**Р Е Ц Е Н З И Я**

на дисертационен труд

за придобиване на образователна и научна степен „**доктор**“

Автор на дисертационния труд **ас. маг. инж. Николина Йорданова Иванова**

Тема на дисертационния труд: „**Изследване на подходи за унифициране на големи данни от различни екологични източници с цел усъвършенстване на техния анализ“**

Рецензент: **проф. д-р Даниела Асенова Гоцева**

Катедра Компютърни системи, ФКСТ, Технически Университет – София

Настоящата рецензия е изготвена в качеството ми на член на научното жури, назначено със Заповед № РД-13-32/19.06.2025 г. на Ректора на Минно-геоложки Университет „Св. Иван Рилски“ гр. София.

Докторантката ас. маг. инж. Николина Йорданова завършва бакалавърска степен по специалност „Компютърни технологии в инженерната дейност“ през 2018 г. във филиала на Минно-геоложки университет „Св. Иван Рилски“ в Кърджали. През 2019 г. придобива магистърска степен по специалност „Компютърни технологии в инженерната дейност“ в Минно-геоложки Университет „Св. Иван Рилски“ в гр. София. От 2022 г. ас. маг. инж. Николина Йорданова е зачислена в редовна докторантура в професионално направление 5.13 Общо инженерство към катедра „Математика и информатика “ в Минно-геоложки Университет „Св. Иван Рилски“ гр. София. Докторантката е изпълнила индивидуалния си план и положила всички изпити от докторантския минимум. С решение на ФС на МЕМФ (Протокол № 17/17.06.2025 г.) по предложение на катедра „Математика и информатика “ (Протокол №14/11.06.2025 г.) е взето решение за разкриване на процедура за публична защита на дисертационния й труд.

Паралелно с научната си дейност, от 2020 г. насам, докторантката е асистент във филиала на Минно-геоложки Университет в гр. Кърджали, а от 2024 г. досега е зам. Директор АСД в същия филиал.

1. **Актуалност на разработвания в дисертационния труд, проблем в научно и научно-приложно отношение**

Представеният дисертационен труд се фокусира върху проблемите, свързани с унифицирането, запълването на липсващи стойности и последващия анализ на данни от реални, практически значими източници, в контекста на Big Data. Тези данни произхождат от пречиствателната станция за отпадни води (ПСОВ) и метеорологичната станция на НИМХ в гр. Кърджали, като представляват типичен пример за несъвместими данни от различно естество - по своята природа, формат, честота на наблюдение и източник. Темата безспорно е една от най-актуалните за индустрията.

За постигане на целите на изследването е изградена цялостна методология за предварителна обработка на данните, включваща почистване на стойности, стандартизация на параметри, синхронизация по времеви интервали и отстраняване на структурни несъответствия между различните източници. Унифицирането и анализът се разглеждат като взаимносвързани етапи от процеса на извличане на стойност от големите данни. Ключов акцент се поставя върху въпроса с липсващите стойности - едно от най- сериозните предизвикателства при анализа на реални данни. В рамките на разработката са проучени, приложени и сравнени различни подходи за импутация, включително интерполационни и авторегресивни методи. Тяхната ефективност е оценена количествено чрез използване на метрики като MAE, RMSE и MAPE, с цел да се идентифицира най-надеждният модел в условията на екологични и метеорологични данни.

Така формулираният подход позволява изграждането на последователен и адаптивен аналитичен модел за обработка на непълни и хетерогенни данни. Предложената методология демонстрира своята приложимост в реални условия и може да послужи като основа за бъдещи изследвания в областта на екологичния мониторинг, управлението на водни ресурси и интегрираните информационни системи за устойчиво развитие.

1. **Степен на познаване състоянието на проблема и творческа интерпретация на литературния материал**

Целта на дисертационния труд е насочена към разработването и прилагането на методологичен подход за предварителна обработка, унифициране и анализ на големи обеми хетерогенни данни от различни източници, с акцент върху оперативни данни от пречиствателни станции за отпадни води и метеорологични наблюдения. Изследването цели да се разкрият статистически зависимости между технологични и климатични фактори, които биха могли да подпомогнат управлението на пречиствателните процеси. Поставените пред дисертационния труд задачи изискват задълбочено проучване и познаване на модели и схеми за съхранение на големи хетерогенни данни, както и за методи и алгоритми, използвани в индустрията за обработката на същите данни. В рамките на работата на докторантката са цитирани 97 литературни източника, посочени в библиографската справка, които свидетелстват за извършения задълбочен анализ и запознаване със спецификата на материята, към която се фокусира дисертационния труд. Разгледани са множество проучвания, научни изследвания, сравнителни анализи и резултати от проведени международни конференции. Това свидетелства за отлично познаване на проблематиката и предметната област.

