

Рішення
разової спеціалізованої вченої ради
про присудження ступеня доктора філософії

Здобувач ступеня доктора філософії: Земляний Олексій Дмитрович, 1998 року народження, громадянин України, освіта вища: закінчив у 2022 році Дніпровський національний університет імені Олеся Гончара за спеціальністю «Інженерія програмного забезпечення», навчається в аспірантурі Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України з 2022 р., виконав акредитовану освітньо-наукову програму «Інженерія програмного забезпечення».

Разова спеціалізована вчена рада, утворена наказом Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України, м. Дніпро, від «06» липня 2026 року № 248, у складі:

голови разової спеціалізованої вченої ради

- Володимира КОРЧИНСЬКОГО, доктора технічних наук, професора, професора кафедри телекомунікаційних систем та мереж Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України;

рецензентів:

- Ольги МАЦУГИ, кандидатки технічних наук, доцентки, доцентки кафедри інженерії програмного забезпечення та інформаційних технологій Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України;
- Світлани КЛИМЕНКО, кандидатки технічних наук, доцентки, завідувачки кафедри кібербезпеки та комп'ютерно-інтегрованих технологій Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України;

офіційних опонентів:

- Андрія ГУБСЬКОГО, кандидата технічних наук, доцента кафедри інформатики та програмної інженерії факультету інформатики і обчислювальної техніки Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» Міністерства освіти і науки України;
- Сергія ГОЛУБА, доктора технічних наук, професора, завідувача кафедри програмного забезпечення автоматизованих систем факультету інформаційних технологій і систем Черкаського державного технологічного університету Міністерства освіти і науки України;

на засіданні «27» серпня 2026 року прийняли рішення про присудження ступеня доктора філософії з галузі знань 12 Інформаційні технології Землянському Олексію Дмитровичу на підставі публічного захисту дисертації «Розроблення методів та програмного забезпечення інтелектуального імпутування пропусків даних» за спеціальністю 121 Інженерія програмного забезпечення.

Дисертацію виконано у Дніпровському національному університеті імені Олеся Гончара Міністерства освіти і науки України, м. Дніпро.

Науковий керівник: БАЙБУЗ Олег Григорович, доктор технічних наук, професор, завідувач кафедри інженерії програмного забезпечення та інформаційних технологій Дніпровського національного університету імені Олеся Гончара.

Дисертацію подано у вигляді спеціально підготовленого рукопису із дотримання вимог пункту 6 Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії, затвердженого постановою Кабінету Міністрів України від 12 січня 2022 року № 44 (зі змінами).

У дисертації вперше розроблено нові методи імпутування пропусків у даних `RegrImputer`, `HybridRegrEntropyImputer` та `CorrelationImputer`, а також запропоновано модифікацію методу `CorrelationImputer` для часових рядів із сезонною складовою, що передбачає окрему обробку даних за місяцями року. Удосконалено методи імпутування `NoNa` та `EntropyImputer` шляхом врахування типів даних, патернів пропусків, змінного порядку обробки ознак, а також застосування послідовних і паралельних процедур зменшення умовної ентропії. Набули подальшого розвитку методи перетворення якісних ознак на кількісні через розроблення модифікованих кодувальників `IgnoreNaNLabelEncoder` та `IgnoreNaNFrequentEncoder`, які зберігають інформацію про пропуски та підтримують зворотне перетворення даних. Запропоновані методи забезпечують підвищення точності імпутування пропусків, покращення якості моделей класифікації та прогнозування, а також зростання обчислювальної ефективності обробки даних.

Практичне значення отриманих результатів полягає в розробленні програмного забезпечення мовою Python відповідно до архітектурних принципів бібліотеки `scikit-learn`, що забезпечує інтеграцію запропонованих методів у стандартні конвеєри аналізу даних. Результати дослідження можуть бути використані для підвищення якості моделей класифікації та прогнозування в задачах медичного та гідрологічного моніторингу. Розроблені методи впроваджено в освітній процес Дніпровського національного університету імені Олеся Гончара під час викладання дисциплін, пов'язаних з аналізом даних, програмуванням та технологіями інтелектуальної обробки інформації.

