

**НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
ІНСТИТУТ ТЕЛЕКОМУНІКАЦІЙ І ГЛОБАЛЬНОГО ІНФОРМАЦІЙНОГО  
ПРОСТОРУ**

**ТЕРЕНТЬЄВ ОЛЕКСАНДР МИКОЛАЙОВИЧ**



УДК 004.89:519.22](043.3)

**МОДЕЛІ, МЕТОДИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ПРОГНОЗУВАННЯ  
НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ В УМОВАХ  
НЕВИЗНАЧЕНОСТІ**

05.13.06 – Інформаційні технології

**АВТОРЕФЕРАТ**  
дисертації на здобуття наукового ступеня  
доктора технічних наук

Київ – 2021

Дисертацією є рукопис.

Робота виконана в Інституті телекомунікацій і глобального інформаційного простору НАН України

**Науковий консультант:** член-кореспондент НАН України,  
доктор технічних наук, професор  
**ТРОФИМЧУК Олександр Миколайович,**  
Інститут телекомунікацій і глобального  
інформаційного простору НАН України, директор

**Офіційні опоненти:** доктор технічних наук, професор  
**НАКОНЕЧНИЙ Олександр Григорович,**  
Київський національний університет імені Тараса  
Шевченка, завідувач кафедри системного аналізу та  
теорії прийняття рішень

доктор технічних наук, професор  
**КОРАБЛЬОВ Микола Михайлович,**  
Харківський національний університет  
радіоелектроніки, професор кафедри комп'ютерних  
інтелектуальних технологій і систем

доктор технічних наук, професор  
**МАЛАХОВ Євгеній Валерійович,**  
Одеський національний університет імені  
І. І. Мечникова, завідувач кафедри математичного  
забезпечення комп'ютерних систем

Захист дисертації відбудеться 27 квітня 2021 року о 13<sup>00</sup> на засіданні спеціалізованої вченої ради Д 26.255.01 Інституту телекомунікацій і глобального інформаційного простору НАН України за адресою: 03186, м. Київ, Чоколівський бульвар, 13, к. 601.

З дисертацією можна ознайомитись у бібліотеці Інституту телекомунікацій і глобального інформаційного простору НАН України за адресою: 03186, м. Київ, Чоколівський бульвар, 13.

Автореферат розісланий 26 березня 2021 року.

Вчений секретар  
спеціалізованої вченої ради  
к.т.н., ст. дослідник



О. Г. Лебідь

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Тенденції, характерні для сучасного етапу розвитку суспільства, концентрують увагу на необхідності вирішення фундаментальних проблем, пов'язаних із дослідженням та прогнозуванням нелінійних нестационарних процесів різної природи, розвиток яких триває в умовах невизначеності та ризику. Тобто зростає потреба у постійному удосконаленні методів прогнозування за рахунок формування принципово нових підходів до формалізації об'єкта прогнозування, побудови моделей (ансамблів моделей) та опрацювання результатів прогнозування, які слугуватимуть підґрунтям для прийняття коректних управлінських рішень.

Побудова моделей та удосконалення методів потребують розробки сучасних інформаційних технологій, які дозволяють автоматизувати вирішення комплексу задач, пов'язаних з аналізом значних обсягів інформації. Остання отримується з різних джерел, часто є неповною чи спотвореною. Існує необхідність виявлення в ній закономірностей, відображення властивостей досліджуваних явищ та процесів, зокрема, циклічності розвитку, сезонності, подібності процесів тощо.

Розвинута у роботі теорія вирішення проблеми прогнозування розвитку нелінійних нестационарних процесів в умовах наявності невизначеностей різного типу ґрунтується на дослідженнях та теоретичних результатах вітчизняних та закордонних вчених: О. М. Трофимчука, С. О. Довгого, О. Г. Наконечного, В. Є. Снитюка, М. З. Швиденка (*технології та системи підтримки прийняття рішень в умовах невизначеності*); В. Я. Данилова, В. М. Азарскова (*теорія та прикладні методи системного аналізу*); В. С. Степашка, С. Г. Удовенка, П. І. Бідюка (*аналіз нелінійних систем та методи прогнозування*); М. М. Корабльова, Є. В. Бодянського., Ю. В. Крака, В. І. Литвиненка, С. А. Положаєнка (*методи інтелектуального аналізу даних*); О. І. Міхальова, Ю. П. Кондратенка, Ю. П. Зайченка, В. В. Кульби (*методи нечіткого аналізу і моделювання*); Т. Байєса, Дж. Перла, А. Л. Тулуп'єва (*основи ймовірнісного підходу*); Є. В. Малахова, С. Ф. Теленика, О. П. Гожого, О. С. Меняйленка, С. В. Цюцюри, Р. О. Коржа (*інформаційні технології*), С. Клайна, Дж. Краскала, А. А. Гухмана, М. Леонарда, Дж. Слоана, Т. Ли, Б. Ельшеймера, С. Шуберта (*теорія подібності процесів*), О. А. Павлова (*методи моделювання складних систем*), Ф. Бєрнштейна, С. Холсаппла, В. Г. Тоценка, О. Ф. Волошина, (*побудова систем підтримки прийняття рішень*), Д. В. Ланде, М. Бері, С. Аггарвала, Х. До Прадо, Е. Фемеди (*Text Mining*) та інших.

Незважаючи на наявність значної кількості досліджень з даної тематики, проблема розробки моделей, методів та інформаційних технологій прогнозування нелінійних нестационарних процесів залишається невирішеною повністю. Зокрема, слід відзначити відсутність загальних підходів до аналізу та попередньої обробки структурованих і неструктурованих даних, які описують нелінійні нестационарні процеси, в тому числі із застосуванням методів інтелектуального аналізу даних. Також нерозв'язаною залишається й низка задач, пов'язаних із виявленням характеру досліджуваних процесів, розробкою універсальних методик побудови моделей та їх ансамблів для отримання прогнозів нелінійних нестационарних процесів.

Вирішення вказаних задач неможливе без створення єдиної науково

обґрунтованої інформаційної технології, яка б охоплювала весь комплекс задач побудови прогнозів високої якості, необхідних для прийняття ефективних управлінських рішень.

На підставі викладеного можна стверджувати, що проблема побудови інформаційних технологій аналізу та прогнозування нелінійних нестационарних процесів, які забезпечують високу обчислювальну ефективність, є актуальною.

Важливою і актуальною науково-прикладною проблемою, що розв'язується в дисертаційній роботі є підвищення ефективності підтримки прийняття рішень управління нелінійними нестационарними процесами в умовах невизначеностей та ризиків різної природи, на основі застосування сучасних методів системного аналізу, моделювання, прогнозування та інформаційних технологій.

**Зв'язок роботи з науковими програмами, планами й темами.** Тема дисертаційної роботи повністю відповідає науковим напрямам, за якими виконуються передові наукові дослідження в Інституті телекомунікацій і глобального інформаційного простору, зокрема, у сфері систем і методів прийняття рішень, прогнозування, побудови інтелектуальних інформаційних систем. Робота виконувалась в Інституті телекомунікацій і глобального інформаційного простору в рамках цільового проекту наукових досліджень НАН України 0120U000000 «Розробка інформаційної технології моделювання і прогнозування розвитку соціально-еколого-економічних систем в умовах невизначеності, нестационарності та ризику» (2020-2021 рр.) та наукової (науково-технічної) роботи 0121U109211 «Розробка інформаційно-аналітичної системи для дослідження спроможності та прогнозування розвитку територіальних громад із застосуванням методів інтелектуального аналізу даних» (2021-2023 рр.), а також в Інституті прикладного системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» в рамках науково-дослідницьких тем 0109U000300 «Побудова системи підтримки прийняття рішень на основі теорії байєсівських мереж для моделювання поведінки складних систем» (2009-2010 рр.), 0111U001241 «Розробка і реалізація методики інтелектуального аналізу даних із використанням теорії мереж Байєса та регресійного аналізу» (2011-2012 рр.), 0113U000650 «Розробка інформаційної технології моделювання та оцінювання фінансово-економічних ризиків із врахуванням невизначеностей різної природи (на основі байєсівських моделей)» (2013-2014 рр.).

**Мета й завдання дисертаційного дослідження.** Метою дослідження є створення нових моделей та методів виконання системних досліджень нелінійних нестационарних процесів різної природи, розроблення на їх основі методологічних засад розробки інформаційних технологій, що дозволить підвищити якість оцінок прогнозів та відповідних рішень в умовах невизначеності та ризику.

Для досягнення поставленої мети в роботі необхідно вирішити такі *завдання*:

- виконати огляд та аналіз сучасних методів моделювання і прогнозування розвитку нелінійних нестационарних процесів різної природи в умовах невизначеності;
- розробити системну методологію побудови математичних моделей для опису та прогнозування процесів, що характеризують розвиток нелінійних нестационарних процесів в умовах невизначеності;
- удосконалити методи прогнозування нелінійних нестационарних процесів різної

природи на основі використання ймовірнісних та комбінованих прогнозів;

- запропонувати метод оцінювання параметрів математичних моделей та їх ансамблів, який дозволяє подолати зміщеність оцінок прогнозів;
- розробити метод розкриття невизначеностей різних типів з метою коректного розв'язання задач прогнозування розвитку обраних процесів;
- побудувати математичні моделі та ансамблі моделей обраних нелінійних нестационарних процесів для прогнозування їх розвитку з метою підтримки прийняття управлінських рішень в умовах інформаційної невизначеності;
- розробити та виконати апробацію запропонованих підходів до моделювання та прогнозування, запропонувати підходи, що забезпечують адекватний опис причинно-наслідкових зв'язків різних груп чинників та визначення можливих варіантів розвитку досліджуваних нелінійних нестационарних процесів;
- виконати аналіз адекватності побудованих моделей та формування висновків на їх основі щодо можливості їх застосування для розв'язання задач оцінювання і прогнозування розвитку нелінійних нестационарних процесів;
- на основі системного підходу побудувати та реалізувати інформаційну технологію для розв'язання задач моделювання та прогнозування нелінійних нестационарних процесів у різних галузях досліджень для її подальшого використання у системах підтримки прийняття рішень.

*Об'єкт дослідження* – нелінійні нестационарні процеси різної природи.

*Предмет дослідження* – моделі, методи та інформаційні технології дослідження та прогнозування нелінійних нестационарних процесів і систем, прийняття рішень щодо їх розвитку в умовах наявності невизначеностей і ризиків.

**Методи дослідження.** *Теоретичну та методологічну основу дослідження* складають метод діалектичного пізнання стану та особливостей розвитку нелінійних нестационарних процесів, системний підхід до визначення факторів і механізмів їх розвитку. *При узагальненні теоретичних та методологічних засад* математичного моделювання та прогнозування розвитку нелінійних нестационарних процесів використані методи індукції та дедукції, аналізу і синтезу, аналогій і зіставлення, формалізації й моделювання, методи порівняння, дослідження часових рядів, кластерного та інтелектуального аналізу даних, сценарний підхід, елементи теорії прийняття рішень. *Для аналізу і обробки інформації* – методи системного, інтелектуального та статистичного аналізу даних, експертного оцінювання; *для моделювання нелінійних нестационарних процесів* – методи ймовірнісно-статистичного та економіко-математичного моделювання; *для побудови, аналізу і оцінювання прогнозів* – методи прогнозування на основі теорії подібності процесів, регресійного аналізу, ймовірнісно-статистичних моделей, дерев рішень та багатовимірних розподілів; *для розроблення інформаційної технології* використано методи системного аналізу, підтримки прийняття рішень, багатокритеріального аналізу та байєсівського підходу; *для побудови практичних реалізацій* – методи та засоби проектування і реалізації систем підтримки прийняття рішень, прикладне програмування.

