

ВІДГУК ОФІЦІЙНОГО ОПОНЕНТА

кандидата технічних наук, доцента, доцента кафедри інформаційних
технологій та комп'ютерної інженерії

Національного технічного університету «Дніпровська політехніка»

Каштан Віти Юріївни

на дисертаційну роботу **Пилипенка Владислава Ігоровича**
**«Моделі та методи прогнозування успішності здобувачів освіти на основі
машинного навчання»**, представлену на здобуття наукового ступеня
доктора філософії (PhD) в галузі знань 12 – Інформаційні технології, за
спеціальністю 122 – Комп'ютерні науки

1. Актуальність теми дисертації та її зв'язок з планами та напрямками науково-дослідних робіт наукових установ і організацій

Масштабна цифровізація вищої освіти та інтеграція платформ управління навчанням забезпечують безперервне накопичення цифрового сліду освітньої діяльності студентів. Це відкриває нові можливості для створення інтелектуальних систем підтримки прийняття рішень, орієнтованих на проактивний моніторинг та прогнозування результатів навчання. Раннє виявлення студентів із групи академічного ризику є важливим для оперативного педагогічного втручання та зниження рівня відрахувань.

Ефективне вирішення цього завдання методами машинного навчання стримується тим, що більшість існуючих підходів обмежуються аналізом статичних чинників, зокрема, поточних оцінок та відвідуваності. З іншого боку динамічні поведінкові характеристики взаємодії здобувачів із цифровим контентом (такими як, мультимедійні та відеоматеріали) залишаються недостатньо вивченими. Відтак, існує чітка науково-практична потреба в розробленні прогностичних методів і моделей на основі розширеного простору ознак цифрової активності студентів. Реалізація цього наукового підходу та обґрунтування відповідних рішень становлять основний зміст дисертаційної роботи Пилипенка Владислава Ігоровича.

2. Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни

Матеріали, що викладені в дисертації Пилипенка Владислава Ігоровича, мають високий рівень обґрунтованості та дозволяють зробити висновок про

результати його наукових досліджень. Використано сучасні методи досліджень, які включають в себе методи машинного навчання, глибокого навчання, математичне моделювання та методи математичної статистики. Вирішена важлива науково-практична проблема прогнозування академічної успішності студентів шляхом дослідження сучасних моделей машинного навчання та розробки нової моделі. Отримані результати можуть бути використані в якості основи для розробки інших моделей і методів прогнозу аналітики в освіті або впроваджені в інформаційно-комунікаційні системи управління навчанням різних типів.

Результати досліджень використано в дослідницькій діяльності кафедри «Комп'ютерних наук» Київського національного університету технологій та дизайну при використанні в межах науково-дослідної теми № 0122U200947, а також їх впроваджено в навчальний процес цього ж закладу вищої освіти та у навчальний процес Хмельницького національного університету технологій.

3. Структура, обсяг роботи

Робота є структурованою, добре обґрунтованою, що дає можливість зрозуміти проблематику роботи, зробити висновки про наукову новизну та практичну цінність. Робота написана українською мовою та відповідає стилю наукових досліджень з використанням термінів і понять, характерних для фахової та наукової термінології.

Дисертація загальним обсягом 211 сторінок, з яких 164 сторінки основного тексту. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел (116 найменувань) та 3 додатків. Робота містить 48 рисунків, 27 таблиць та 3 додатки, у які винесені фрагменти програмного коду та два акти впровадження.

Дисертаційна робота оформлена відповідно до вимог наказу МОН України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

4. Мова та стиль викладення результатів

Дисертаційна робота написана українською мовою, характеризується чіткістю викладу матеріалу, логічною послідовністю та використанням сучасної термінології галузі знань 12 Інформаційні технології за спеціальністю

122 Комп'ютерні науки. Зміст повністю узгоджується зі Стандартом вищої освіти для зазначеної спеціальності та тематично охоплює актуальний напрям машинного навчання для інтелектуальних інформаційно-комунікаційних систем. Матеріал викладено послідовно та зрозуміло. Результати дослідження належним чином ілюстровані таблицями, рисунками та результатами експериментальних досліджень, що сприяє їх сприйняттю та аналізу.

5. Новизна основних наукових положень, висновків і рекомендацій, сформульованих в дисертації

Дисертаційна робота та опубліковані роботи здобувача дозволяють визначити наукову новизну, а також практичне значення.

