Облікова картка дисертації

I. Загальні відомості

Державний обліковий номер: 0823U100148

Особливі позначки: відкрита

Дата реєстрації: 10-03-2023

Статус: Захищена

Реквізити наказу МОН / наказу закладу:

II. Відомості про здобувача

Власне Прізвище Ім'я По-батькові:

1. Белас Андрій Олегович
2. Belas Andrii O.

Кваліфікація:

Ідентифікатор ORCID ID: Не застосовується

Вид дисертації: доктор філософії

Аспірантура/Докторантура: так

Шифр наукової спеціальності: 122

Назва наукової спеціальності: Комп'ютерні науки

Галузь / галузі знань:

Освітньо-наукова програма зі спеціальності: Не застосовується

Дата захисту: 09-03-2023

Спеціальність за освітою: Системний аналіз

Місце роботи здобувача:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

III. Відомості про організацію, де відбувся захист

Шифр спеціалізованої вченої ради (разової спеціалізованої вченої ради): ДФ 26.002.11

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Перемоги, буд. 37, м. Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

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

Повне найменування юридичної особи: Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Перемоги, буд. 37, м. Київ, 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

V. Відомості про дисертацію

Мова дисертації:

Коди тематичних рубрик: 28.29.27, 81.14.10.07

Тема дисертації:

1. Моделі і методи інтелектуального аналізу даних для прогнозування нелінійних нестаціонарних процесів
2. Models and methods of intellectual data analysis for forecasting nonlinear nonstationary processes

Реферат:

1. У дисертаційному дослідженні розв’язана задача підвищення адекватності математичних моделей нелінійних нестаціонарних фінансово-економічних процесів та точності відповідних прогнозів, шляхом застосування сучасних методів інтелектуального аналізу даних до статистичних даних, представлених у вигляді часових рядів. В роботі вибрано, розглянуто і описано класи нелінійних нестаціонарних процесів, що використані для моделювання та прогнозування, а також виконано опис математичних моделей та підходів, що на базі статистичних даних у вигляді часових рядів, використовуються для опису їх динаміки. Вибрано типи процесів для дослідження, такі як: інтегровані, гетероскедастичні процеси, процеси Леві, процеси зі стохастичним трендом, процеси логістичного типу. В роботі вибрано клас фінансово-економічних процесів, сформульовано задачу їх прогнозування. Однак розроблену методику можливо застосовувати і в інших системах (технічних, медичних, тощо) з відповідною визначеною динамікою. Для прогнозування розглянуто

підхід предиктивної аналітики з використанням методів інтелектуального аналізу даних. Методологічною основою роботи є сучасна аналітична методологія SEMMA. Зібрано статистичні дані для експериментів на базі продажів магазину Walmart, продажів протидіабетичних препаратів в Австралії, продажу пального в США для застосування розробленої методики на реальних статистичних даних. Для первинного аналізу даних запропоновано підходи статистичного та графічного аналізу, а також статистичні тести для визначення нелінійності та нестаціонарності процесу: вибрано тест Уайта для перевірки на нелінійність та тест KPSS для перевірки на нестаціонарність як основні. Розглянуто методи для виявлення та обробки аномальних та пропущених значень. Для виділення шумової складової з часового ряду розглянуто фільтр Калмана та метод експонеційного згладжування. Підходи цифрової фільтрації варто застосовувати обережно, не завжди як необхідний етап процесу побудови моделі, а як можливий варіант, обов'язково перевіряючи в кінці процесу моделювання якість отриманих прогнозів з та без використання попередньої фільтрації. В роботі запропоновано метод побудови моделей нелінійних процесів, що відрізняється застосуванням окремих процедур для оптимізації структури лінійної та нелінійної складових моделі з їх наступним адитивним об'єднанням в єдину модель, що забезпечує підвищення адекватності моделі і точності прогнозів у цілому. Розглянуто різні методи комбінування оцінок прогнозів різних моделей, вибрано метод бустінгу як основний для використання в роботі. Описано підхід для виділення та моделювання лінійної складової процесу на базі регресійних моделей, а також підхід з використання отриманої моделі AR для формування оцінок прогнозів, розглянуто і вибрано критерії для оцінки адекватності отриманих моделей, такі як BIC. Для оцінок якості прогнозів побудовано критеріальну базу на базі критеріїв MSE, MAE, RMSLE. Для опису нелінійної складової розглянуто авторегресивний підхід на базі ARIMA, з алгоритмом автоматичної побудови моделі, а також підходи на базі рекурентних (RNN) і згорткових (CNN) нейронних мереж, проаналізовано їх переваги та недоліки. Для нейронних мереж проаналізовано підходи до оптимізації параметрів моделі, запропоновано алгоритм Adam, як найбільш ефективний. Розглянуто різні підходи для багатокрокового прогнозування за допомогою нейронних мереж, вибрано підхід із застосуванням multi-output мереж як основний. Розглянуто необхідність адаптивної побудови моделей для прогнозування нелінійних нестаціонарних процесів, основні принципи адаптації таких моделей. Розглянуто відповідні підходи до адаптації як для лінійних моделей, так і для нейронних мереж. Для адаптації лінійних моделей удосконалено метод оцінювання параметрів на основі методу Монте-Карло для марківських ланцюгів. Проведено порівняльний аналіз отриманих прогнозів з результатами використання відомих підходів та методів. За результатами всіх практичних експериментів було показано, що використання розробленого підходу AR-CNN дозволяє отримувати адекватні моделі та точні прогнози при відносній простоті побудові та невеликих обчислювальних втратах.