## **Наукова новизна отриманих результатів:**

### ***Вперше:***

– розроблено новий метод опрацювання невизначеностей, який ґрунтується на застосуванні теорії подібності процесів, який відрізняється робастністю результатів аналізу, що забезпечує отримання оцінок прогнозів високої якості за наявності неповних або спотворених даних;

– запропоновано новий метод побудови регресійних та ймовірнісно-статистичних моделей у формі мереж Байєса, який відрізняється можливістю врахування нестаціонарності і нелінійності стосовно змінних, що забезпечує високу адекватність моделей і якість прогнозів процесів досліджуваного типу;

– розроблено метод моделювання нелінійних нестаціонарних процесів різної природи в умовах невизначеності, який відрізняється від відомих урахуванням різних типів невизначеностей, що підвищує адекватність моделей і якість оцінок прогнозів за лінійними та нелінійними моделями;

– побудовано та досліджено ансамблі моделей для формального опису нелінійних нестаціонарних процесів, які відрізняються модифікованою комплексною структурою та високою адекватністю, що дозволяє підвищити якість оцінювання прогнозів розвитку досліджуваних процесів;

– запропоновано метод прогнозування, оснований на використанні адаптивного підходу до моделювання у поєднанні із статистичним та ймовірнісним моделюванням, що дає можливість урахувати структурно-параметричні невизначеності і забезпечує адекватний опис причинно-наслідкових зв'язків і можливих варіантів розвитку процесів різної природи під впливом груп внутрішніх та зовнішніх чинників;

– запропоновано нові моделі і методи створення інформаційних технологій розв'язування задач побудови математичних моделей для прогнозування нелінійних нестаціонарних процесів, які ґрунтуються на принципах багатомодельного та багатокритеріального підходів, інтеграції різнотипної інформації і засновані на системному використанні методів інтелектуального аналізу даних, ймовірнісно-статистичного моделювання, теорії подібності процесів, прогнозування і підтримки прийняття рішень, що підвищує обґрунтованість прийняття рішень в умовах наявності невизначеностей та ризиків різних типів;

– розроблено інформаційну технологію, в основу якої покладено поєднання принципів системного аналізу, методів обробки та оцінювання якості даних, прогнозного моделювання із використанням нових моделей та їх композицій, запропонованих критеріїв адекватності моделей, оцінок якості прогнозів, яка забезпечує високу якість проміжних та остаточних результатів дослідження нелінійних нестаціонарних процесів.

### ***отримали подальший розвиток:***

– метод аналізу інформації на основі засобів статистично-ймовірнісного моделювання, який дає змогу зменшити суб'єктивізм відбору найбільш значимих чинників в задачах прогнозування процесів різної природи, що створює передумови для прийняття обґрунтованих управлінських рішень;

– метод побудови моделей у формі байєсівської мережі для оцінювання розвитку

динаміки процесів різної природи, який відрізняється коректністю формального опису за байєсівським інформаційним критерієм, що забезпечує обчислення високоякісних ймовірнісних оцінок прогнозів розвитку досліджуваних процесів;

– системна методологія побудови адаптивних моделей процесів різної природи, яка відрізняється удосконаленням існуючих методів моделювання в умовах невизначеності і нестаціонарності даних, методом структурно-параметричної адаптації, яка забезпечує підвищення адекватності моделей та якості оцінок прогнозів;

**удосконалено:**

– інформаційну технологію розв’язання задач прогнозування розвитку досліджуваних процесів нелінійних нестаціонарних процесів, яка створює підґрунтя для прийняття ефективних рішень;

– метод оцінювання параметрів математичних моделей, який відрізняється комплексним застосуванням теорії оцінювання та байєсівського підходу, що забезпечує подолання проблеми зміщеності оцінок;

– інформаційну технологію, призначену для реалізації у системах підтримки прийняття рішень на основі системного підходу, множини методів ідентифікації і врахування невизначеностей, регресійного та інтелектуального аналізу даних, яка забезпечує побудову адекватних моделей досліджуваних процесів і обчислення високоякісних оцінок прогнозів.

**Практичне значення отриманих результатів** полягає у тому, що результати теоретико-методологічного та емпіричного досліджень доведені до використання в науково-практичній діяльності, а також апробовані та впроваджені за безпосередньою участю автора в практичну діяльність підприємств та установ, що підтверджується відповідними довідками: Державної служби України з лікарських засобів та контролю за наркотиками; ТОВ «Картезіан-Європа» та Smart Arbitrage Technologies Limited. Результати використовуються у навчальному процесі Інституту прикладного системного аналізу Національного технічного університету «Київський політехнічний інститут імені Ігоря Сікорського»

**Особистий внесок здобувача.** Усі теоретичні та практичні результати, що виносяться на захист, отримані автором самостійно. Пошук та аналіз літературних джерел за тематикою дисертаційного дослідження виконано автором особисто. Наукові положення, висновки, рекомендації, що викладені в дисертації, розроблені особисто здобувачем. У працях, виконаних у співавторстві, автору дисертації належать: розроблено системну методологію моделювання і прогнозування нелінійних нестаціонарних процесів в енергетиці [6, 7]; запропоновано критеріальну базу для аналізу адекватності лінійних і нелінійних моделей досліджуваних процесів [1-4, 6,7, 11, 12, 15, 16, 18, 19, 20]; розроблено метод моделювання і прогнозування нелінійних нестаціонарних процесів з використанням мереж Байєса [2-5, 8, 9, 12-14, 17,]; запропоновано метод ідентифікації невизначеностей ймовірнісно-статистичного типу [1,2,4-14,17,19,20]; розроблено архітектуру системи підтримки прийняття рішень для аналізу та прогнозування нелінійних нестаціонарних процесів [6, 9, 21]; створено функціональну схему системи підтримки прийняття рішень із використанням запропонованої інформаційної технології [2, 6, 7, 12, 14, 16, 19, 20].

**Апробація результатів дисертації.** Основні положення дисертації було подано

і обговорено більше, ніж на 17 міжнародних, всеукраїнських та науково-технічних конференціях: 12-й міжнародній науково-технічній конференції «Системний аналіз та інформаційні технології» (м. Київ, 24 квітня 2012 р.); 15-й міжнародній науково-технічній конференції «Системный анализ и информационные технологии» (м. Київ, 27–31 травня 2013 р.); 15-й міжнародній конференції «International conference on control and optimization with industrial applications» (м. Баку, 27-29 серпня 2015 р.); науково-технічній конференції «Геоинформационные системы и компьютерные технологии эколого-экономического мониторинга – 2016» (м. Дніпропетровськ, 13–15 квітня 2016 р.); міжнародній конференції «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту» (м. Залізний порт, 24-28 травня 2016 р.); 18-й міжнародній науково-технічній конференції “Системний аналіз та інформаційні технології” (м. Київ, 30 травня – 2 червня 2016 р.); міжнародній конференції «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту» (м. Залізний порт, 22-26 травня 2017 р.); 11-й всеукраїнській науково-практичній конференції «Моделювання та прогнозування економічних процесів» (м. Київ, 6-8 грудня 2017 р.); IEEE First Ukraine Conference on Electrical and Computer Engineering (м. Київ, 29 травня – 2 червня 2017 р.); 16-й міжнародній науково-практичній конференції «Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях» (м. Київ, Пуща-Водиця, 3-4 жовтня 2017 р.); 6-й міжнародній конференції «Control and Optimization with Industrial Applications» (м. Баку, 11-13 липня 2018 р.); 6-й міжнародній конференції «Advanced Information Systems and Technologies, AIST 2018» (Суми, 16-18 травня 2018 р.); 1st International Conference on Computer Science, Engineering and Education Applications (м. Київ, 18 -20 січня 2018 р.); 14-й міжнародній науково-практичній конференції «Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: актуальні питання» (м. Київ, 06-07 жовтня 2020р.); 4-й міжнародній науково-практичній конференції «Прикладні системи та технології в інформаційному суспільстві» (м. Київ, 30 вересня 2020 р.); 14-й міжнародній конференції «Application of information and communication technologies» (м. Ташкент, 07-09 жовтня 2020 р.); 9-й міжнародній конференції «Information Control Systems & Technologies» (м. Одеса, 24–26 вересня 2020 р.).

**Публікації.** Основні наукові результати дисертації опубліковані у 42 наукових працях, у тому числі, 2 колективних монографіях, 18 статтях у фахових виданнях, в тому числі 5 статей у закордонних фахових виданнях, з них 4 – у виданнях, що індексуються у наукометричній базі Scopus, 15 – у збірниках наукових праць, матеріалах і тезах міжнародних і національних конференцій. Також опубліковано 7 робіт, що додатково відображають наукові результати дисертації: 3 навчальні посібники, 4 авторських свідоцтва на комп’ютерну програму.

**Структура та обсяг роботи.** Робота складається із вступу, 7 розділів, списку використаних джерел та додатків. Загальний обсяг роботи складає 293 сторінки, з яких основного тексту 285 сторінок, 77 ілюстрацій, 58 таблиць, 12 додатків. Список використаних джерел налічує 274 найменування на 32 сторінках.



## ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано актуальність теми роботи, розкрито суть і стан проблеми, сформульовано мету й завдання дослідження, об'єкт, предмет та методи дослідження, достовірність отриманих результатів, зв'язок із науковими програмами, планами, темами, сформульовано наукову новизну роботи та практичне значення одержаних результатів, подано кількість публікацій за темою роботи, виділено особистий внесок здобувача.

У **першому розділі** виконано системологічний аналіз проблеми розробки аналітичного інструментарію для створення інформаційних технологій прийняття рішень на основі прогнозування нелінійних нестационарних процесів різної природи.

Враховуючи те, що розвиток більшості нелінійних нестационарних процесів відбувається в умовах невизначеностей, проаналізовано основні типи невизначеностей та методи їх подолання, наведено класифікацію основних типів невизначеностей, характерних для аналізу нелінійних нестационарних процесів.

Виконано аналіз стану робіт у галузі інформаційних технологій підтримки прийняття рішень і з'ясовано, що відсутня загальна теорія розв'язання задач прийняття рішень щодо розвитку нелінійних нестационарних процесів різної природи з урахуванням невизначеностей різних типів.

Як показав проведений аналіз, наявні методи й технології розв'язання задач прогнозування розвитку нелінійних нестационарних процесів мають певні переваги і недоліки, що створює передумови для їх удосконалення, створення нових методів, алгоритмів, інформаційних технологій, які дозволять автоматизувати більшість вказаних процесів.

Окрім того, відсутній метод побудови інформаційних технологій, який би дозволив реалізувати сучасні підходи до прогнозного моделювання, що ґрунтуються на використанні методів інтелектуального аналізу даних, інтеграції й комбінованого використання методів аналізу даних та побудови моделей нелінійних нестационарних процесів, методів прогнозування і прийняття рішень для вирішення завдань управління розвитком нелінійних нестационарних процесів. Все це підкреслює актуальність досліджуваної проблеми та необхідність створення сучасних інформаційних технологій.

Враховано те, що нелінійні нестационарні процеси, що мають місце у різних сферах діяльності, мають особливості розвитку та протікають під впливом багатьох специфічних факторів. Тому надано характеристику та подано класифікацію нелінійних нестационарних процесів; відзначено, що кожний клас потребує відповідного формального опису використовуючи математичні моделі різних типів та інтелектуальний аналіз даних. Значну увагу приділено дослідженню процесів, що містять нестационарності, інтегрованим процесам, гетероскедастичним та коінтегрованим процесам.

Розглянуто процеси різної природи та функціонуючі в різних умовах, наведено їх класифікацію, відзначено особливості характеристик їх окремих типів, проаналізовано існуючі методи та моделі, що використовуються для їх прогнозування, подано загальну схему оцінювання прогнозів нелінійних нестационарних процесів (рис. 1).

