Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0823U101138

Особливі позначки: відкрита

Дата реєстрації: 06-11-2023

Статус: Наказ про видачу диплома

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Ямковий Клим Сергійович

2. Klym Yamkovyi

Кваліфікація:

Ідентифікатор ORCID ID: 0000-0001-9512-4150

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 122

Назва наукової спеціальності: Комп'ютерні науки

Галузь / галузі знань: інформаційні технології

Освітньо-наукова програма зі спеціальності: Комп'ютерні науки

Дата захисту: 20-12-2023

Спеціальність за освітою: Прикладна математика

Місце роботи здобувача: Національний технічний університет "Харківський політехнічний інститут"

Код за ЄДРПОУ: 02071180

Місцезнаходження: вул. Кирпичова, буд. 2, Харків, Харківський р-н., 61002, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): ДФ 64.050.103-2969

Повне найменування юридичної особи: Національний технічний університет "Харківський політехнічний інститут"

Код за ЄДРПОУ: 02071180

Місцезнаходження: вул. Кирпичова, буд. 2, Харків, Харківський р-н., 61002, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

IV. Відомості про підприємство, установу, організацію, в якій було виконано дисертацію

Повне найменування юридичної особи: Національний технічний університет "Харківський політехнічний інститут"

Код за ЄДРПОУ: 02071180

Місцезнаходження: вул. Кирпичова, буд. 2, Харків, Харківський р-н., 61002, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

V. Відомості про дисертацію

Мова дисертації: Українська

Коди тематичних рубрик: 28.23.29, 28.23.35, 20.54.06

Тема дисертації:

1. Інформаційні технології побудови композитних індикаторів на основі методів машинного навчання

2. Information technologies for building composite indicators based on a machine learning approach

Реферат/анотація:

1. У дисертаційній роботі вирішено науково-практичне завдання розробки методів та інформаційних технологій побудови композитних індикаторів на основі ядерних методів машинного навчання та оптимального узгодження експертної та статистичної інформації. Об'єкт дослідження – процеси побудови композитних індикаторів у задачах ранжування і багатокритеріального оцінювання та вибору. Предмет дослідження – методи та інформаційні технології побудови композитних індикаторів з на основі оптимального узгодження експертної і статистичної інформації та агрегації даних. Мета і задачі дослідження – розробка методів та інформаційних технологій побудови композитних індикаторів основі ядерних методів машинного навчання та оптимального узгодження експертної та статистичної інформації з метою підвищення точності отриманих моделей та обмеження їх складності. У вступі обґрунтовано актуальність

теми дисертаційної роботи, зазначено зв'язок роботи з науковими темами, сформульовано мету, задачі та цілі дослідження, визначено об'єкт, предмет та методи дослідження, показано наукову новизну та практичне значення отриманих результатів, наведено інформацію про практичне використання, особистий внесок здобувача, апробацію результатів дослідження та їх висвітлення у публікаціях. Приводяться відомості щодо структури та обсягу дисертаційної роботи. У першому розділі проведено аналіз задачі побудови композитних індикаторів та огляд різних підходів до їх побудови, зокрема, методів машинного навчання. Наведено приклади використання композитних індикаторів у багатьох сферах для побудови різних узагальнених показників: людського розвитку, екологічної ефективності, інвестиційного портфеля тощо. Обґрунтовано вибір мети і задач роботи. У другому розділі задача побудови композитного індикатора сформульована в термінах машинного навчання, отримано розв'язання задачі побудови нелінійної моделі композитного індикатору на основі ядерної рідж-регресії. Проаналізовано методи узгодження різнорідної експертної інформації, що дозволяють знайти компроміс між експертними оцінками композитних індикаторів та статистичними оцінками часткових показників. Обґрунтовано запропонований метод оптимального узгодження експертної і статистичної інформації за допомогою регуляризації ядерної регресії з використанням апріорної інформації щодо важливості часткових показників, що суттєво підвищує точність отриманих моделей. У третьому розділі наведено принципи концепції великих даних та описані проблеми, які виникають при збільшені обсягів інформації, що використовується для побудови композитних індикаторів. Запропоновано використовувати методи агрегування даних для зменшення складності ядерної моделі. Розглянуто методи групування та кластеризації для агрегації даних. Для підвищення точності та ефективності кластеризації запропоновано використовувати регуляризацію за допомогою цільової змінної на етапі розрахунку відстані між точками в просторі ознак, викладено запропонований метод регуляризованої кластеризації. Визначено проблему недостатньої розмітки даних, яка особливо часто виникає при збільшенні кількості даних. Для розв'язання цієї проблеми запропоновано використовувати методи напівкерованого навчання на основі графової регуляризації та ядерного трюку під час оптимізації нелінійної функції переваг. Для розв'язання зазначених проблем розроблено двоетапний алгоритм агрегації даних, який використовує як глобальні, так і локальні патерни у структурі набору під час агрегації. Такий підхід дозволяє значно скоротити розмір вибірки при збереженні всіх властивостей та шаблонів. У четвертому розділі наведено опис запропонованої інформаційної технології для побудови композитних індикаторів методами машинного навчання, яка реалізує розроблені у роботі методи та алгоритми. Розроблена інформаційна технологія реалізована у вигляді бібліотеки мовою програмування Python з відкритим похідним кодом та наслідуванням інтерфейсів бібліотеки scikit-learn та відповідає усім вимогам методологій розробки проєктів у сфері машинного навчання та аналізу даних, а саме KDD та CRISP-DM. Перевірено працездатність розробленої інформаційної технології, точність запропонованих алгоритмів та проаналізовано отримані результати дослідження. Для цього використано декілька багатомірних наборів даних, які представляють різні домені області. Результати показали працездатність та ефективність запропонованих у роботі методів та алгоритмів. У висновках наведено основні результати дисертаційної роботи щодо вирішення поставлених наукових задач дослідження.