Дисертація виконана державною мовою із дотриманням вимог до оформлення дисертації, встановлених МОН України. Обсяг основного тексту дисертації відповідає нормам, встановленим освітньо-науковою програмою «Інженерія програмного забезпечення» за спеціальністю 121 Інженерія програмного забезпечення Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України.

Здобувач має 11 наукових публікацій за темою дисертації, з яких 4 статті у наукових фахових виданнях України, 2 статті в інших збірниках наукових

Дисертацію виконано у Дніпровському національному університеті імені Олеся Гончара Міністерства освіти і науки України, м. Дніпро.

Науковий керівник: БАЙБУЗ Олег Григорович, доктор технічних наук, професор, завідувач кафедри інженерії програмного забезпечення та інформаційних технологій Дніпровського національного університету імені Олеся Гончара.

Дисертацію подано у вигляді спеціально підготовленого рукопису із дотримання вимог пункту 6 Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії, затвердженого постановою Кабінету Міністрів України від 12 січня 2022 року № 44 (зі змінами).

У дисертації вперше розроблено нові методи імпутування пропусків у даних `RegrImputer`, `HybridRegrEntropyImputer` та `CorrelationImputer`, а також запропоновано модифікацію методу `CorrelationImputer` для часових рядів із сезонною складовою, що передбачає окрему обробку даних за місяцями року. Удосконалено методи імпутування `NoNa` та `EntropyImputer` шляхом врахування типів даних, патернів пропусків, змінного порядку обробки ознак, а також застосування послідовних і паралельних процедур зменшення умовної ентропії. Набули подальшого розвитку методи перетворення якісних ознак на кількісні через розроблення модифікованих кодувальників `IgnoreNaNLabelEncoder` та `IgnoreNaNFrequentEncoder`, які зберігають інформацію про пропуски та підтримують зворотне перетворення даних. Запропоновані методи забезпечують підвищення точності імпутування пропусків, покращення якості моделей класифікації та прогнозування, а також зростання обчислювальної ефективності обробки даних.

Практичне значення отриманих результатів полягає в розробленні програмного забезпечення мовою Python відповідно до архітектурних принципів бібліотеки `scikit-learn`, що забезпечує інтеграцію запропонованих методів у стандартні конвеєри аналізу даних. Результати дослідження можуть бути використані для підвищення якості моделей класифікації та прогнозування в задачах медичного та гідрологічного моніторингу. Розроблені методи впроваджено в освітній процес Дніпровського національного університету імені Олеся Гончара під час викладання дисциплін, пов'язаних з аналізом даних, програмуванням та технологіями інтелектуальної обробки інформації.

Дисертація виконана державною мовою із дотриманням вимог до оформлення дисертації, встановлених МОН України. Обсяг основного тексту дисертації відповідає нормам, встановленим освітньо-науковою програмою «Інженерія програмного забезпечення» за спеціальністю 121 Інженерія програмного забезпечення Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України.

Здобувач має 11 наукових публікацій за темою дисертації, з яких 4 статті у наукових фахових виданнях України, 2 статті в інших збірниках наукових

праць та 5 тез доповідей у збірниках матеріалів наукових конференцій, які відповідають вимогам пунктів 8, 9 Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії, зокрема:

1. Земляний О.Д., Байбуз О.Г. Огляд методів інтелектуального аналізу даних та методів машинного навчання при прогнозуванні ішемічної хвороби серця. *Актуальні проблеми автоматизації та інформаційних технологій*, т. 27, Дніпро, 2023, с. 109–129. DOI: <http://dx.doi.org/10.15421/432311> URL: <https://actualproblems.dp.ua/index.php/APAIT/article/view/236> (фахове видання категорії «Б»).

2. Земляний О.Д., Байбуз О.Г. Методи імпутовання пропусків у даних про ішемічну хворобу серця. *Системні технології. Регіональний міжвузівський збірник наукових праць*, вип. 2(151), Дніпро, 2024, с. 33–49. DOI: <https://doi.org/10.34185/1562-9945-2-151-2024-04> URL: <https://journals.nmetau.edu.ua/index.php/st/article/view/1708/1008> (фахове видання категорії «Б»).