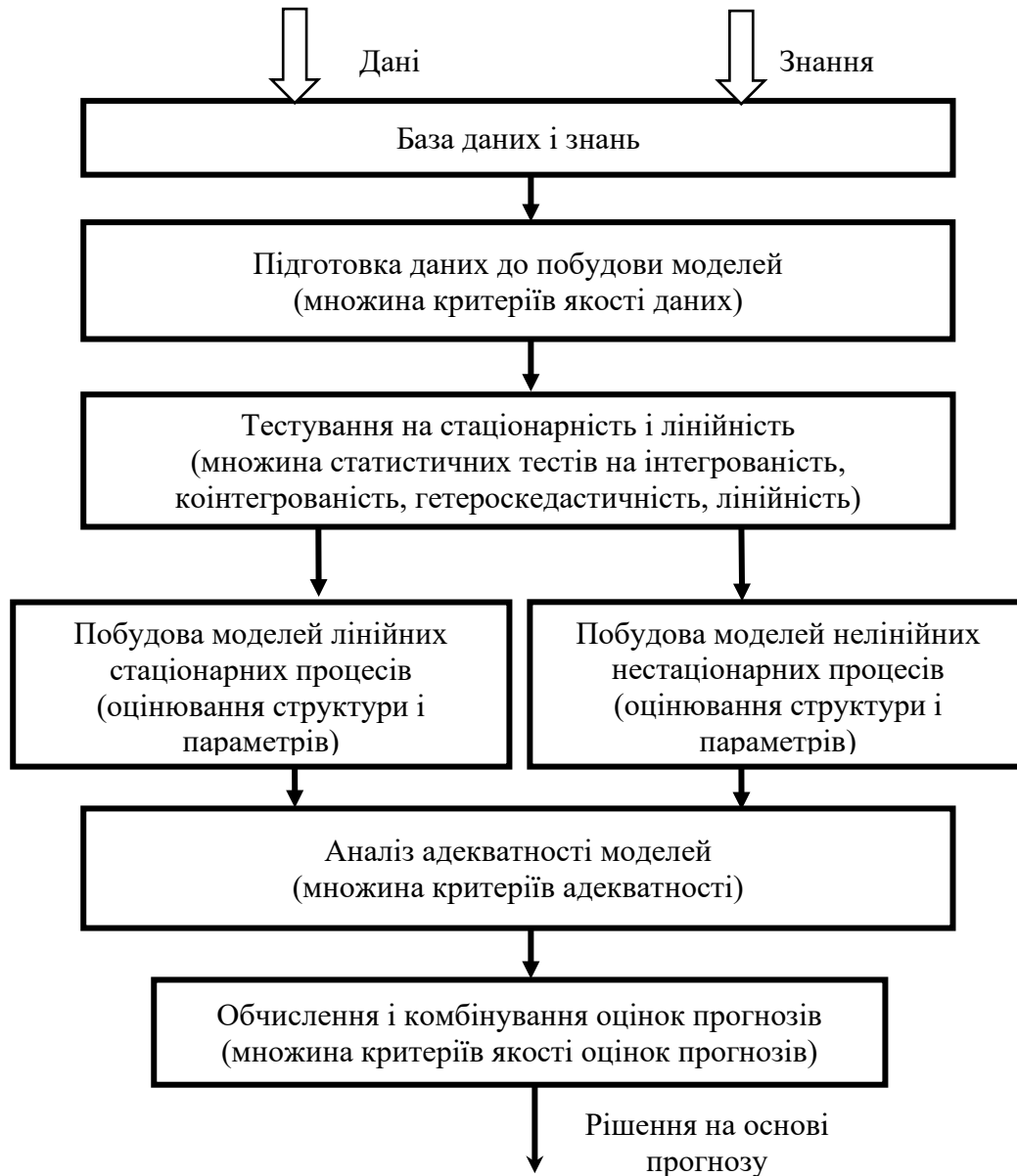


Рисунок 1 – Загальна схема аналізу нелінійних нестаціонарних процесів та оцінювання прогнозів

Відзначено, що якість результатів прогнозування значною мірою залежить від повноти та достовірності вхідної інформації, яка використовується для формування часових рядів.

Заповнення пропусків даних, незважаючи на наявність значної кількості досліджень, залишається невирішеною проблемою. На основі аналізу існуючих методів та моделей узагальнено класифікацію пропусків; для визначення найбільш поширених причин виникнення пропусків в часових рядах даних було проведено чисельні експерименти та запропоновано власну методика, яка дозволяє опрацьовувати часові ряди, пропуски у яких мають різні механізми формування та різну кількість пропущених значень (рис. 2).

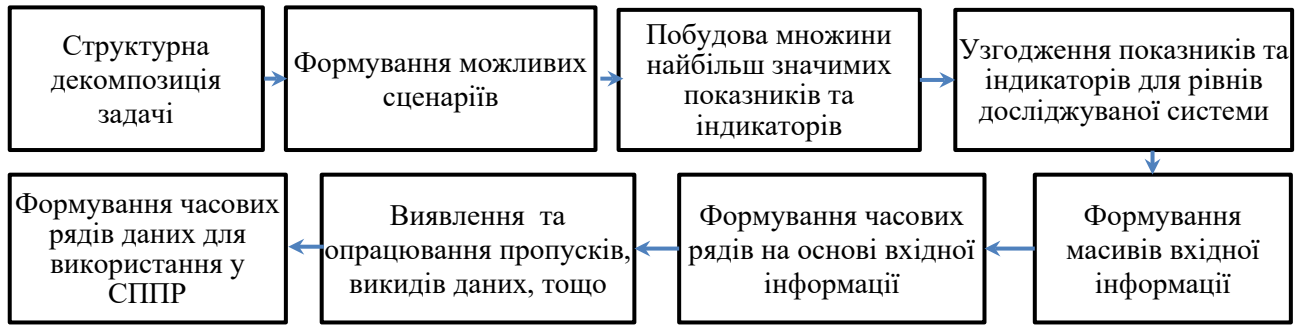


Рисунок 2 – Методика обробки вхідної інформації на етапі попереднього аналізу даних та заповнення пропусків

Як показали виконані чисельні експерименти, при виборі методу для заповнення пропусків, потрібно визначити чи відповідає обраний метод конкретній аналітичній моделі досліджуваного процесу. Крім того, заповнення пропусків випадковими або детермінованими значеннями на основі використання, наприклад, аналітичної моделі або середнім за вибіркою, є штучним підходом, що не дозволяє виявити наявність шуму в даних та спотворює уяву про характер досліджуваного процесу. Тому для заповнення пропусків у часових рядах запропоновано використовувати методiku на основі аналітичної моделі, побудова та реалізація якої здійснюється у два (за потреби у три) етапи. На першому етапі виконується аналіз наявності пропусків у досліджуваних часових рядах, на другому – аналіз характеру виникнення, кількість пропусків в даних та аналіз на наявність шаблонів, на третьому – виявлення допоміжних змінних за допомогою кореляційного аналізу даних та експертних знань стосовно області дослідження (даний етап виконується у випадку необхідності). Для визначення ефективних методів заповнення пропусків даних, було виконано порівняльний аналіз існуючих методів заповнення пропущених значень у довгих часових рядах даних та експериментально перевірено ефективність застосування різних алгоритмів за різної кількості пропусків. В ході виконання значної кількості експериментів було побудовано прогнози на основі часових рядів, які містили до 20% пропущених значень. Пропуски були заповнені із використанням різних методів за десятьма сценаріями наявності до 20% пропусків даних у досліджуваних часових рядах. Виявлено, що навіть, якщо часовий ряд містить 20% пропущених значень, прогноз, отриманий після їх заповнення значеннями, отриманими за запропонованою методикою, дав кращі результати. Зокрема, відхилення прогнозного значення показника відрізнялось від фактичного в межах 5-10%, а значення MAPE становило від 8,5 до 24%, MSE – від 5,4 до 11% (залежно від використаного методу заповнення пропусків), в той час у випадку, коли заповнення пропусків не виконувалось, значення вказаних показників становили 61-78%.

Для вирішення окреслених проблем були виділені та застосовані основні аспекти використання методології системного аналізу з метою розв'язання задач прогнозування нелінійних нестационарних процесів різної природи. Представлена загальна методика обробки інформації при розв'язанні задач формування інформаційної бази для їх дослідження, в тому числі ідентифікацію та опрацювання невизначеностей різних типів при формуванні часових рядів вхідних даних, визначено

методи та моделі для прогнозування нелінійних нестационарних процесів та шляхи їх реалізації в рамках відповідної інформаційної технології.

У другому розділі представлено розроблену методіку попереднього опрацювання значних обсягів різнорідних даних, отриманих із різних інформаційних джерел. Вказана методіка заснована на системному використанні методів статистичного аналізу та автоматизованого інтегрування різнотипної інформації, методів моделювання, прогнозування та багатокритеріального прийняття рішень. Для подолання можливих невизначеностей статистичного, структурного та параметричного типів запропоновано множину методів, залежно від типу невизначеності. Запропонована методіка попереднього опрацювання даних вирізняється тим, що при формуванні масивів вхідних даних їх інформативність, синхронність та коректність забезпечуються на етапі попередньої обробки, яка виконується для приведення їх до форми, що забезпечить можливість коректного застосування методів оцінювання параметрів моделі та отримання їх статистично значущих оцінок. Зокрема, розроблено методіку корегування значних імпульсних (екстремальних) значень, нормування вимірів у заданих межах, логарифмування великих значень та фільтрації шумових складових, виявлення та усунення мультиколінеарності.

Встановлено, що для побудови математичних моделей у задачах моделювання нелінійних нестационарних процесів доцільно використовувати підхід, що ґрунтується на розділенні всієї вибірки вхідних даних на тренінгову, валідаційну та тестову. Проблемним питанням побудови математичних моделей нелінійних нестационарних процесів є так зване «прокляття розмірності», коли необхідно будувати модель на коротких часових вибірках даних. Тобто використання існуючих методів прогнозування в задачах прийняття рішень, які ґрунтуються на аналітичних процедурах, логічних правилах та раціональному експертному оцінюванні, у багатьох випадках не дає бажаного результату стосовно якості оцінок прогнозів, а тому в процесі дослідження зазвичай вирішується проблема системного використання альтернативних методів моделювання і прогнозування для підвищення якості оцінок прогнозів для кожного конкретного випадку. Зокрема, для моделей нелінійних нестационарних процесів важливо знайти підмножину вхідних змінних (які називаються ознаками або атрибутами), що якнайкраще характеризуватимуть предметну область. Для цього застосовуються три стратегії – стратегія фільтра (наприклад, накопичення ознак), стратегія обгортання (наприклад, пошук згідно заданої точності) і стратегія вкладення (вибираються ознаки для додавання або видалення в міру побудови моделі, заснованої на аналізі похибок прогнозування).

Для застосування техніки відбору ознак при побудові моделей нелінійних нестационарних процесів у роботі використано такі підходи: спрощення моделей, щоб зробити їх простішими для інтерпретації дослідниками / користувачами; зменшення часу на тренування (навчання) моделі; зменшення розмірності вхідних даних; покращене узагальнення шляхом уникнення перенавчання (формально, це зменшення дисперсії).

Пропонована методіка попереднього аналізу даних призначена для аналізу даних на наявність зв'язків та аномалій під час побудови моделей нелінійних нестационарних процесів складається з таких етапів:

Етап 1. Графічний аналіз даних із використанням діаграм розсіювання включає використання графіків та кореляційного аналізу.

Етап 2. Діагностика екстремальних значень та викидів, включає аналіз діаграм розсіювання та графіків на наявність викидів, за потреби, з'ясовуються причини появи викидів.