1. **Съответствие на избраната методика на изследване с поставената цел и задачи на** **дисертационния труд**

Избраната методика за провеждане на научните изследвания се базира на сравнителен и функционален анализ, практически изследвания и анализ на текущи предизвикателства, на основата на които е направен обоснован избор на средства за решаване на поставените задачи.

Обработката на данните започва с унифициране, включващо почистване на стойности, стандартизация на параметри, синхронизация по времеви интервали и отстраняване на структурни несъответствия между различните източници. След извършеното унифициране и запълване на първичните данни, изследването се насочва към прилагане на аналитични методи за разкриване на зависимости между показателите. Извършен е корелационен и крос-корелационен анализ, чрез които се изследват връзките между метеорологичните условия и показателите за качество на водата в пречиствателната станция. Приложен е и факторен анализ, който позволява намаляване на размерността и извеждане на скрити (латентни) фактори, обединяващи сходни по поведение показатели. Получените факторни оценки се използват за статистическо тестване на хипотези, свързани с причинно-следствените зависимости в наблюдаваната система. Така формулираният подход позволява изграждането на последователен и адаптивен аналитичен модел за обработка на непълни и хетерогенни данни.

Предложената методология демонстрира своята приложимост в реални условия и може да послужи като основа за бъдещи изследвания в областта на екологичния мониторинг, управлението на водни ресурси и интегрираните информационни системи за устойчиво развитие.

Дисертационния труд следва логическа последователност, проведените научни изследвания са напълно съобразени с предложената методика за провеждане на изследвания и експерименти, като успешно са изпълнени поставените цели и задачи преди дисертационния труд.

1. **Кратка аналитична характеристика и оценка на достоверността на материала, върху който се градят приносите на дисертационния труд**

Дисертационният труд съдържа увод, четири глави, заключение, списък с приноси, списък с публикации на докторанта по дисертационния труд, библиографска справка на използваните литературни източници и три приложения. Общият обем на дисертационния труд е 145 страници. Дисертационният труд съдържа 28 таблици и 26 фигури. Цитирани са 97 литературни източника, посочени в библиографската справка.

Първа глава на дисертационния труд предоставя общ преглед на концепцията за Big Data, като разглежда нейните основни характеристики и типовете данни, които обхваща. Описани са основните предизвикателства при анализа на големи данни, включително необходимостта от оценка на качеството на данните и тяхното унифициране за последващ анализ. Представени са съвременните аналитични подходи, приложими към комплексни и разнородни масиви от данни, като се подчертава значението на всеки от тях в контекста на Big Data. Главата завършва с преглед на основните техники и методи, използвани в анализа на данни.

Във втора глава е направено подробно описание на източниците на данни – Пречиствателната станция за отпадни води и метеорологичната станция към НИМХ в гр. Кърджали. Анализирана е структурата на данните, тяхната честота на измерване, характеристиките на наблюдаваните показатели, както и проблемите, свързани с дигитализацията и предварителната обработка. Главата представя сравнение на структурните различия между двата набора от данни и предприетите действия за осигуряване на съвместимост между тях.

Трета глава е изцяло концентрирана върху анализа и обработката на липсващите стойности в данните от ПСОВ. Разгледани са различни механизми на възникване на пропуски в данните, както и методите за тяхното откриване. Разработена е методология за запълване на липсващите стойности чрез няколко подхода, включително интерполационни и авторегресивни методи, приложени в среда MATLAB. Анализирани са резултатите от запълването чрез подходящи метрики за грешка (MAE, RMSE и др.), като е извършено експериментално сравнение на различни алгоритми по отношение на тяхната точност и стабилност.

Четвърта глава описва основните аналитични изследвания, проведени върху обработените и запълнени данни. Извършени са корелационен и крос-корелационен анализ между показателите от двата източника, с цел откриване на потенциални взаимовръзки. Основната част от главата разглежда прилагането на факторен анализ който позволява идентифициране на латентни зависимости и обединяване на наблюдаваните променливи в логически обосновани фактори. Представено е сравнение между модели с различен брой фактори, като е обоснован изборът на оптималната структура. В заключителната част, на база получените факторни оценки, се формулират и проверяват статистически хипотези, чрез които се изследва връзката между обобщените фактори и ключови показатели на входа и изхода на пречиствателната станция. и правила за валидация.