3. Земляний О.Д., Байбуз О.Г. Алгоритми імпутовання пропусків у даних на основі ентропії. *Системні технології. Регіональний міжвузівський збірник наукових праць*, вип. 6(155), Дніпро, 2024, с. 133–149. DOI: <https://doi.org/10.34185/1562-9945-6-155-2024-12> URL: <https://journals.nmetau.edu.ua/index.php/st/article/view/1920/1194> (фахове видання категорії «Б»).

4. Земляний О.Д., Байбуз О.Г. Імпутовання пропусків у даних гідрологічного моніторингу. *Актуальні проблеми автоматизації та інформаційних технологій*, т. 28, Дніпро, 2024, с. 147–160. DOI: <http://dx.doi.org/10.15421/432413> URL: <https://actualproblems.dp.ua/index.php/APAIT/article/view/258> (фахове видання категорії «Б»).

У дискусії взяли участь голова і члени спеціалізованої ради та присутні на захисті фахівці:

Голова спеціалізованої вченої ради – Корчинський Володимир Михайлович, доктор технічних наук (05.01.01 – Прикладна геометрія, інженерна графіка), професор, професор кафедри телекомунікаційних систем та мереж Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України. Зауважень немає.

Офіційний опонент – Губський Андрій Миколайович, кандидат технічних наук (05.13.07 – Автоматизація процесів керування), доцент кафедри інформатики та програмної інженерії факультету інформатики і обчислювальної техніки Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» Міністерства освіти і науки України.

Зауваження:

1. У першому розділі доцільно було б систематизувати результати аналізу існуючих методів імпутації та виокремити їхні переваги і недоліки. Це дозволило б більш обґрунтовано сформулювати постановку задач дослідження.

2. У другому розділі на сторінках, де описуються алгоритми імпутування, бажано надати більше деталей щодо вибору конкретних моделей машинного навчання (наприклад, чому саме вибрано ті чи інші класифікатори/регресори).

3. У розділі 3 доцільно було б обґрунтувати вибір архітектури scikit-learn для реалізації методів.

4. У розділі 4 доцільно було б надати пояснення, чому саме такі датасети було вибрано для оцінки ефективності запропонованих методів.

5. Автор розглядає кількісні та якісні ознаки, але було б цікаво дізнатися, чи запропоновані методи впораються з бінарними даними, у яких пропуски можуть мати специфічні патерни.

Офіційний опонент – Голуб Сергій Васильович, доктор технічних наук (05.13.06 – Інформаційні технології), професор, завідувач кафедри програмного забезпечення автоматизованих систем факультету інформаційних технологій і систем Черкаського державного технологічного університету Міністерства освіти і науки України.

Зауваження:

1. У тексті дисертації потрібно було б навести результати перевірки адекватності моделей, покладених в основу імплементацій, особливо в контексті їхнього застосування до різних типів даних (медичних, гідрологічних тощо).

2. Недостатньо розкрито процес проектування програмного забезпечення для нових методів імпутування. Додатковий опис етапів розробки, вибору архітектурних рішень та обґрунтування вибору мови програмування Python (наприклад, порівняно з іншими мовами, такими як C++ або R) допомогло б краще зрозуміти технічні рішення, прийняті автором.

3. Бажано було б детальніше описати метрики, використані для оцінки моделей. Наприклад, які саме метрики (окрім точності) застосовувалися для оцінки ефективності імпутування та їхнього впливу на кінцеві результати.

4. Не вказано, які інструменти використовувалися для проведення досліджень. Наприклад, можна було б описати, які інтегровані середовища чи спеціалізовані бібліотеки застосовувалися.

5. Не зрозуміло, які із рисунків дисертації відображають інформаційну технологію, яка розроблялася у цій роботі.