Етап 3. Аналіз впливу екстремальних значень, студентизовані залишки; D-статистика Кука; статистика DFFITS; статистика DFBETAS.

Етап 4 Дослідження даних на наявність колінеарності, включає використання VIF-статистики, індекс стану та пропорцію варіації. Перевірка пропорції варіації включає дослідження студентизованих залишків, D-статистика Кука та DFFITS статистика.

Етап 5. Аналіз категоріальних даних, включає дослідження їх розподілу, проведення попереднього дослідження наявності зв'язку між змінними (використовується статистика  $\chi^2$ -квадрат Пірсона, V-статистика Крамера, співвідношення шансів,  $\chi^2$ -квадрат тест Мантел-Гаензеля)

Етап 6. Стандартизація даних використовує методи стандартизації, узагальнення яких представлено в роботі.

Згідно з принципами багатомодельного та багатокритеріального підходів розроблено метод застосування інформаційних технологій для розв'язування задач прийняття рішень щодо управління розвитком нелінійних нестационарних процесів, який ґрунтується на інтеграції різнотипної інформації й заснований на системному використанні методів аналізу даних, моделювання, методів прогнозування і методів прийняття рішень.

Для вирішення цієї задачі запропоновано виконувати аналіз неструктурованої інформації, в тому числі розміщеної в мережі Інтернет, використовуючи засоби SAS Textual Analytics. Методика використання засобів текстової аналітики, запропонована у роботі, передбачає послідовну реалізацію множини кроків, які охоплюють як попередню обробку текстової інформації, формування множини правил для аналізу, відбір найбільш значимих критеріїв та цілей, так і візуалізацію одержаних результатів. Використання засобів текстової аналітики, як це було запропоновано у роботі, дає можливість найбільш повно сформулювати альтернативи та визначити цільову настанову. У випадках, коли інформація, що описує досліджувані процеси має суттєві протиріччя, пропуски, похибки і затримки, може бути сумнівною щодо достовірності, доцільно застосовувати методику визначення подібності процесів.

Для реалізації запропонованої методики обробки та аналізу неструктурованої інформації, отриманої з мережі Інтернет, запропоновано алгоритм латентно-семантичного аналізу, реалізований засобами мови програмування SAS Base. Пропонована методика вирізняється тим, що на етапах інтерпретації результатів та виявлення найбільш значимих факторів можуть бути застосовані дерева рішень, кластерний аналіз, ймовірно-статистичні моделі, експертні оцінки. Узагальнена методика використання неструктурованих даних для побудови моделей нелінійних нестационарних процесів представлена на рис. 3.

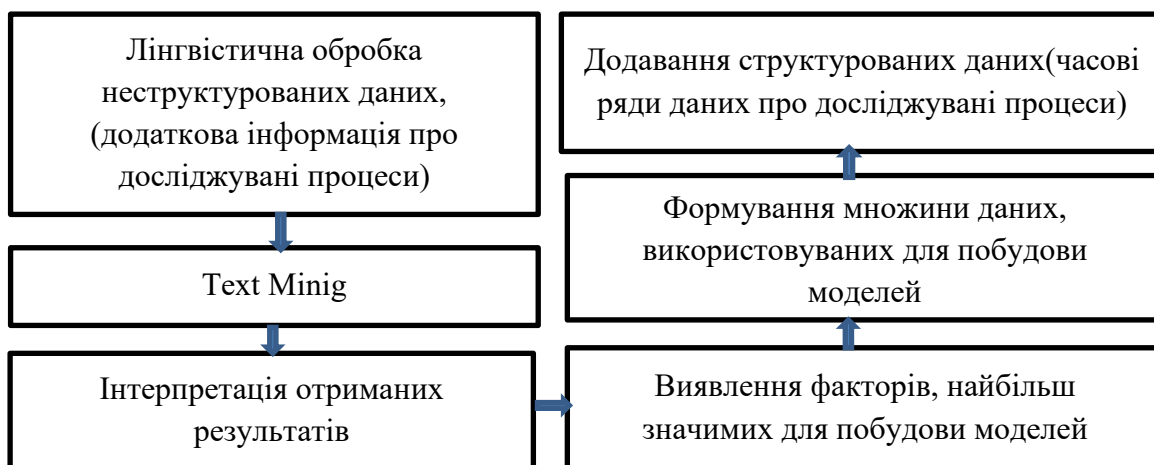


Рисунок 3 – Методика опрацювання неструктурованих даних із використанням Text Mining

Таким чином, представлені методи попереднього аналізу даних формують цілісну методику підготовки статистичних даних для подальшого використання при побудові математичних моделей нелінійних нестационарних процесів та використанні цих моделей для оцінювання прогнозів.

У **третьому розділі** запропоновано використання теорії подібності до аналізу нелінійних нестационарних процесів різної природи. Цей метод надає особливі переваги при вирішенні задач моделювання, що мають неповні або спотворені вхідні дані. Пропонується новий робастний непараметричний метод на основі аналізу нелінійних нестационарних процесів за їх подібністю. Суть методу полягає у пошуку часових рядів (процесів) досліджуваного типу зі схожою статистичною поведінкою, що в результаті допомагає зробити висновки про наявність поведінкових шаблонів досліджуваних процесів. Для врахування різної поведінки рядів у часі за допомогою аналізу подібності розраховується відстань між першою часовою послідовністю та другою із урахуванням впорядкування. Зсув часу (ковзання) послідовно зміщує перший часовий ряд уздовж часової шкали з метою пошуку подібних зразків, які можуть виникати в довільні моменти часу. Перетворення часового ряду дозволяє «розтягувати» або «стискати» часовий вимір для випадків незначної відмінності у часі між схожими шаблонами.

Формально постановку задачі виявлення подібності часових рядів можна сформулювати таким чином. Нехай в наявності є два часових ряди  $Y$  та  $X$ , які математично записуються як послідовності значень  $Y = \{y_1, \dots, y_n\}$  та  $X = \{x_1, \dots, x_m\}$ .  $Y$  – цільовий ряд значень, що містить  $n$  елементів, а  $X$  – вхідний ряд значень, що містить  $m$  елементів. При описі алгоритмів для послідовного перебору всіх значень цільового ряду  $Y$  використовується індекс  $j$  який змінюється від 1 до  $n$ , а для перебору всіх значень вхідного ряду  $X$  використовується індекс  $i$  який змінюється від 1 до  $m$ . В загальному випадку вводиться деякий функціонал  $d_{ij}$ :  $d_{ij} = d(x_i, y_j)$ , що відображає відстань між двома точками  $x_i$  та  $y_j$ .

Для обчислення відстані між часовими рядами для їх порівняння, з метою оцінювання подібності, можуть використовуватися різні формули, зокрема, формула

модуля різниці, формула квадрата різниці, варіанти мір Мінковського тощо.

При проведенні порівняння часових рядів доцільно виконувати нормування даних. Це доцільно робити тому, що може виникати проблема несумісності одиниць вимірювань змінних, що може суттєво вплинути на кінцевий результат. Найбільш популярними підходами до нормування даних є нормування за діапазоном та стандартне. З метою зменшення області допустимих значень як цільового, так і вхідного рядів, використовують різні аналітичні підходи, зокрема, логарифмування даних, логістичні перетворення, обчислення квадратного кореня, перетворення Бокса-Кокса.

Матриця відстаней  $D$  містить усі можливі між елементами цільового та вхідного часових рядів попарні комбінації. Матриця відстаней, що має розмірність  $n \times m$ ,  $n$  рядків на  $m$  стовпчиків, кожен елемент матриці – це значення міри подібності  $d_{ji}$  між відповідними точками  $y_j$  та  $x_i$  ( $j$  – номер рядка (значення цільового часового ряду), а  $i$  – номер стовпчика (значення вхідного часового ряду), представлена у табл. 1.

Таблиця 1 – Матриця відстаней у загальному вигляді

	$x_1$	$x_2$	...	$x_m$
$y_1$	$d_{11} = d(y_1, x_1)$	$d_{12} = d(y_1, x_2)$	...	$d_{1m} = d(y_1, x_m)$
$y_2$	$d_{21} = d(y_2, x_1)$	$d_{22} = d(y_2, x_2)$	...	$d_{2m} = d(y_2, x_m)$
...				
$y_n$	$d_{n1} = d(y_n, x_1)$	$d_{n2} = d(y_n, x_2)$	...	$d_{nm} = d(y_n, x_m)$

Шлях в матриці відстаней  $D$  – це переміщення по елементах матриці відстаней, починаючи з елемента  $D[1,1]$  та закінчуючи елементом  $D[n,m]$ . При побудові шляху в матриці відстаней зазвичай задають обмеження на розширення та стиснення цільового ряду. В залежності від налаштувань параметрів обмежень  $RL$  (обмеження на розширення) та  $CL$  (обмеження на стиснення) виконується пошук оптимального шляху серед допустимих значень матриці відстаней.

Для деякого шляху  $\pi \in \Pi$  вводиться функція шляху, яка позначається так (1):

$$Path_{\pi} = \{Path_{\pi}(p) : p = 1, \dots, P\} = \{Path_{\pi}(1), \dots, Path_{\pi}(P)\}, \quad (1)$$

де аргумент  $p$  – це індекс функції шляху, який в загальному вигляді може приймати значення  $p = 1, \dots, P$ ,  $P$  – це довжина шляху в матриці відстаней  $D$ , тобто кількість переходів між початковим та кінцевим елементами шляху  $\pi$ .

Діапазон допустимих значень довжини шляху визначається інтервалом  $\max(n, m) \leq P \leq (n + m - 1)$ .

Для того щоб функція шляху відповідала властивостям часового ряду, повинні виконуватися такі умови: якщо  $Path_{\pi}(p) = (j_p, i_p)$  та  $Path_{\pi}(p+1) = (j_{p+1}, i_{p+1})$ , то  $(j_{p+1} - j_p) \in \{0,1\}$ ,  $(i_{p+1} - i_p) \in \{0,1\}$ .

Для оцінювання ступеня близькості часових рядів  $Y$  та  $X$  було запропоновано метрику, названу  $ScoreSim$ . Вона розраховується за наступною методикою.

Нехай для цільового ряду значень  $Y = \{y_1, \dots, y_n\}$  формується додатковий вхідний ряд  $Y_0$ , що містить обернені (помножені на мінус одиницю) значення  $Y$ , тобто  $Y_0 = \{-y_1, \dots, -y_n\}$ . Послідовність кроків для розрахунку метрики:

Крок 1. Побудувати матрицю відстаней  $D_0$ , розмірністю  $n \times n$ , на основі цільового  $Y$  та  $Y_0$  вхідного рядів даних;

Крок 2. Задати обмеження  $RL = 0$  та  $CL=0$ ;

Крок 3. Побудувати оптимальний для  $D_0$  шлях  $\pi_0$ , який буде мати діагональний вигляд  $\pi_0 = \{(1,1), \dots, (n, n)\}$ ;

Крок 4. Обчислити значення функції вартості шляху  $ScoreAvg_0 = ScoreAvg(\pi_0) = \frac{1}{n} \cdot \sum_{j=1}^n D_0[j, j]$ .

Отримане значення  $ScoreAvg_0$  буде використовуватися в якості значення-нівеліру для визначення подібності досліджуваних рядів.

Після цього для цільового ряду  $Y$  та вхідного  $X$ , за вищеописаним алгоритмом

1) будується матриця відстаней  $D_1$ ,

2) враховуються обмеження  $RL$  та  $CL$ ,

3) будується оптимальний для  $D_1$  шлях  $\pi_1$ ,

4) обчислюються значення функції вартості шляху  $ScoreAvg_1 = ScoreAvg(\pi_1)$ .