В заключението е представено обобщение на основните резултати. Формулирани са изводи и препоръки, както и насоки за бъдещи изследвания. Представен е списък с научни публикации по темата и забелязани цитирания.

От предоставеното изложение, поставените пред дисертационния труд задачи и извършените анализи и изследвания, може да се направи заключение, че докторантката добре познава тематиката и състоянието на разглежданите проблеми. На база извършения анализ на предизвикателствата пред ИТ сектора по отношение на проблеми свързани с управление и обработка на огромни хетерогенни масиви от информация свързани с ПСОВ и на база предложените и експериментално оценени методи за решаването на гореописаните предизвикателства са изведени изводи и оценки, които ми дават основание да определя получените резултати и свързаните с тях приноси като достоверни и практически полезни.

1. **Научни и/или научно-приложни приноси по дисертационния труд**

Приемам формулираните от докторантката приноси, постигнати в резултат на извършените научни изследвания. Приносите могат да се отнесат към категориите обогатяване на съществуващото научно знание и научни постижения в практиката, както и създаване на нови и модифициране на съществуващи методи, подходи и модели за решаване на поставените в дисертационния труд задачи.

Формулираните в дисертационния труд приноси, са групирани в три категории, както следва:

Приноси с научен характер

1. **Предложени са допълнителни критерии за усъвършенстване на интегрираната методология за унифициране на хетерогенни екологични и метеорологични** **данни**, включваща синхронизация на измерванията по дати, стандартизация на променливите и запълване на липсващи стойности чрез специализирани импутационни методи. Методологията е приложена успешно върху реални оперативни данни от ПСОВ и метеорологична станция и може да бъде адаптирана към други типове хетерогенни данни.
2. **Адаптирани са статистически подходи за извършване на факторен анализ в контекста на екологични системи**, чрез използване на метода на главните компоненти (PCA) и Varimax ротация. Извлечени са стабилни латентни фактори с ясно интерпретируемо съдържание - температурно-химичен, атмосферно-метеорологичен и хидравличен, което разкрива дълбочинно залегналата структура в сложни екологични данни.
3. **Разработена е експериментална процедура за сравнителна оценка на методи за запълване на липсващи стойности**, основана на зануляване на избрани интервали с известни стойности от данните и последващо възстановяване чрез PCHIP, SPLINE и авторегресивен модел. Точността на методите е оценена с метриките MAE, RMSE и MAPE, което осигурява обективен критерий за избор на подход при импутация.

Приноси с научно-приложено характер

1. **Валидирана е хипотеза за съществуването на отложени (лагови) зависимости между атмосферни и водни показатели**, чрез прилагане на крос-корелационен анализ. Установени са значими зависимости със закъснение от няколко дни, което обогатява разбирането на динамиката в пречиствателните системи и има потенциал за разработване на прогнозни модели.
2. **Създадена е аналитична база от зависимости между климатични и технологични променливи**, която позволява по-задълбочено моделиране на поведението на пречиствателната система при различни атмосферни условия и оперативни натоварвания.
3. **Предложен е подход за оптимизация на мониторинговата мрежа** **в ПСОВ**, чрез обобщение на факторни товари и анализ на обяснена дисперсия. Извлечената редукция е съставена от 15 променливи, които запазват 77,36% от общата обяснена дисперсия. Това позволява значително намаляване на броя на наблюдаваните показатели без съществена загуба на информация и създава предпоставки за оптимизация на измервателните мрежи в ПСОВ.
4. **Поставена е основа за прилагане на прогнозни модели, базирани на факторни оценки и атмосферни данни**, което би подпомогнало оперативното управление и би повишило ефективността на пречиствателните процеси чрез по-информирано вземане на решения.
5. **Предложени са оригинални методи за обработка и анализ на Big Data в контекста на ПСОВ**, чрез съчетаване на техники от статистическия анализ и машинното обучение, което разширява съществуващите изследвания и дава основа за разработване на практически приложими алгоритми за анализ на реални данни от непрекъснат мониторинг.
6. **Систематизирана и апробирана е методология за интегриран анализ на екологични и технологични данни**, която обединява процесите на унифициране, редукция на размерността чрез факторен анализ и проверка на хипотези. Тази рамка е приложима за широк спектър от изследвания и демонстрира как съчетаването на класически статистически техники с Big Data подходи може да доведе до значими интердисциплинарни резултати.
7. **Оценка на степента на личното участие на докторантката в приносите**

