**Р Е Ц Е Н З И Я**

от доц. д-р Мариана Трифонова Драганова, МГУ „Св. Иван Рилски“, София

върху дисертационния труд на **маг. инж. Николина Йорданова Иванова**,

представен за присъждане на образователна и научна степен „доктор“

по докторска програма „Компютърни технологии в инженерната дейност“,

професионално направление 5.13. „Общо инженерство“.

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

**Основание за представяне на рецензията:** решение от Първото заседание на Научно жури, утвърдено със заповед № РД-13-32/19.06.2025 г. на Ректора на МГУ „Св. Ив. Рилски“, състояло се на 30.06.2025 г.

# Информация за докторанта

Маг. инж. Николина Йорданова Иванова е зачислена за редовен докторант към катедра „Математика и информатика“ със заповед на Ректора на МГУ „Св. Ив. Рилски” № Р-86/14.02.2022 г. На 18.11.2024 г. с ректорска заповед № Р-797 е променен научния ръководител и за такъв е определен доц. д-р Веселин Христов.

Докторантката е положила успешно (среден успех отличен 5.85) предвидените в индивидуалния учебен план пет на брой изпити съгласно удостоверение № 378-2025 от 05.06.2025 г.

Отчислена е с право на защита със заповед на Ректора на МГУ „Св. Ив. Рилски” – № УД-16-19/04.04.2025 г.

Дисертационният труд е разгледан и обсъден на разширен катедрен съвет (съгласно Ректорска заповед № РД-13-28 от 28.05.2025 г.) на катедра „Математика и информатика“, проведен на 11.06.2025 г.

По процедурата за защита маг. инж. Николина Йорданова Иванова е представила всички необходими документи и материали в съответствие с изискванията на Закона за развитие на академичния състав в Република България, Правилника за неговото прилагане и Правилата и процедурите за приемане и обучение на докторанти и придобиване на ОНС „Доктор“ и научната степен „Доктор на науките“ на Минно-геоложки университет „Св. Иван Рилски“.

# Актуалност на темата

През последните години рязко нарасна обема от данни (структурирани и неструктурирани), генерирани от бизнес-процесите и свързаните с тях системи. Това определя и значението на инструментите и техниките, използвани за систематично извличане, обработка и анализ на получените данни с цел предоставяне на полезна информация и генериране на тенденции, което подпомага идентифициране на проблеми, оптимизация на процеси, вземане на решения и т.н.

Дисертационният труд, в контекста на Big Data, разглежда проблемите, свързани с унифициране на данните, особено хетерогенни данни, запълване на липсващи стойности в тях, както и последващия им анализ. Като основа за изследването служат данни от реални, практически значими източници. В случая тези данни са типичен пример за несъвместими данни.

# Обща характеристика на дисертационния труд

Представеният дисертационен труд се състои от 145 страници текст, включително 26 фигури и 28 таблици. Текстът е структуриран в увод, 4 глави, обобщени изводи и заключение, списък на приносите, списък на публикациите, забелязани цитирания по публикациите, списък на фигури и таблици, списък на използваните съкращения, библиография и три приложения. От общо 97 литературни източници, 6 са електронни, един е на български език и останалите 90 – на английски език.

Основната цел на дисертационния труд е насочена към разработване и прилагане 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.

За постигането на тази цел са дефинирани и успешно решени следните 5 задачи:

1. Изследване на характеристиките и проблемите при обработка и анализ на хетерогенни и непълни масиви, включващи екологични и метеорологични данни с различна резолюция, структура и качество.
2. Разработване на методология за унифициране на данни с цел постигане на съвместимост между различните източници. Проучване и сравнение на съществуващи методи и подходи за почистване и стандартизиране на данни, както и интеграцията им от различни формати и структури.
3. Изследване на методи за запълване на липсващи стойности. Прилагане и сравнение на различни импутационни техники, оценка точността на възстановените данни и тяхното влияние върху аналитичните резултати.
4. Избор на подходящи аналитични методи за обработка на унифицирани и запълнени данни: корелационен, крос-корелационен и факторен анализ, както и оценка на устойчивостта на изведените модели при различни условия.
5. Приложение на разработената методология върху реалните данни от ПСОВ и метеорологична станция – Кърджали с цел извличане на значими зависимости между климатичните и технологични показатели. Анализ на резултатите и оценка на приложимостта на методите в реална работна среда.