На основі отриманих значень функцій вартостей обчислюється коефіцієнт подібності за формулою (2):

$$ScoreSim(Y, X) = 100\% \cdot \left(1 - \frac{ScoreAvg_1}{ScoreAvg_0}\right) = 100\% - 100\% \times \left( \left( \frac{1}{(\max(n, m) - (P_1 - \max(n, m)))} \cdot \sum_{k=1}^{P_1} D_1[j_p, i_p] \right) / \left( \frac{1}{n} \cdot \sum_{k=1}^{P_0} D_0[j, j] \right) \right) \quad (2)$$

де  $P_0$  – довжина шляху  $\pi_0$ ;  $P_1$  – довжина шляху  $\pi_1$ ,  $\max(n, m)$  – повертає максимальне серед  $n$  та  $m$  значення.

За своєю суттю коефіцієнт  $ScoreSim(Y, X)$  відображає ступінь схожості між часовими рядами  $Y$  та  $X$ , виконуючи порівняння значення середньої вартості переходу по матриці відстаней між рядами  $Y$  та  $X$ , та середньої вартості максимально несхожих рядів  $Y$  та  $Y_0$ . Чим більше ряди  $Y$  та  $X$  несхожі один на одного, тим ближчим буде відношення  $\frac{ScoreAvg_1}{ScoreAvg_0}$  буде до одиниці, тому в наведеній формулі для  $ScoreSim(Y, X)$  пропонується саме вираз  $1 - \frac{ScoreAvg_1}{ScoreAvg_0}$ , що дозволяє отримати значення, яке буде сходиться до нуля у випадку несхожесті рядів  $Y$  та  $X$ , а у випадку максимальної близькості сходиться до одиниці (рис. 4).



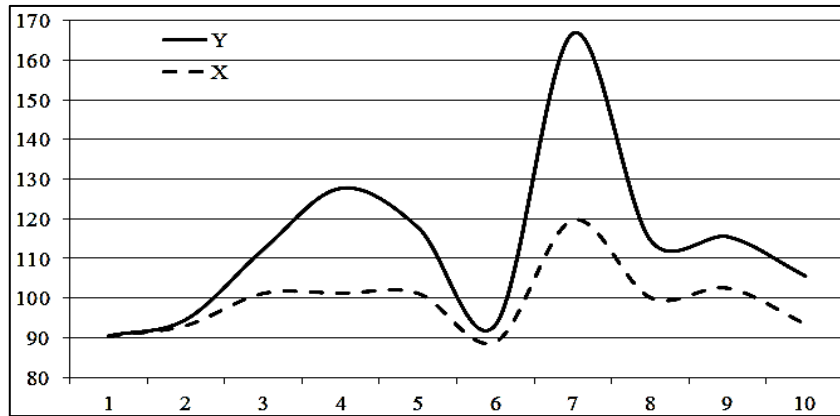


Рисунок 4 – Графіки цільового та вхідного рядів даних після виконання операцій перетворення з розширення та стиснення

Застосування запропонованого підходу дозволяє аналізувати та враховувати невизначеності у задачах прогнозування розвитку нелінійних нестационарних процесів, виявляючи поведінкові шаблони у досліджуваних часових рядах, що дає змогу визначати характер та динаміку досліджуваного процесу навіть за наявності значної кількості пропусків та спотворених чи неповних даних.

У четвертому розділі представлено розроблені моделі у формі мереж Байєса, призначені для урахування структурно-параметричних невизначеностей і адекватного опису причинно-наслідкових зв'язків та можливих варіантів розвитку подій під впливом різних груп детермінованих та випадкових чинників (табл.2).

Крім того, застосування запропонованого підходу доцільно застосовувати для оцінювання невідомих параметрів моделей, при побудові яких опрацьовуються короткі часові ряди даних.

Таблиця 2 – Методика застосування ймовірнісно-статистичних моделей для опрацювання коротких часових рядів та таких, що містять неповні чи спотворені дані

Етап	Пропоновані методичні підходи
Попередня підготовка даних	Збір інформації, попередня обробка даних, підготовка даних до аналізу
	Обробка пропусків, пошук аномалій
	Виявлення закономірностей у досліджуваних процесах, зв'язків у даних
	Зменшення простору вхідних змінних
Моделювання причинно-наслідкових зв'язків	<p>Етап 1. Скорочення розмірності задачі моделювання.</p> <p>Етап 2. Масштабування розподілів даних з метою їх приведення до зручної для подальшого використання форми й дискретизація неперервних змінних.</p> <p>Етап 3. Формулювання семантичних обмежень.</p> <p>Етап 4. Пошук структур моделей-кандидатів. На цьому кроці з множини можливих структур моделей вибирається декілька кращих моделей-кандидатів за допомогою відповідних критеріїв якості та оцінюються їх параметри.</p> <p>Етап 5. На цьому етапі виконується порівняння характеристик ймовірнісних моделей-кандидатів з метою вибору кращої для опису досліджуваного процесу</p>

Етап	Пропоновані методичні підходи
Прогнозування на основі адаптивного підходу з комбінованим використанням регресійних та ймовірнісно-статистичних моделей у формі мереж Байєса	<p>Етап 1. Представлення адитивної мережної моделі:</p> <p>Етап 2. Представлення вимірів незалежних змінних у вигляді вектора</p> <p>Етап 3. Представлення адитивної моделі із використанням довільних функцій</p> <p>Етап 4. Формування динамічної мережної моделі, за якою буде обчислюватись прогноз. Головною особливістю двокрокової методики моделювання є те, що параметри декомпозиції обчислюються повторно після отримання нових вимірів. За даної методики змінна <math>Y(k)</math> залежить від вимірів незалежних змінних у часі.</p> <p>Етап 5. Обчислення умовної ймовірності змінної <math>Y(k)</math> з використанням адитивної декомпозиції ймовірнісної моделі</p> <p>Етап 6. Отримання остаточного результату. На основі моделі розглянутого типу за узагальненим алгоритмом формується ймовірнісний висновок.</p>
Алгоритм ймовірнісного висновку	<ol style="list-style-type: none"> <li>1 Виконується адитивна декомпозиція мережі Байєса на окремі складові загальної мережі.</li> <li>2 Вивід окремих підмножин вузлів основної моделі виконується за L-S алгоритмом.</li> <li>3 Для кожної підмножини (кліки) вузлів обчислюється спільний розподіл ймовірностей. З цієї метою обчислюються ймовірності кількості значень категорійної змінної.</li> </ol>

Особливістю методики є використання ймовірнісно-статистичних методів для створення та застосування уніфікованих за структурою моделей у просторі станів. Для цього спочатку вирішується задача оцінювання елементів структури і параметрів моделі. Структура моделі оцінюється на підставі дослідження закономірностей протікання процесу, застосування статистичних тестів для перевірки наявності нелінійності, інтегрованості, гетероскедастичності, аналізу кореляційних функцій, візуального аналізу даних. При цьому вибирається кілька найбільш ймовірних структур моделей-кандидатів. Потім обчислюються оцінки параметрів моделей-кандидатів і обирається краща з них, використовуючи відповідні статистичні характеристики адекватності моделей.

Для побудови мережі Байєса в умовах наявності значної кількості невизначеностей, зокрема для випадку коли невідома структура, а спостереження є неповними, запропоновано метод побудови топології мережі із прихованими вершинами, в основу якого покладено модифікацію алгоритму максимізації математичного очікування (ЕМ-алгоритм).

Пропонована методика передбачає обчислення параметрів мережі Байєса за умови неповної вхідної інформації і відомої топології мережі, є комплексною системою, яка повністю описує весь процес знаходження невідомих параметрів прихованих вершин і складається з таких кроків:

Крок 1. Побудова мережі Байєса за навчальними даними або «вручну». Необхідно визначити топологію та знайти значення параметрів всієї мережі. Можливі два варіанти: коли структура мережі нам відома, і коли є лише навчальні дані. В другому випадку побудова структури здійснюється за два кроки: на першому кроці будується топологія мережі з використанням, наприклад, евристичного алгоритму, а на другому – визначаються параметри мережі, які максимально правдоподібні до навчальних даних.

Крок 2. Генерування вибірки за заданою структурою мережі (застосовується у випадку відсутності навчальних даних). У випадку, коли навчальні дані не задані або їх недостатньо, відбувається псевдовипадкове генерування вибірки за побудованою на першому етапі структурою мережі. Генерація відбувається наступним чином. Спочатку формується ймовірнісний висновок у мережі без інстанційованих вершин  $P(S_{ij})$ ,  $i = 1, \dots, N$ ;  $j = 1, \dots, S$ . Далі обирається вершина  $N_{i^*}$  і інстанціюється один із її станів  $S_{ij}$  із ймовірністю цього стану  $P(S_{ij})$ . Перераховуються ймовірності станів вершин після інстанціювання  $P(S_{ij} | S_{i^*} = S_{i^*j})$ . Далі обирається наступна вершина. Ця операція повторюється доки не залишаться неінстанційовані вершини. Інстанційовані стани утворюють запис у вибірці  $(S_{i_1} \dots S_{i_n})$ ,  $i_k \in (1, \dots, S)$ . Дії повторюється доки не буде створено необхідної кількості записів у вибірці.

Крок 3. Додавання прихованих вершин до структури мережі. Додавання прихованої вершини до мережі. У випадку, коли вершина вставляється між декількома існуючими, то видаляються попередні дуги між ними і створюються нові, які пов'язані з прихованим вузлом. Якщо необхідно додати приховану батьківську вершину, то просто створюється вершина і відповіді дуги.

Крок 4. Початкова ініціалізація невідомих параметрів мережі. Параметри прихованих вершин ініціалізуються початковими значеннями. Це можуть бути як випадково згенеровані значення, так і значення, що вказані експертами.

Крок 5. Обчислення параметрів мережі на основі згенерованих даних з використанням EM-алгоритму. На даному етапі запускається ітераційний процес EM-алгоритму, який використовує попередньо згенеровану вибірку даних і результатом якого є оцінки невідомих параметрів прихованих вершин мереж Байєса.

Для оцінювання якості роботи методу побудови ймовірнісного висновку запропоновано використовувати KL-відстань. Відстань  $D_K$  між значенням ймовірності  $P(X^{(i)}|e)$  та оцінкою  $\hat{P}(X^{(i)}|e)$  визначається за формулою (3):

$$D_K(P(X^{(i)}|e), \hat{P}(X^{(i)}|e)) = \sum_{\forall x^{(i)} \in A^{(i)}} \left( P(x^{(i)}|e) \cdot \log \left( \frac{P(x^{(i)}|e)}{\hat{P}(x^{(i)}|e)} \right) \right). \quad (3)$$

Ефективність застосування даного методу перевірено у процесі розв'язання практичних задач. Кращі результати прогнозування нелінійних нестационарних процесів отримані в результаті застосування адаптивно-оптимізаційного підходу до моделювання та прогнозування з використанням ансамблів та регресійних, узагальнених лінійних та ймовірнісно-статистичних моделей у формі мереж Байєса.

**У п'ятому розділі** розроблено системну методологію побудови моделей нелінійних нестационарних процесів на основі адаптивного підходу до моделювання з комбінованим використанням сценарного моделювання, регресійних та ймовірнісно-статистичних моделей у формі мереж Байєса, багатовимірних розподілів та узагальнених лінійних моделей, яка відрізняється комплексним застосуванням множини ймовірнісно-статистичних методів виявлення та врахування невизначеностей (фільтрація, оцінювання структури і параметрів математичних моделей), і забезпечує підвищення якості проміжних і остаточних результатів обробки даних, побудови математичних моделей і оцінювання прогнозів (рис. 5).