Представеното съдържание, проведените анализи и експерименти, в рамките на дисертационни труд, показват отличното познаване на предметната област, в която се фокусира труда. Дисертационният труд се откроява с отлична последователност по отношение на съдържанието, извършените анализи и проведените експерименти. Докторантката има четири научни публикации с анализи и резултатите получени в рамките на дисертационното си изследване, като три от публикациите са публикувана в годишника на Минно-геоложки Университет „Св. Иван Рилски“, а една от тях е докладвани на международна научна конференция. Считам, че личния принос на докторантката при постигане на поставените резултати и формулирането на приносите и безспорен.

1. **Преценка на публикациите по дисертационния труд**

В периода 2021-2024 докторантката е публикувала 4 научни статии с анализи и резултати получени в рамките на дисертационното изследване. Три от статиите са публикувани в годишника на Минно-геоложки Университет „Св. Иван Рилски“ гр. София, а четвъртата е докладвана на международна научна конференция CIEES и е публикувана в IEEE. Направените публикации покриват тематиката на дисертационни труд и отразяват основни теми, анализи и резултати от същия. Една от статиите е самостоятелна, останалите са в съавторство с други автори. Едната от статиите е с едно, а друга е с четири цитирания.

1. **Използване на резултатите от дисертационния труд в научната практика**

Резултатите постигнати в извършените по дисертационния труд изследвания са направени от докторантката, като някои от тях са подкрепени от проект „Прогнозни модели в минната промишленост, използвайки техники на Data mining и Big Data“, Договор МЕМФ-176/2023 г.

Постигнатите научни и приложни резултати могат да намерят приложение и при създаване на актуално учебно съдържание, което да бъде използвано в учебния процес със студенти. Резултатите също така могат да намерят приложение и в практиката, съобразно целта и задачите на дисертационни труд.

1. **Оценка на съответствието на автореферата с изискванията за изготвянето му, както и на адекватността на отразяване на основните положения и приносите на дисертационния труд**

Авторефератът към дисертационни труд отговаря на изискванията и съответства с дисертационния труд и поставените пред него цели и задачи. В авторефератът е разгледано съдържанието по глави и постигнатите приноси. Авторефератът е оформен съгласно изискванията на Минно-геоложки Университет „Св. Иван Рилски“ гр. София и считам че адекватно отразява същността на дисертационния труд.

1. **Мнения, препоръки и бележки**

Дисертационния труд изследва множество научни публикации и анализи свързани с предизвикателства в обработка и съхранение на хетерогенни големи данни в областта на ПСОВ. Трудът е оформен старателно и свидетелства за високопрофесионални познания на докторантката в областта. Извършените анализи и експериментални изследвания са проведени с голяма задълбоченост и желание за идентифициране на адекватно решение на поставените цели и задачи. От друга страна поставените цели и задачи целят решаване на реални проблеми и имат безспорно практическо приложение.

Препоръките ми към докторантката са да продължи своята научна работа и изследвания по темата, както и по други теми свързани с предизвикателства пред науката и индустрията, и да насочи усилия към внедряване на предложените методи и в други ПСОВ.

1. **Заключение**

На база гореизложеното, считам че дисертационния труд на ас. маг. инж. Николина Иванова напълно отговаря на изискванията Закона за развитие на академичния състав на Република България, Правилника за прилагане на закона и съответния правилник на Минно-геоложки Университет „Св. Иван Рилски“ гр. София и на изискванията за придобиване на образователната и научна степен „доктор“. Приложеният автореферат отговаря на изискванията и съответства с дисертационния труд и поставените пред него цели и задачи.

Проведените научни изследвания са в актуална област и третират проблематика, която е обект на засилен изследователски и индустриален интерес. Получените резултати съдържат научни и научно-приложни приноси, които са разпространени и направени достояние на заинтересованата научна общност чрез публикуването им в научни статии и представянето им на специализирани научни форуми и конференции.

**След като се запознах напълно с дисертационния труд и предметната област и на база гореизложеното, убедено давам своята положителна оценка за представения дисертационен труд и предлагам на уважаемите членове на Научното жури да бъде присъдена образователна и научна степен „доктор“ на ас. маг. инж. Николина Йорданова Иванова по научна специалност „Компютърни технологии в инженерната дейност“ в професионално направление 5.13 Общо инженерство.**

07.08.2025 г. Рецензент:

Гр. София /проф. д-р Даниела Гоцева/