2. The scientific and practical task of developing methods and information technologies for building composite indicators based on kernel methods of machine learning and optimal concordance of expert and statistical information is solved in the dissertation work. Research object – the processes of building composite indicators in the tasks of ranking and multi-criteria evaluation and selection. Research subject – methods and information technologies for building composite indicators based on optimal concordance of expert and statistical information and data aggregation. The purpose and objectives of the research – the development of methods and information technologies for building composite indicators based on kernel methods of machine learning and optimal concordance of expert and statistical information, to increase the accuracy of the obtained models and limit their complexity. The introduction substantiates the relevance of the topic of the dissertation, indicates the connection of the work with scientific topics, formulates the goal, tasks and objectives of the research, defines the object, subject and methods of the research, shows the scientific novelty and practical significance of the obtained results,

provides information about practical use, personal contribution of the recipient, the approbation of research results and their coverage in publications. Information on the structure and scope of the dissertation work is given. In the first chapter, an analysis of the task of constructing composite indicators and an overview of various approaches to their construction, in particular, machine learning methods, was carried out. Examples of the use of composite indicators in many areas for the construction of various generalized indicators are given: human development, environmental efficiency, investment portfolio, etc. The choice of the goal and tasks of the work is substantiated. The second chapter formulates the task of constructing a composite indicator in terms of machine learning, and a solution to the task of constructing a nonlinear model of a composite indicator based on kernel ridge regression is obtained. The methods of concordance of disparate expert information, which allow finding a compromise between expert assessments of composite indicators and statistical assessments of partial indicators, are analyzed. The proposed method of optimal concordance of expert and statistical information using kernel regression regularization with the use of a priori information on the importance of partial indicators is justified, which significantly increases the accuracy of the obtained models. The third chapter provides the principles of the concept of big data and describes the problems that arise when the amount of information used to construct composite indicators increases. It is proposed to use data aggregation methods to reduce the complexity of the kernel model. Methods of grouping and clustering for data aggregation are considered. To increase the accuracy and efficiency of clustering, it is proposed to use regularization with the help of a target variable at the stage of calculating the distance between points in the feature space, and the proposed method of regularized clustering is outlined. The problem of insufficient data marking, which especially often arises when the amount of data increases, is identified. To solve this problem, it is proposed to use semi-supervised learning methods based on graph regularization and kernel trick during the optimization of the nonlinear preference function. To solve these problems, a two-stage data aggregation algorithm was developed, which uses both global and local patterns in the set structure during aggregation. This approach allows to significant reduce the size of the sample while preserving all properties and patterns. The fourth chapter describes the proposed information technology for building composite indicators using machine learning methods, which implement the methods and algorithms developed in the work. The developed information technology is implemented in the form of a library in the Python programming language with open-source source code and inherits scikit-learn library interfaces and meets all the requirements of project development methodologies in the field of machine learning and data analysis, namely KDD and CRISP-DM. The functionality of the developed information technology, the accuracy of the proposed algorithms, and the obtained research results were analyzed. For this, several multidimensional data sets representing different domains of the area were used. The results showed the efficiency and effectiveness of the methods and algorithms proposed in the work. In the conclusions, the main results of the dissertation work on the solution of the set scientific research problems are presented.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки: Інформаційні та комунікаційні технології