# Структура и съдържание на дисертационния труд

Дисертационният труд е изготвен прецизно в стройна логическа последователност.

**Уводът** кратко и ясно представя проблема, неговата актуалност и значимостта на темата. Формулирана е целта на дисертационния труд, както и задачите, които трябва да се решат за постигането й. Дадено е кратко резюме на последващите глави на дисертацията.

В **първа глава** е направен задълбочен теоретичен обзор на понятието Big Data: характеристики, източници на данни, качество на данните, видове анализи на данните и приложенията им в екологията. Акцентирано е върху действията, свързани с осигуряване на надеждност и съвместимост на данните, включително необходимостта от предварителна обработка, унифициране и валидация. Представени са различни подходи за обработка и анализ на данните: статистически методи, Machine Learning (машинно обучение), Data Mining (извличане на знания от данни), както и средствата за тяхното извършване – (Python, R, Excel, MATLAB, RapidMiner и др.). Задълбоченият преглед показва, че докторантката е навлязла в дълбочина в разглеждания проблем. В резултат на този обзор са изведени изводи, които са в основата на разработваната методология за анализ и интерпретация на използваните реални данни, представена в следващите глави на дисертацията.

**Втора глава** представя подробно описание на източниците на данни. Те са два: Пречиствателната станция за отпадни води и метеорологичната станция към НИМХ в гр. Кърджали. Извършена е огромна работа по дигитализация на данните. Анализирана е тяхната структура, честота на измерване, характеристики на наблюдаваните показатели и др. Анализът показва структурни различия между двата набора от данни, поради което са предприети действия за осигуряване на съвместимост между тях: осредняване на метеорологичните стойности, унифициране на форматите, синхронизация по дати, запълване на липсващите стойности. В резултат е създаден съвместим и надежден аналитичен масив, който осигурява стабилна основа за всички последващи анализи.

**Трета глава** е посветена на анализа и обработката на липсващи стойности. Дадена е класификация на различни подходи за справяне с липсващи данни като изтриване, импутация, моделиране на данните и ETL процеси. Направен е анализ на различни методи и механизми за запълване на липсите като тест на Литъл, логистична регресия, кубична интерполация на Ермит, сплайн-интерполация, авторегресионен модел. Анализът е направен с помощта на софтуерния продукт MATLAB. Предвид спецификата на данните, които съдържат липсващи стойности с различна продължителност и структура, по-подробно – количествено и графично са сравнени три подхода за запълване, а именно кубична интерполация на Ермит (PCHIP-интерполация), SPLINE интерполация и ARM (авторегресионен модел). В резултат на проведения анализ за запълване на липсите е избран авторегресионния метод, който в случая осигурява най-добро съотношение между точност и стабилност. По принцип изборът на подхода при предварителната обработка на данните е особено важен, за да няма изкривяване крайните резултати от изследването.

**Четвърта глава** съдържа 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. На база на тези резултати е направена предварителна редукция на първоначалния набор от 59 количествени показатели до 32 чрез елиминиране на взаимнозависими или слабоинформативни променливи. Върху така редуцирания набор е приложен факторен анализ, ориентиран към разкриване на зависимости между различните показатели и редуциране на размерността на данните без значителна загуба на информация. Разгледани са различни варианти на анализа, включващи модели с 2, 3, 4 и 5 фактора като броят на оптималните фактори за изследване и анализ е направен чрез критерий на Кайзер, Scree Plot и обяснена дисперсия. Анализът на моделите с повече от три фактора е показал, че добавянето на допълнителни фактори не подобрява значително анализа, тъй като се наблюдава значителна корелация между част от факторите, в резултат на което обосновано са избрани три от тях, които съвкупно обясняват 100% от дисперсията.

На база на резултатите от факторния анализ са формулирани и проверени 4 хипотези относно:

* влияние на температурно-химичния фактор като водещ регулатор на процесите в ПСОВ;
* влияние на атмосферните условия върху ПСОВ през преходните сезони пролет и есен;
* въздействие на хидравличния фактор върху концентрациите на азот и фосфор.
* влияние на редукцията на броя на наблюдаваните променливи върху достоверността на анализа и информационната цялост.