1. Вперше розроблено 2-рівневу стекінгову модель прогнозування академічної успішності здобувачів освіти, яка відрізняється від відомих моделей використанням на першому рівні ансамблю алгоритмів логістичної регресії, випадкового лісу та нейронної мережі, а на другому рівні – алгоритму градієнтного бустингу, що забезпечує підвищення достовірності прогнозування;

2. Удосконалено метод розширення простору ознак для задач прогнозування академічної успішності, за рахунок включення характеристик взаємодії здобувачів освіти з навчальними відеоматеріалами в LMS, що дозволило підвищити інформативність навчальних вибірок та достовірність прогнозування;

3. Удосконалено методи збору та підготовки даних для задач прогнозування успішності здобувачів освіти, що передбачає автоматизоване отримання і накопичення даних про взаємодію користувачів з навчальними відеоматеріалами та їх інтеграцією у процес побудови моделей машинного навчання;

4. Набули подальшого розвитку методи аналізу освітніх даних, зокрема щодо оцінювання впливу поведінкових характеристик роботи користувачів електронного освітнього середовища на результати прогнозування академічної успішності.

Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності

Методи та підходи, що були використані здобувачем у дослідженнях є такими, що відповідають задачі та відповідають проблематиці, а також дозволяють вирішувати поставлені завдання.

У першому розділі дисертації (с. 27–51) виконано аналітичний огляд наукових підходів та інформаційних технологій у сфері прогнозування успішності в системах дистанційного навчання, зокрема LMS Moodle. Визначено основні категорії освітніх даних та проаналізовано існуючі моделі прогнозування, що дозволило виявити недостатню дослідженість поведінкових характеристик взаємодії студентів із відеоматеріалами. На основі цього сформульовано метру, задачі, об'єкт і предмет дослідження, а також обґрунтовано доцільність розроблення ансамблевих моделей.

В другому розділі (с. 52–99) розглянуто сучасні алгоритми машинного навчання для задач класифікації та прогнозування академічної успішності, досліджено їхні характеристики. Проведено порівняльний аналіз моделей на основі логістичної регресії, наївного Баєса, методу опорних векторів, випадкового лісу та нейронних мереж на однаковому наборі даних. Отримані результати досліджено на предмет точності їхньої роботи за метриками Accuracy, Sensitivity, Specificity, F1-score та ROC-AUC, що дозволило визначити випадковий ліс та нейронні мережі як найбільш ефективні архітектури та обґрунтувати необхідність розширення ознакового простору.

Третій розділ (с. 100–141) присвячено роботі моделей машинного навчання з розширеним набором поведінкових ознак взаємодії студентів із відеонавчальним контентом, отриманих за допомогою розробленого плагіну VideoPlayer. В розділі проведено дослідження точності роботи методів, за результатами якого визначено, що додавання поведінкових характеристик забезпечує приріст загальної точності приблизно на 10%. Отримані дані застосовано для побудови запропонованої у розділі дворівневої ансамблевої стекінгової моделі на основі базових класифікаторів та мета-моделі градієнтного бустингу, що дало змогу підвищити точність роботи методів прогнозування академічної успішності до 90.2%.

В четвертому розділі (с. 142–179) розглянуто функціональну та програмну реалізацію системи збору й аналізу поведінкової освітньої аналітики, а також інтеграцію розробленого плагіну відеоплеєра в систему управління навчанням Moodle. Визначено принципи організації зберігання та обробки даних у межах реалізованої системи, в якій апробовано запропоновані методи та моделі машинного навчання. Отримані дані та розроблений програмний застосунок для формування аналітичної звітності дають змогу проводити моніторинг навчальної діяльності та завдяки цьому забезпечувати підтримку процесів аналізу результатів прогнозування.

У висновках (с. 180–181) наведено загальні результати що отримані в процесі досліджень.

6. Повнота викладення результатів досліджень в опублікованих працях

Результати досліджень опубліковано у 23 працях, зокрема: десять статей – у наукових фахових періодичних виданнях України; одна стаття – у наукових періодичних виданнях іншої держави (Scopus, Q3); дванадцять публікацій – у матеріалах міжнародних та всеукраїнських наукових, науково-технічних, науково-практичних конференцій (три із них входять до наукометричної бази Scopus).

Публікації здобувача містять опис проведених досліджень та їх результатів. Таким чином публікації повною мірою відображають матеріали дисертації.