Рецензент – Мацуга Ольга Миколаївна, кандидатка технічних наук (05.13.06 – Автоматизовані системи управління та прогресивні інформаційні технології), доцентка, доцентка кафедри інженерії програмного забезпечення

та інформаційних технологій Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України.

Зауваження:

1. Дискусійним є підхід, за якого у підрозділі 1.2, де описано первинний аналіз даних гідрологічного моніторингу, значення одного календарного місяця за різні роки розглядаються як статистична вибірка для застосування критеріїв однорідності. Доцільно було б обґрунтувати коректність такого підходу з урахуванням можливої часової залежності даних та змін статистичних характеристик часових рядів.

2. Опис патернів пропусків на стор. 52 потребує уточнення та узгодження з алгоритмами 2.1–2.3. Зокрема, за буквального тлумачення пункти (a) та (b) можуть сприйматися як взаємовиключні: пункт (a) передбачає відсутність повністю заповнених рядків або стовпців, тоді як пункт (b) – наявність повністю заповнених стовпців. Водночас в описі алгоритмів ключовим етапом є формування підмножини `full_data`, яка складається з повністю заповнених рядків, та не цілком зрозуміло, яким чином враховується патерн (b). У зв'язку з цим доцільно уточнити зміст патернів (a)–(d) та детальніше пояснити, яким чином кожен із них враховується в наведених алгоритмах.

3. У роботі варто уточнити умови, за яких запропонований метод `RegImputer` є коректним для категоріальних ознак, а також пояснити, у яких випадках коригування прогнозованого регресійною моделлю значення забезпечує коректне відновлення категоріальної змінної.

4. Для запропонованого методу `CorrelationImputer` не цілком зрозуміло, яким чином метод працює у випадку, коли жодна з побудованих регресійних моделей не забезпечує прийнятної якості апроксимації. Також доцільно було б обґрунтувати вибір саме восьми квазілінійних моделей.

5. У підрозділі 3.2 переваги векторизації обчислень продемонстровано на прикладі окремого фрагмента програмного коду. Доцільно було б коротко охарактеризувати, наскільки аналогічні підходи були використані під час реалізації інших розроблених класів імпутерів.

6. У підрозділі 3.3 під час порівняння послідовного, багатопоточного та багатопроцесорного підходів доцільно було навести оцінку варіабельності часу виконання (наприклад, стандартне відхилення поряд із середнім), а також для забезпечення коректності порівняння проводити експерименти за однакових масок пропусків і фіксованого випадкового компонента алгоритму імпутації.

7. У підрозділах 3.3 та 4.3 швидкодію реалізованих методів варто було оцінювати окремо для кожного відсотка пропусків, а не лише узагальнено, оскільки обчислювальна складність алгоритмів може істотно залежати від кількості пропущених значень.

8. У підрозділах 4.3 та 4.5 у ході дослідження методів імпутування на даних гідрологічного моніторингу серед базових методів порівняння

використано методи SimpleImputer, IterativeImputer, KNNImputer, які не враховують часову структуру даних. У разі нестационарності часового ряду (тренду, структурних зсувів) застосування таких методів без попереднього дослідження стаціонарності може призводити до систематичних похибок імпутації, оскільки заповнене значення не враховує положення пропуску в часі. Тому доцільно було також порівняти запропоновані методи з підходами, що враховують часову структуру даних (наприклад, сезонну чи лінійну інтерполяцію за часом), або обґрунтувати, чому застосування методів, які не враховують часову структуру, є достатнім для поставленої задачі.

9. Для підтвердження переваги запропонованих методів над базовими доцільно було не лише порівнювати середні значення показників (RMSE, accuracy, recall тощо), а й навести оцінку їх варіабельності та застосувати статистичні тести (наприклад, тест знакових рангів Вілкоксона або, за виконання відповідних припущень, парний t-тест).

Рецензент – Клименко Світлана Володимирівна, кандидатка технічних наук (05.13.06 – Інформаційні технології), доцентка, завідувачка кафедри кібербезпеки та комп'ютерно-інтегрованих технологій Дніпровського національного університету імені Олеся Гончара Міністерства освіти і науки України.