Стратегічний пріоритетний напрям інноваційної діяльності: Розвиток сучасних інформаційних, комунікаційних технологій, робототехніки

Підсумки дослідження: Нове вирішення актуального наукового завдання

Публікації:

- K. Yamkovyi, “Adaptation of LambdaMART model to semi-supervised learning,” Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології, 2023, №1(9), с. 76—81. (Б)
- L. Lyubchyk, K. Yamkovyi, “Comparative Analysis of Modified Semi- Supervised Learning Algorithms on a Small Amount of Labelled Data,” System Research & Information Technologies, 2022, № 4, с. 34—43. (A, Scopus)

- O. Akhiiezer, G. Grinberg, L. Lyubchyk, K. Yamkovyi, “Failure rate regression model building from aggregated data using kernel-based machine learning methods,” Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології, 2022, № 2 (8), с. 51—56. (Б)
- K. Yamkovyi, “Development and comparative analysis of semi-supervised learning algorithms on a small amount of labeled data,” Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології, 2021, № 1 (5), с. 98—103. (Б)
- L. Lyubchyk, O. Akhiiezer, G. Grinberg, K. Yamkovyi, “Machine Learning-Based Failure Rate Identification for Predictive Maintenance in Industry 4.0,” 2022 12th International Conference on Dependable Systems, Services and Technologies (DESSERT), IEEE, Athens, Greece , 2022, с. 1—5. (Scopus, Греція)
- L. Lyubchyk, G. Grinberg, K. Yamkovyi, “Integral Indicator for Complex System Building Based on Semi-Supervised Learning,” 2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC), IEEE, Kyiv, Ukraine, 2018, с. 1—5. (Scopus, Україна)

Наукова (науково-технічна) продукція: методи, теорії, гіпотези; програмні продукти, програмно-технологічна документація

Соціально-економічна спрямованість: підвищення автоматизації виробничих процесів; забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг

Охоронні документи на ОПІВ:

Впровадження результатів дисертації: Впроваджено

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Любчик Леонід Михайлович

2. Leonid Lyubchyk

Кваліфікація: д.т.н., професор, 05.13.07

Ідентифікатор ORCID ID: 0000-0003-0237-8915

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет "Харківський політехнічний інститут"

Код за ЄДРПОУ: 02071180

Місцезнаходження: вул. Кирпичова, буд. 2, Харків, Харківський р-н., 61002, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Недашківська Надія Іванівна

2. Nadezhda Nedashkovskaya

Кваліфікація: д. т. н., доц., 01.05.04

Ідентифікатор ORCID ID: 0000-0002-8277-3095

Додаткова інформація:

Повне найменування юридичної особи: Навчально-науковий комплекс "Інститут прикладного системного аналізу" Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 25408067

Місцезнаходження: проспект Перемоги, буд. 37 , корпус 35, Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Машталір Сергій Володимирович

2. Sergii Mashtalir

Кваліфікація: д. т. н., професор, 05.13.06

Ідентифікатор ORCID ID: 0000-0002-0917-6622

Додаткова інформація:

Повне найменування юридичної особи: Харківський національний університет радіоелектроніки (ХНУРЕ)

Код за ЄДРПОУ: 02074197

Місцезнаходження: , 61166

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Чередніченко Ольга Юріївна

2. Olga Cherednichenko

Кваліфікація: д. т. н., професор, 05.13.06

Ідентифікатор ORCID ID: 0000-0002-9391-5220

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет "Харківський політехнічний інститут"

Код за ЄДРПОУ: 02071180

Місцезнаходження: вул. Кирпичова, буд. 2, Харків, Харківський р-н., 61002, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

Власне Прізвище Ім'я По-батькові:

1. Галуза Олексій Анатолійович
2. Oleksii Haluza

Кваліфікація: д. ф.-м. н., професор, 01.04.07

Ідентифікатор ORCID ID: 0000-0003-3809-149X

Додаткова інформація:

Повне найменування юридичної особи: Національний технічний університет "Харківський політехнічний інститут"

Код за ЄДРПОУ: 02071180

Місцезнаходження: вул. Кирпичова, буд. 2, Харків, Харківський р-н., 61002, Україна

Форма власності: Державна

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR:

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради	Шаронова Наталія Валеріївна
Власне Прізвище Ім'я По-батькові головуючого на засіданні	Шаронова Наталія Валеріївна
Відповідальний за підготовку облікових документів	Ямковий Клим Сергійович
Реєстратор	УкрІНТЕІ
Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності	Юрченко Тетяна Анатоліївна