7. Відповідність вимогам академічної доброчесності

Грунтовний аналіз тексту дисертації Пилипенка Владислава свідчить про відсутність порушення здобувачем вимог академічної доброчесності. У дисертації не виявлено ознак академічного плагіату, самоплагіату, фабрикації, фальсифікації чи інших порушень, що могло б поставити під сумнів самостійний характер виконаної роботи. У дисертації є відповідні посилання на інформаційні джерела у тих випадках, де використано результати досліджень інших авторів.

8. Недоліки та зауваження до дисертаційної роботи

1. У розділі «Особистий внесок здобувача» автор наводить загальний опис своєї ролі в опублікованих працях, виконаних у співавторстві. Проте для об'єктивної оцінки наукової самостійності дисертанта доцільно було б чітко розписати конкретний внесок автора окремо по кожній сумісній статті із прямим посиланням на відповідну працю з бібліографічного списку.

2. У тексті зазначено, що базові моделі мали «обмежену здатність до розпізнавання студентів групи ризику». Хоча стекінгова модель покращує загальні метрики, доцільно детальніше описати, за рахунок яких саме

математичних чи чинникових механізмів мета-модель компенсує помилки базових класифікаторів саме в розрізі виявлення слабких студентів (групи ризику), оскільки для практичного впровадження метрика чутливості (Sensitivity) до незадовільних результатів є важливим аспектом. Зокрема, залишається невисвітленим питання, чи застосовувалися в мета-моделі специфічні методи балансування класів або зважування функцій втрат для підвищення точності розпізнавання саме міноритарного класу "групи ризику", що є важливим для систем раннього попередження академічної неспішності.

3. Зміст роботи вказує на використання даних електронного журналу, бази даних LMS Moodle та розробленого плагіну VideoPlayer (кількість переглядів, пауз, перемотувань тощо). Проте залишається не до кінця зрозумілою безпосередня методологія формування фінального датасету.

4. В роботі не зазначено фінальний обсяг вибірки (кількість студентів та унікальних курсів), використаний для експериментів у Розділі 2 та Розділі 3, адже в анотації наведено лише відсоткові прирости точності, але бракує кількісних характеристик генеральної сукупності.

5. Автор досліджував вплив розміру навчальної вибірки на стабільність моделей та застосовував ансамблевий підхід для зменшення похибки узагальнення. Однак для складних моделей (таких як MLPClassifier та Градієнтний бустинг у стекінгу) ризик перенавчання на специфічних даних конкретного закладу освіти є дуже високим. Необхідно чітко специфікувати у тексті, які саме регуляризаційні методи (наприклад, L1/L2 регуляризація, обмеження глибини дерев у бустингу, крос-валідація за замовчуванням) застосовувалися для запобігання перенавчанню. Доцільно продемонструвати порівняння точності моделей на тренувальних і тестових вибірках для підтвердження відсутності оверфіттингу.

6. В роботі бракує аналізу того, чи змінюється вага ознак (наприклад, важливість перемотувань чи пауз) залежно від складності або спрямування навчального курсу.

Вище вказані зауваження мають окремий характер та не впливають на високий науковий рівень роботи, а також на розуміння практичної цінності проведених досліджень.

9. Висновок про дисертаційну роботу

Вважаю, що дисертаційна робота здобувача ступеня доктора філософії Пилипенка Владислава Ігоровича на тему «Моделі та методи прогнозування

успішності здобувачів освіти на основі машинного навчання», виконана на належному науковому рівні та є завершеним науковим дослідженням. Отримані результати характеризуються науковою новизною, теоретичною та практичною цінністю, відповідають спеціальності 122 Комп'ютерні науки та мають важливе значення для розвитку освітньої аналітики, систем підтримки прийняття рішень та інтелектуальних інформаційних систем.

Дисертаційна робота відповідає вимогам п. 6-9 «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 р. № 44, а її автор Пилипенко Владислав Ігорович заслуговує на присудження ступеня доктора філософії в галузі знань 12 Інформаційні технології за спеціальністю 122 Комп'ютерні науки.

Офіційний опонент

Кандидат технічних наук, доцент,
доцент кафедри інформаційних технологій
та комп'ютерної інженерії
Національного технічного університету
«Дніпровська політехніка»



Віта КАШТАН

Підпис Каштан В
засвідчую:
вчений секретар
Вченої ради
Галюшня І