Зауваження:

1. У дисертаційній роботі запропоновано методи імпутування UnifiedClassRegrImputer та RegrImputer, які ґрунтуються на використанні моделей класифікації та регресії. Разом з тим, опис процедури вибору конкретних моделей машинного навчання та критеріїв їх налаштування представлено недостатньо детально. Більш ґрунтовне обґрунтування вибору базових моделей та аналіз їх впливу на результати імпутування дозволили б повніше оцінити універсальність і можливості практичного застосування запропонованих підходів.

2. У підрозділі 2.5.4 (с. 57) автором зазначено про проведення рефакторингу програмного коду та його оптимізації. Водночас більш детальне висвітлення причин проведення рефакторингу, особливостей реалізації внесених змін та їх впливу на швидкість й масштабованість розробленого програмного забезпечення дозволило б повніше оцінити ефективність запропонованих архітектурних рішень.

3. У підрозділі 2.5 дисертації автор приділяє значну увагу питанням інтеграції розроблених методів до архітектури бібліотеки scikit-learn. Водночас представлення архітектури програмного забезпечення могло б бути доповнене UML-діаграмами класів або компонентів із відображенням взаємозв'язків між реалізованими імплементами, засобами кодування ознак та програмними інтерфейсами бібліотеки scikit-learn. Це сприяло б кращому розумінню особливостей реалізації та повторного використання програмних компонентів.

4. На сторінці 83 (табл. 3.2) наведено результати порівняння ефективності різних реалізацій ентропійного методу імпутування (послідовної, багатопотокової та багатопроцесної). Водночас потребують додаткового пояснення причини відмінностей в отриманих результатах для різних способів реалізації одного й того ж алгоритму. Доцільним було б більш детально проаналізувати вплив особливостей паралельних обчислень, використання генераторів випадкових чисел та програмної реалізації алгоритму на відтворюваність і точність отриманих результатів.

5. Автором виконано оптимізацію програмної реалізації запропонованих методів, зокрема шляхом проведення рефакторингу та векторизації програмного коду. Проте в роботі недостатньо уваги приділено опису проведених змін та кількісній оцінці їх впливу на підвищення продуктивності програмного забезпечення. Наведення більш детального аналізу ефективності застосованих підходів дозволило б повніше оцінити внесок виконаної оптимізації у загальну швидкодію розроблених алгоритмів.

6. У підрозділі 3.3 наведено результати оцінювання якості моделей класифікації та прогнозування після застосування розроблених методів імпутування. Разом із тим робота могла б виграти від більш детального аналізу обчислювальної складності запропонованих алгоритмів та їх масштабованості залежно від розмірності наборів даних. Такий аналіз дозволив би повніше оцінити перспективи використання розробленого програмного забезпечення при роботі з великими обсягами даних.

7. У розділі 3 наведено результати експериментальних досліджень ефективності запропонованих методів імпутування на даних медичного та гідрологічного моніторингу. Разом із тим було б доцільно більш детально обґрунтувати вибір параметрів експериментів, зокрема частки пропущених значень та механізмів їх генерації. Наведення результатів досліджень для різних сценаріїв виникнення пропусків (MCAR, MAR та MNAR) дозволило б повніше оцінити особливості застосування розроблених методів у практичних задачах.

8. У роботі реалізовано програмне забезпечення відповідно до архітектурних принципів бібліотеки scikit-learn та забезпечено можливість його інтеграції щодо обробки даних. Разом із тим у дисертації практично не розглянуто питання тестування розробленого програмного забезпечення. Наведення результатів модульного та інтеграційного тестування програмних компонентів дозволило б повніше оцінити їх надійність, коректність функціонування та готовність до практичного використання.

Результати відкритого голосування:

«За» – 5 членів ради,

«Проти» – немає.

На підставі результатів відкритого голосування разова спеціалізована вчена рада присуджує **ЗЕМЛЯНОМУ Олексію** ступінь доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення.

Відеозапис трансляції захисту дисертації додається.

Голова разової спеціалізованої
вченої ради



Володимир КОРЧИНСЬКИЙ