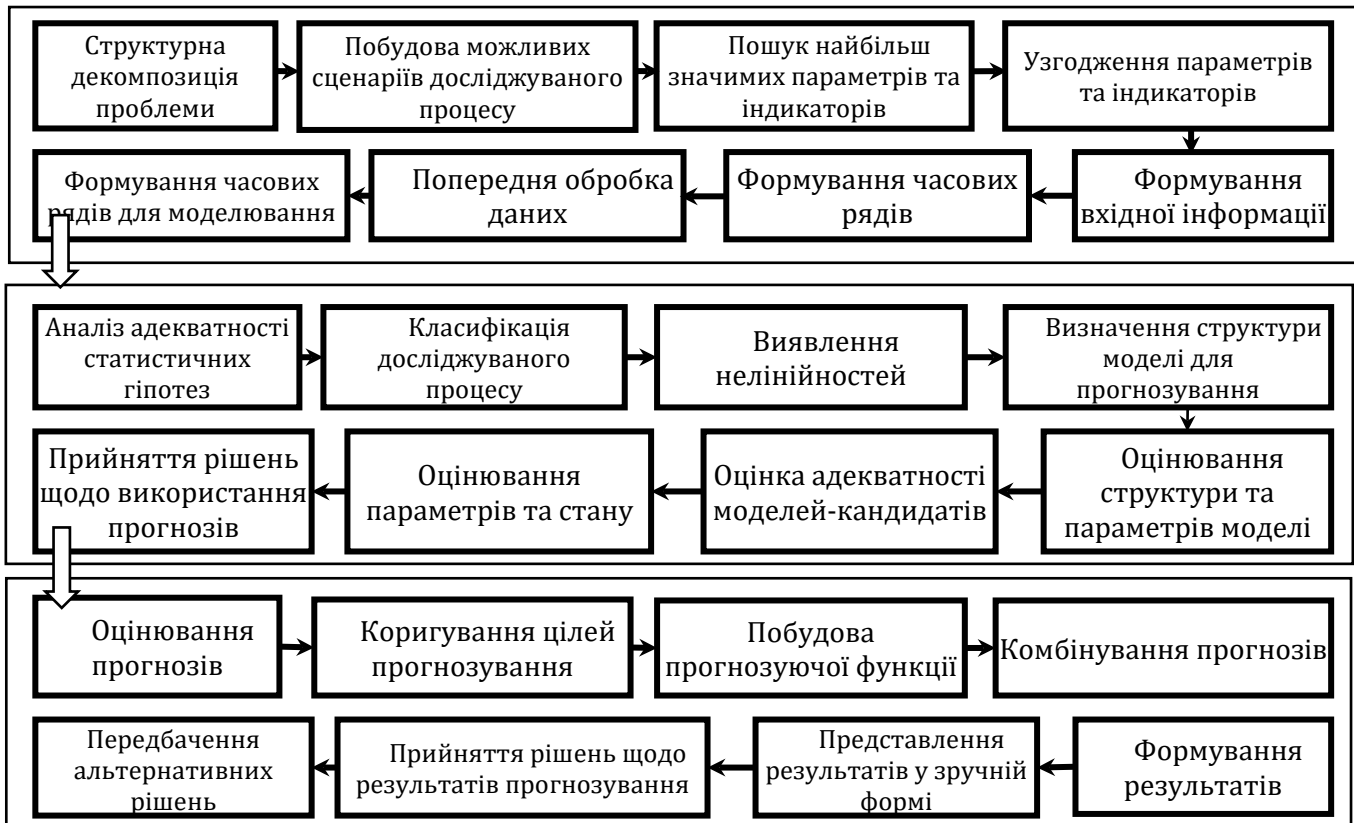


Рисунок 5 – Загальна схема застосування пропонованого багатомодельного підходу для моделювання нелінійних нестационарних процесів

Часто нелінійні нестационарні процеси характеризуються тим, що дані, які їх описують, можуть бути неповними або недостовірними. Для розв'язання даної задачі запропоновано багатомодельний підхід; в основу цього методу покладено інтегроване використання теорії подібності процесів та ймовірно-статистичного моделювання (рис. 6).

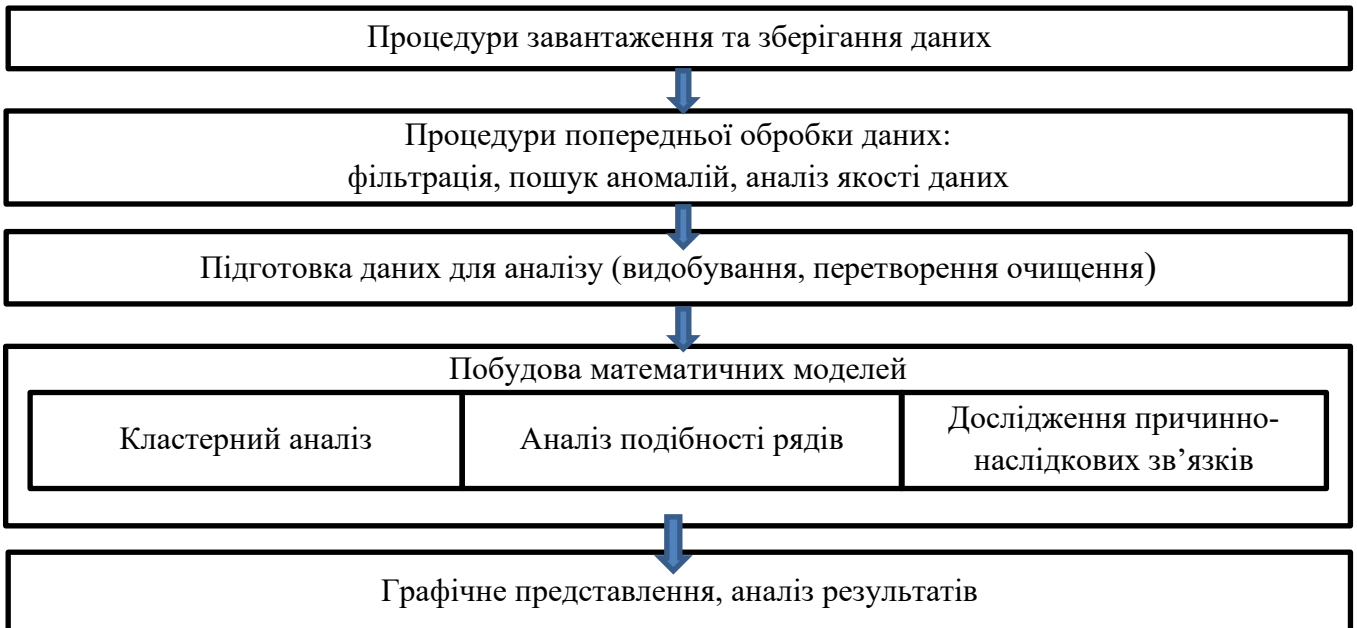


Рисунок 6 – Схема багатомодельного підходу із використанням теорії подібності процесів та ймовірно-статистичного моделювання

На основі запропонованої загальної методики побудови прогнозів на основі запропонованих підходів до моделювання нелінійних нестационарних процесів розроблено методику застосування моделей різного типу та їх комбінацій для прогнозного моделювання залежно від горизонту прогнозування. Передбачено, що в загальному випадку математична модель окрім історичних даних також повинна враховувати різні найбільш значимі чинники, суттєві для аналізу досліджуваних процесів, які необхідно включати в модель відповідно до побудованого сценарію (рис. 7).

Етап	Характеристика етапу			
<b>Завантаження вхідних даних</b>	Завантаження часових ряди історичних даних			
<b>Діагностика даних</b>	Визначення аномалій та їх обробка (виключення, згладжування).			
<b>Обробка пропусків</b>	Заповнення пропусків			
<b>Формування набору додаткових чинників</b>	Додаються чинники, суттєві для аналізу досліджуваних процесів (дані, що надаються за бажанням користувача, довідкових дані, даних щодо екзогенних параметрів процесу, експертні оцінки, дані, що доповнюють відомості про аналізований процес)			
<b>Строк</b>	Дуже короткострокове	Короткострокове	Середньострокове	Довгострокове
<b>Особливості побудови прогнозу</b>	на кожну годину, враховуються цикли всередині доби.	на добу (тиждень) вперед, враховуючи цикли та закономірності в середині тижня.	Трендова складова, тренд та комбінований фактор з іншими регресорами, похідні показники	Враховуються додаткові чинники, користувача (на довгострокову перспективу)
<b>Включення додаткових чинників</b>	Додаткові регресори, що описують процес	Додаткові регресори, що описують процес.	Враховуються додаткові чинники, користувача	В залежності від предметної області
<b>Горизонт прогнозування</b>	від 1 до 24 годин	від доби до місяця	від 1 місяця до 3 років	більш 3 років
<b>Сценарій</b>	оптимістичний, реалістичний, песимістичний			

<b>Побудова моделей-кандидатів</b>	На кожному горизонті прогнозування будується модель за історичними даними та за даними користувача
<b>Етап 1</b>	Одноступеневі моделі (експоненційні, регресійні, авторегресійні, узагальнені лінійні моделі)
<b>Етап 2</b>	Двоступеневі моделі (регресійні та авторегресійні моделі, моделі із включенням трендової складової та урахуванням залишків (різниця між реальним та прогнозним значенням), що включаються до моделі у вигляді ковзного середнього, за умови, що між залишками та цільовою змінною є кореляція (автокореляція). Моделі класу експоненційного згладжування – Хольта, Тейла-Вейджа, Брауна, Вінтерса (з адитивною або мультиплікативною сезонною складовою), з урахуванням демпфуючого тренду та інші модифікації, Узагальнені лінійні моделі, нейронні мережі, ймовірнісні моделі та нечіткі методи.
<b>Вибір кращої моделі-кандидата</b>	Вибір за результатами виконання етапу 1 та етапу 2, на основі: середньої абсолютної процентної похибки (MAPE), максимального MAPE, коефіцієнт детермінації ( $R^2$ ), середньоквадратична похибка (RMSE).
<b>Побудова прогнозу</b>	Використовуються кращі моделі-кандидати, вхідні дані: часові ряди досліджуваних процесів та додаткові чинники, що є значимими на даному горизонті прогнозування.

Рисунок 7 – Методика застосування моделей різного типу та їх комбінацій для прогнозного моделювання залежно від горизонту прогнозування

Для оцінювання якості прогнозів, обчислених за побудованими моделями, запропоновано процедуру автоматизованого вибору кращої прогнозної моделі, у якій для вибору було запропоновано інтегральний критерій якості, який включає 2-5 окремих статистичних критеріїв якості. Процедура автоматизації забезпечує можливість побудови, аналізу та вибору кращої із множини можливих моделей, кількість яких може сягати кількох сотень. Витрати часу залишаються при цьому цілком прийнятними для практичного використання запропонованої методики у автоматизованій системі, призначеній для прогнозного моделювання нелінійних нестационарних процесів.

У шостому розділі представлено інформаційну технологію моделювання і прогнозування нелінійних нестационарних процесів в умовах невизначеності, нестационарності та ризику на основі використання сучасних методів інтелектуального аналізу даних, моделювання та прогнозування. Зокрема, формулювання та розв'язання задач прогнозування нелінійних нестационарних процесів ґрунтується на застосуванні методології системного аналізу, яка ґрунтується на комплексному використанні методів попередньої обробки і статистичного аналізу даних, математичного і статистичного моделювання, прогнозування та оптимального оцінювання станів процесів довільної природи. Використання запропонованої концепції забезпечує отримання високоякісних (за точністю) коротко- та середньострокових прогнозів за умови наявності інформативних даних, а також формування на їх основі альтернативних оптимальних і раціональних рішень, що також передбачає універсальність застосування запропонованої концепції до аналізу широкого класу нелінійних нестационарних процесів. На рис. 8 подано структурну схему системного підходу до організації процесу прогнозування.

