniosek końcowy

W podsumowaniu stwierdzam, że rozprawa doktorska mgr. inż. Czesława Horynia, mimo pewnych uwag krytycznych, **spełnia wymagania stawiane kandydatom do stopnia naukowego doktora**, określone w art. 187 ust. 1 i 2 Ustawy z dnia 20 lipca 2018r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2020r. poz. 85, z późn. zm.). W mojej ocenie zawiera **oryginalne rozwiązanie problemu naukowego**. Doktorant osiągnął postawione cele, wykazując się niezbędną wiedzą teoretyczną w dyscyplinie „Informatyka” i umiejętnościami w zakresie prowadzenia badań naukowych z wykorzystaniem metod informatycznych. **Wnoszę zatem o dopuszczenie jej do publicznej obrony.**