Особено значимо и полезно е потвърждаването на последната хипотеза, според която редуциран поднабор от 15 променливи, избрани чрез факторния анализ (по 5 на фактор), на база обединени данни от ПСОВ и метеорологичната станция на НИМХ, обяснява 77,36% от общата дисперсия. Това показва, че чрез подходящ избор на индикатори може да се постигне съществено опростяване на измервателния процес и намаляване на разходите за измервания без съществена загуба на точност при интерпретиране на резултатите.

Като цяло четвърта глава демонстрира ефективността и интердисциплинарната приложимост на избрания аналитичен подход, както и способността му да извлича значима, оперативно приложима информация от сложни екологични данни.

# Публикации по дисертационния труд, авторство на получените резултати

По дисертационния труд са представени четири публикации, три са в съавторство (от двама до четирима автори). Докторантката е на първо място в три от тях. Общият брой точки от четирите публикации е 41.67, което надхвърля минимума от 30 точки по показатели от Група В.

Публикациите отразяват основни резултати от изследванията, представени в дисертацията. Те са в издания от „отворен достъп“ и може да се счита, че е постигната необходимата публичност пред научно-инженерната общност.

Към момента са забелязани три цитирания на статиите. Самостоятелната статия е цитирана от преподаватели от Индонезия.

Авторефератът в обем от 32 страници отразява коректно и точно съдържанието на дисертационния труд, получените резултати и приноси.

Може да се направи изводът, че маг. инж. Николина Йорданова Иванова отговаря напълно на задължителните количествени наукометрични показатели за придобиване на ОНС „Доктор“.

# Приноси и значимост на дисертационния труд

Формулирани са 9 приноса, три научни и 6 научно-приложни. Съгласна съм по същество с формулировките, предложени от докторантката, но по-скоро 8 от тях класифицирам като научно-приложни.

Признавам изложените по-горе приноси като заслуги на докторантката, постигнати в процеса на дисертационното изследване.

Получените резултати потвърждават значимостта на разработката за практиката по отношение на устойчивото управление на водните ресурси.

Не съм установила плагиатство или друго неправомерно използване на чужди идеи в представената по конкурса разработка.

# Въпроси, критични бележки и препоръки

По същество нямам критични бележки и конкретни препоръки.

Целта и задачите на изследването са реализирани успешно и в пълен обем.

Дисертационният труд е оформен старателно, следва ясна логика, като всяка глава е последователно развита и свързана със следващата. Коректно са представени научните изследвания.

Маг. инж. Николина Йорданова Иванова е извършила значителна работа по темата на дисертацията, която е довела до убедителни научно-приложни резултати. Имам основание да смятам, че докторантката притежава задълбочени теоретични знания по разглежданата проблематика, умее да използва правилно литературния материал при провеждане на научни изследвания, при работа със специализирани програмни продукти, при анализ и тълкуване на постигнатите резултати, което показва способност за творческа интерпретация на зависимостите – необходимо и важно качество за един научен работник.

# Заключение

След като се запознах с дисертационния труд на тема „Изследване на подходи за унифициране на големи данни от различни екологични източници с цел усъвършенстване на техния анализ“ и автореферата към него, смятам, че те отговарят напълно на изискванията на Закона за развитие на академичния състав в Република България, Правилника за неговото прилагане, както и на Правилата и процедурите за приемане и обучение на докторанти и придобиване на ОНС „Доктор“ и научната степен „Доктор на науките“ на Минно-геоложки университет „Св. Иван Рилски“.

Докторантката е провела задълбочено изследване в актуална и приложно-значима област и е получила резултати, които съдържат съществени научно-приложни приноси, отразени в представените научни публикации. То показва, че маг. инж. Николина Йорданова Иванова притежава необходимите теоретични знания по тематиката, овладяла е нови методи и подходи за изследване, може да планира и провежда самостоятелно научни изследвания, да интерпретира и представя получените резултати.

***Въз основа на гореизложеното и представените материали давам положителна оценка на дисертационния труд и предлагам на уважаемите членове на Научното жури на маг. инж. Николина Йорданова Иванова да бъде присъдена образователната и научна степен „ДОКТОР“ в професионално направление 5.13. „Общо инженерство“, докторска програма „Компютърни технологии в инженерната дейност“.***

12.08.2025 г. Изготвил рецензията:

гр. София /доц. д-р Мариана Трифонова/