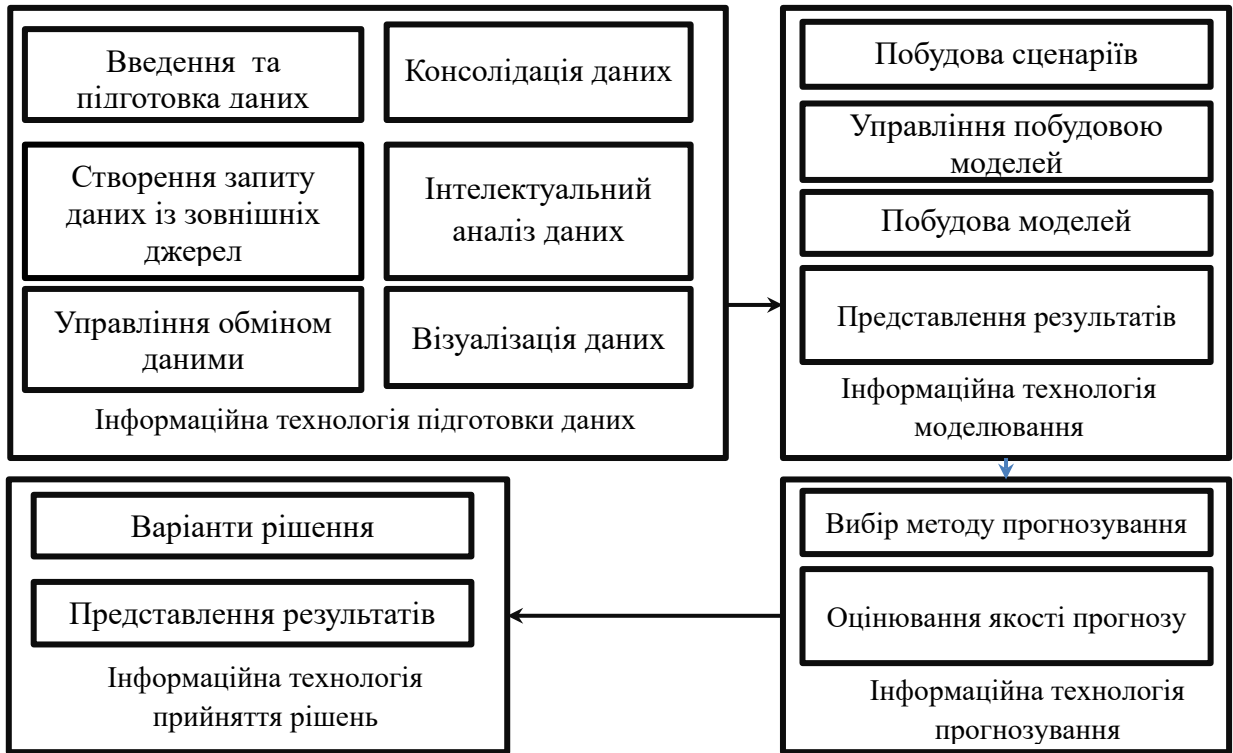


Рисунок 8 – Структурна схема системного підходу до організації процесу прогнозування при вирішенні завдань дослідження

Структура інформаційної технології підтримки прийняття рішень подана на рис. 9.

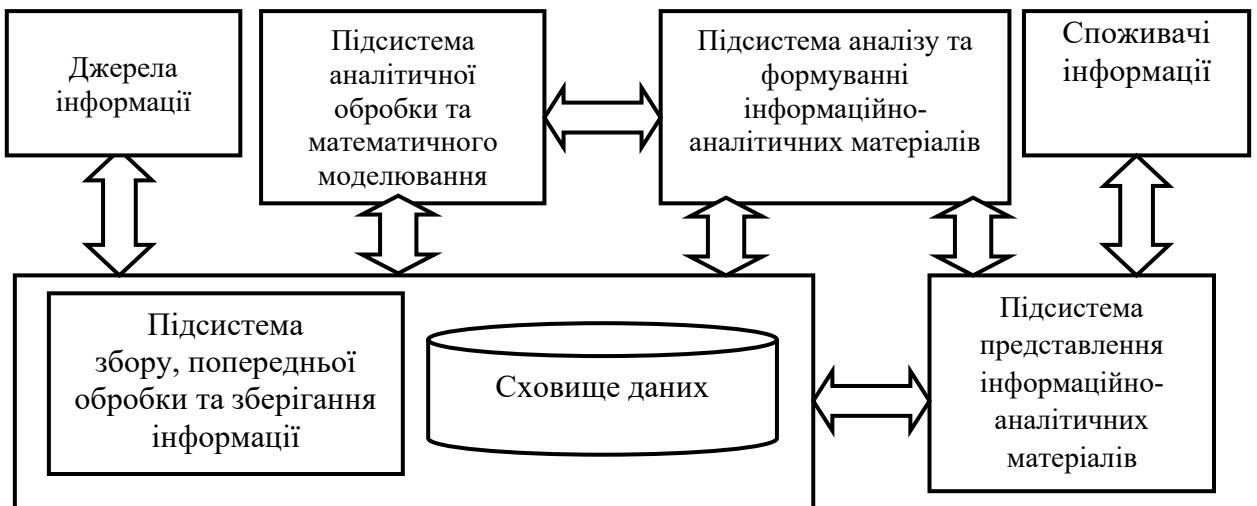


Рисунок 9 – Структура інформаційної технології підтримки прийняття рішень

Пропонований підхід ґрунтується на аналізі досліджуваного процесу, встановленні типів наявних характерних невизначеностей, оцінюванні структури і параметрів моделі та прогнозів. Для розв'язання цього комплексу задач спроектовано СППР з відповідною архітектурою (рис.10).



Рисунок 10 – Спрощена концептуальна схема СППР

Пропонована архітектура СППР дозволяє будувати додатки, за допомогою яких вирішуються комплекс задач прогнозування нелінійних нестационарних процесів. Представлені в роботі СППР мають гнучку модульну структуру, призначені для аналізу даних у вигляді часових рядів сформованих на основі структурованої та неструктурованої інформації.

Блок збору та підготовки даних орієнтований, в тому числі, на використання API-інтерфейсу, що дає можливість опрацьовувати великі обсяги даних, отримуваних з мережі Інтернет, виконувати весь комплекс робіт попередньої обробки даних, в тому числі і з використанням інструментів інтелектуального аналізу даних. У блоці зберігання даних можуть бути використані системи управління базами даних такі як, MySQL, PostgreSQL та аналогічні. У блоці прогнозного моделювання передбачено використання таких аналітичних систем як SAS, R, Python, є можливість інтеграції до інших аналітичних платформ. Блок прогнозного моделювання має модуль тестової перевірки ідей та гіпотез, що дає можливість виконувати значну кількість чисельних експериментів для апробації пропонованих методик прогнозного моделювання. Також у блоці прогнозного моделювання передбачено можливість вибору горизонту прогнозування та побудови сценаріїв. Особливістю модуля представлення



результатів є можливість користувачам самостійно генерувати звіти про якість прогнозів.

Пропонована архітектура СППР має переваги інтеграції з іншими технологічними рішеннями (аналітичними платформами) (рис. 11).

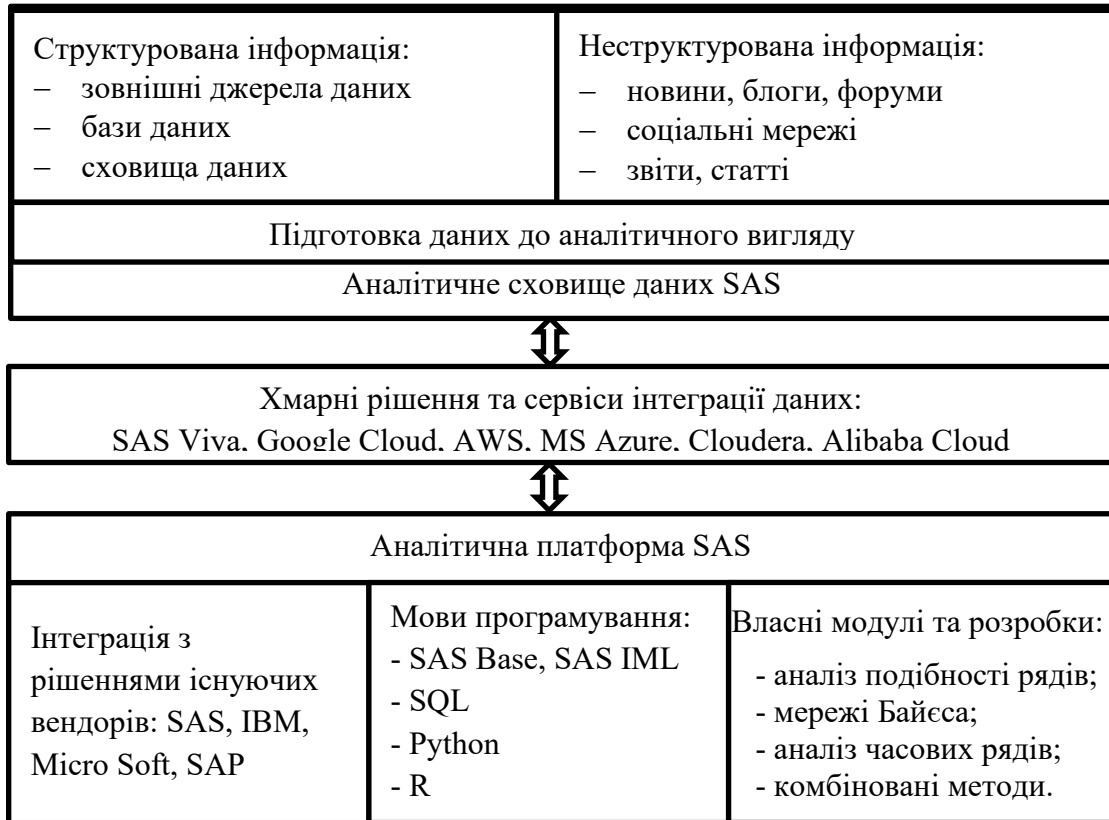


Рисунок 11 – Архітектура СППР на базі SAS Foundation

Така конфігурація дозволяє багатьом клієнтам з'єднуватися з серверами SAS Foundation. У конфігурації клієнт-сервер до загального набору конфігурації «Клієнтський комп'ютер» додається компонент «Аналітична платформа» (Analytics Platform). Аналітична платформа забезпечує зв'язок між клієнтом SAS Enterprise Miner, SAS Foundation, серверами Metadata Server.

Аналітичне сховище даних призначене для зберігання та інтеграції як структурованої так і неструктурованої інформації у вигляді, придатному для аналізу та моделювання. Система формує набори метаданих, призначені для використання у аналітичній системі. Для певних типів аналізу можуть існувати інші специфічні метадані. Схема визначення метаданих в аналізі для вибірки дозволяє повідомляти SAS Enterprise Miner передати таку інформацію:

- налаштування ролі кожної змінної в аналізі (цільова або регресорна змінна);
- рівень вимірювань кожної змінної – розрізняє безперервні чисельні і категоріальні змінні;
- роль набору даних в аналізі – повідомляє системі, як використовувати вибраний набір даних в аналізі (навчальний, валідаційний, тестовий, скорінговий, транзакційний, корпус текстів).

Аналітична платформа дозволяє використовувати як готові аналітичні рішення від існуючих вендорів, так і власні методи та підходи реалізовані у вигляді модулів програм на загальних та спеціалізованих мовах програмування. Хмарні рішення та сервіси інтеграції даних дозволяють виконувати задачі по розгортанню аналітичних рішень, баз та сховищ даних, налаштування програмних інтерфейсів для запуску побудованих аналітичних моделей та обміну з ними даними від різноманітних сервісів та рішень.