2. In the dissertation study, the problem of increasing the adequacy of mathematical models of nonlinear nonstationary financial and economic processes and the accuracy of corresponding forecasts is solved by applying modern methods of intellectual data analysis to statistical data presented in the form of time series. In the work, the classes of nonlinear nonstationary processes used for modeling and forecasting was selected, considered, and described, as well as a description of mathematical models and approaches that are used to describe their dynamics based on statistical data in the form of time series. Types of processes are selected for research, such as integrated, heteroskedastic processes, Levy processes, processes with a stochastic trend, and logistic-type processes. In the work, a class of financial and economic processes was selected, and the problem of their forecasting was formulated. However, the developed methodology can be applied in other systems (technical, medical, etc.) with the corresponding defined dynamics. For forecasting, the approach of predictive analytics using methods of machine learning is considered. The methodological basis of the work is the modern analytical methodology of SEMMA. Collected statistical data for experiments based on sales of the Walmart store, sales of antidiabetic drugs in Australia, and sales of fuel in the USA to apply the developed methodology to real statistical data. For exploratory data analysis, statistical and graphical analysis approaches, as well as statistical tests for determining process nonlinearity and nonstationarity were proposed: White's test for nonlinearity and KPSS test for nonstationarity were chosen as the main ones. Methods for the detection and processing of anomalies and missing values were considered. To extract the noise component from the time series, the Kalman filter and the

exponential smoothing method were considered. Digital filtering approaches should be used carefully, not always as a necessary stage of the model building process, but as a possible option, necessarily checking at the end of the modeling process the quality of the obtained forecasts with and without the use of preliminary filtering. The thesis proposes a method of building models of nonlinear processes, which is distinguished by the use of separate procedures for optimizing the structure of linear and nonlinear components of the model with their subsequent additive combination into a single model, which ensures an increase in the adequacy of the model and the accuracy of forecasts in general. Different methods of combining estimates of forecasts of different models were considered, and the boosting method was chosen as the main one for use in the work. An approach for the selection and modeling of the linear component of the process based on regression models was described, as well as an approach for using the obtained AR model to form forecast estimates, criterion for assessing the adequacy of the obtained models, such as BIC, was considered and selected. A criterion base based on the MSE, MAE, and RMSLE was built to assess the quality of forecasts. To describe the nonlinear component, an autoregressive approach based on ARIMA, with an algorithm for automatic model building, as well as approaches based on recurrent (RNN) and convolutional (CNN) neural networks were considered, and their advantages and disadvantages were analyzed. For neural networks, approaches to optimization of model parameters were analyzed, and the Adam algorithm was proposed as the most effective. Different approaches for multi-step forecasting using neural networks were considered, and the approach using multi-output networks was chosen as the main one. The need for adaptive construction of models for forecasting nonlinear nonstationary processes, the main principles of adaptation of such models were considered. Appropriate approaches to adaptation for both linear models and neural networks were considered. To adapt linear models, the parameter estimation method based on the Monte Carlo method for Markov chains has been improved. A comparative analysis of the obtained forecasts with the results of the use of known approaches and methods was carried out. According to the results of all practical experiments, it was shown that the use of the developed AR-CNN approach allows for obtaining adequate models and accurate forecasts with relative ease of construction and small computational losses.

Державний реєстраційний номер ДіР:

Пріоритетний напрям розвитку науки і техніки:

Стратегічний пріоритетний напрям інноваційної діяльності:

Підсумки дослідження:

Публікації:

Наукова (науково-технічна) продукція:

Соціально-економічна спрямованість:

Охоронні документи на ОПІВ:

Впровадження результатів дисертації:

Зв'язок з науковими темами:

VI. Відомості про наукового керівника/керівників (консультанта)

Власне Прізвище Ім'я По-батькові:

1. Бідюк Петро Іванович

2. Bidyuk Petro Ivanovych

Кваліфікація: д.т.н., 05.13.03

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

VII. Відомості про офіційних опонентів та рецензентів

Офіційні опоненти

Власне Прізвище Ім'я По-батькові:

1. Цюцюра Світлана Володимирівна
2. Tsyutsyura Svitlana Volodymyrivna

Кваліфікація: д.т.н., 05.13.22

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Погорілий Сергій Дем'янович
2. Pogorelyy Sergiy Demianovich

Кваліфікація: д.т.н., 05.13.13

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Рецензенти

Власне Прізвище Ім'я По-батькові:

1. Зайченко Олена Юріївна
2. Zaychenko Olena Yuryivna

Кваліфікація: д.т.н., 05.13.06

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

Власне Прізвище Ім'я По-батькові:

1. Мухін Вадим Євгенійович
2. Mukhin Vadym

Кваліфікація: д. т. н., 05.13.05

Ідентифікатор ORCID ID: Не застосовується

Додаткова інформація:

Повне найменування юридичної особи:

Код за ЄДРПОУ:

Місцезнаходження:

Форма власності:

Сфера управління:

Ідентифікатор ROR: Не застосовується

VIII. Заключні відомості

Власне Прізвище Ім'я По-батькові голови ради Панкратова Наталія Дмитрівна

Власне Прізвище Ім'я По-батькові головуючого на засіданні Панкратова Наталія Дмитрівна

Відповідальний за підготовку облікових документів

Реєстратор

Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності

Юрченко Т.А.