Аналітичне сховище даних призначене для зберігання та інтеграції як структурованої так і не структурованої інформації у вигляді, придатному для аналізу та моделювання. Аналітична платформа дозволяє використовувати як готові аналітичні рішення від існуючих вендорів, так і власні методи та підходи реалізовані у вигляді модулів програм на загальних та спеціалізованих мовах програмування. Хмарні рішення та сервіси інтеграції даних дозволяють виконувати задачі по розгортанню аналітичних рішень, баз та сховищ даних, налаштування програмних інтерфейсів для запуску побудованих аналітичних моделей та обміну з ними даними від різноманітних сервісів та рішень.

Окрім цього розділ містить інформацію щодо створення СППР для вирішення завдань моделювання та прогнозування нелінійних нестационарних процесів.

**У сьомому розділі** наведено низку практичних прикладів використання запропонованої інформаційної технології у СППР для моделювання нелінійних нестационарних процесів на практичних прикладах.

**Приклад 1.** Застосування багатомодельного підходу, в основу якого покладено використання комбінацій регресійних моделями та байєсівських мереж для прогнозування нелінійних нестационарних процесів у еколого-економічних системах. Для застосування даного підходу запропоновано методику, реалізовано за допомогою наступних кроків:

Крок 1. Завантаження даних.

Крок 2. Розділення даних на 2 підвибірки: навчальну та перевірочну.

Крок 3. Побудова моделей.

Крок 4. Порівняння моделей за статистичними критеріями якості.

Крок 5. Побудова загального звіту за виконаним обчислювальним експериментом.

Прогнозування нелінійних нестационарних процесів виконано на матеріалах урожайності зернових. Нестационарність досліджуваного процесу підтверджено розрахунками значення математичного сподівання показника урожайності, яке відрізняється у аналізовані роки у 1,62 рази, а дисперсія – у 6,13 рази. Вхідними даними для побудови моделей є часові ряди значень NDVI та погодних факторів. Побудована модель лінійної регресії має вигляд (4):

$$\begin{aligned} \text{Productivity} = & 25,352 \cdot \text{Intercept} + 89,646 \cdot \text{ndvi}_{17\_18} - 29,044 \cdot \text{ndvi}_{19\_20} + \\ & 74,230 \cdot \text{ndvi}_{23\_24} - 89,578 \cdot \text{ndvi}_{25\_26} + 52,028 \cdot \text{ndvi}_{27\_28} + 0,330 \\ & \text{temp}_{17\_18} + 0,711 \cdot \text{temp}_{19\_20} - 4,118 \cdot \text{temp}_{21\_22} \end{aligned} \quad (4)$$

Регресори типу  $ndvi_{17\_18}$  позначають різницю між значенням нормалізованого відносного індексу вегетації (NDVI) на 18-у та 17-у тижнях; відповідно,  $temp_{17\_18}$  позначає різницю між температурою навколишнього середовища на 18-у та 17-у тижнях вегетації.

В ході дослідження було побудовано низку регресійних моделей з різними комбінаціями регресорів – факторів, що впливають на урожайність зернових (табл. 3).

Таблиця 3 – Статистичні характеристики регресійних моделей, побудованих для прогнозування урожайності озимої пшениці

Фактори моделі множинної регресії	RMSE	MAPE	DW
Тільки добрива	1,19	24,78	2,25
Тільки культура попередник	1,24	26,39	2,07
Добрива та культура попередник	1,17	24,59	2,31
Добрива, культура-попередник, добрива-культура - попередник	1,17	24,59	2,32
Тільки температура	0,93	20,58	1,41
Тільки опади	1,05	19,43	1,51
Тільки температура та опади	0,69	13,62	1,17
Добрива, культура попередник, температура та опади	0,52	10,01	1,95

Також були побудовані моделі за методом LARS з вибором моделі за значенням ASE та SBC на перевіірочній вибірці, методом PLS, моделювання залежності урожайності озимої пшениці від факторів навколишнього середовища за лінійною регресією з відбором значущих змінних за допомогою дерева рішень, за методом градієнтного бустингу, дві моделі у формі мережі Байєса. Для побудови другої мережі Байєса, робились такі перетворення:

1. Обчислені перші різниці від показників NDVI, температури та кількості опадів.
2. Дані перетворені в дискретну форму за методом рівних проміжків

Таблиця 4 – Значення показників якості регресійних моделей залежності врожайності озимої пшениці від NDVI і погодних факторів

Модель	Вибірка					
	Навчальна			Перевірочна		
	RMSE	SSE	MAX E	RMSE	SSE	MAX E
Регресійна	5,21	977,74	13,64	5,54	459,81	11,73
LARS	5,21	977,73	13,64	5,57	465,68	11,74
LARS за критерієм мінімального SBC	5,54	1104,23	15,74	6,3	595,91	12,65
PLS	5,21	977,73	13,64	5,57	465,68	11,74

Градiєнтного бустингу	3,63	475,25	10,58	4,94	366,78	13,04
-----------------------	------	--------	-------	------	--------	-------

Як видно з табл. 4 найкращою виявилась модель, побудована за методом градієнтного бустингу. Досліджуючи отримані результати моделювання залежності врожайності озимої пшениці від NDVI та погодних факторів на основі регресійних моделей, можна зробити висновок, що найкраща модель отримана за методом градієнтного бустингу. Показники якості на навчальній вибірці такі: RMSE = 3,63; SSE = 475,25; MAX E = 10,58; на перевіірочній вибірці: RMSE = 4,94; SSE = 366,78; MAX E = 13,04). Інші моделі показали гірші і схожі результати: RMSE від 5,21 до 5,54; SSE від 977,73 до 1105,32; MAX E від 13,64 до 15,74. На перевіірочній вибірці отримано такі значення: RMSE від 5,54 до 6,42; SSE від 459,81 до 595,91; MAX E від 11,73 до 13,16.

Також побудовані мережі Байєса за двома методами: за вихідними даними, перетвореними в інтервальні значення за методом оптимального розбиття на групи з використанням алгоритму дерева рішень. Для побудови мереж використано спеціальний компонент Optimal Binning в системі SAS Enterprise Miner за вихідними даними, переведеними в різницеву форму і в інтервали за методом рівних інтервалів.

Для порівняння якості побудованих регресійних моделей з мережами Байєса, розраховано такі показники:

- для мереж Байєса розраховано відношення правильно класифікованих значень врожайності і вибиралась та мережа Байєса, у якої це показник вищий;
- для регресійних моделей вхідні дані показника врожайності і показника врожайності, розрахованого за відповідною моделлю, перетворено в інтервали відповідно до мережі Байєса, з якою виконувалось порівняння.

Отримані мережі Байєса порівнювались за загальною точністю класифікації, показники якої склали 55,36% та 70,98% для кукурудзи, і 78,43% та 47,06% для пшениці.

Порівнюючи результати прогнозування врожайності для обох досліджуваних культур, отримані за регресійними моделями та байєсівськими мережами, кращими є отримані методом градієнтного бустингу (відсоток класифікації на тестовому наборі даних склав: мережі Байєса – 78,43%, метод градієнтного бустингу – 86,11%, регресійні моделі, PLS, – 69,44, PLS, LARS – 69,44%).

Серед регресійних моделей урожайності пшениці, кращою виявилась модель, отримана за методом градієнтного бустингу. Показники якості на навчальній вибірці такі: RMSE = 3,63, SSE = 475,25, MAX E = 10,58; на перевіірочній вибірці: RMSE = 4,94, SSE = 366,78, MAX E = 13,04. Серед байєсівських мереж, кращими за показником точності класифікації виявилися такі: для кукурудзи – модель, побудована за вихідними даними, переведеними в інтервальний вимір за методом рівних інтервалів (точність класифікації – 70,98%); для озимої пшениці – мережа, побудована за вихідними даними, переведеними в інтервальний вимір за методом градієнтного бустингу (точність класифікації – 78,43%).

**Приклад 2.** Математичні методи та підходи та методи довго-, середньо-, короткострокового та дуже короткострокового прогнозування навантаження енергосистеми були застосовані для вирішення задачі прогнозування нелінійних

нестационарних процесів в енергетиці. Для проведення обчислювальних експериментів та оцінювання коефіцієнтів рівнянь було використано реальні статистичні дані однієї з енергетичних компаній України. Схема застосованої методики представлена на рис. 7.

Для побудови довгострокової математичної моделі на річних інтервалах було використано такі макроекономічні чинники, як значення ВВП, загально чисельності населення та чисельність працюючих на виробництві, а також індекс, що відображає стан розвитку економіки та кількість населення. Отримана модель має характеристики  $R^2 = 0,98$ ,  $MAPE=0,2\%$ .

При середньостроковому прогнозуванні, на місячних відрізках даних, було побудовано модель авторегресії дванадцятого порядку та Вінтерса з мультиплікативною сезонністю, для яких значення статистики  $MAPE$  становить 2,64% та 2,6% відповідно.

Менш точні результати було отримано при короткостроковому прогнозуванні на кварталних даних, похибка авторегресії четвертого порядку сягнула значення статистики  $MAPE$  в 6,9%.

Для випадку короткострокового прогнозування на годинних часових проміжках, навчання моделей відбувалося на даних з 1 січня 2015 року по 31 вересня 2017 року, а тестування на відрізок від 1 жовтня до 31 грудня. Перед початком моделювання було досить ретельно виконано графічний та статистичний аналіз даних на наявність типових патернів та відмінностей поведінки за різних умов. Було виявлено періоди різкого зростання та спаду споживання електроенергії, що використано при побудові моделей в якості додаткових чинників.

Аналіз даних в добовому та місячному розрізах температури та навантаження виявив залежність у вигляді оберненої фази в залежності від доби року. Як наслідок, температуру необхідно враховувати в моделі у вигляді поліному третього порядку разом із спеціалізованими комбінованими змінними на основі температури та місяця, температури та години доби, температури та дня неділі.

Було побудовано чотири моделі для короткострокового прогнозування у вигляді нейронної мережі, експоненційної моделі з мультиплікативною сезонністю, GLM моделі ARIMAX та GLM моделі із урахуванням свят, статистична характеристика  $MAPE$  дорівнює 3.19%, 2.55%, 2.43% та 2.39% відповідно.

Більш детальний аналіз результатів показав, що загалом 93% попереднього значення навантаження враховується в наступному, при цьому присутній тренд в часових даних, а підмішування в модель регресорів що описують свята підвищує точність прогнозування за критерієм  $MAPE$  на 0,04%.

Узагальнення результатів обчислювального експерименту представлено в таблиці 5, критерієм для порівняння якості побудованих моделей використано показник середньої похибки в процентах.

Таблиця 5 – Значення середньої похибки в процентах ( $MAPE$ ) для різних класів моделей за реалістичного сценарію на різних горизонтах прогнозу, %

Назва моделі	Горизонт прогнозування
--------------	------------------------

	Дуже коротко- строкова модель, на 24 години вперед	Коротко- строкова модель, на 2 тижні вперед	Середньо- строкова модель, на 3 місяці вперед	Довгострокова модель
GLM-B	3,62	2,97	1,46	1,61
GLM-BR	3,32	2,74	2,15	1,67
GLM-BRW	3,32	2,74	2,15	1,67
GLM-BRWH	3,27	2,79	2,09	1,83
GLM UCM	2,68	2,05	1,70	3,13
Нейронна мережа	3,19	2,59	2,08	3,88
Експоненційне згладжування	2,55	2,01	1,86	3,15
GLM модель ARIMAX	2,39	1,97	1,57	3,09
Комбінована модель	2,44	2,03	1,75	1,82

Застосування мультимодельного підходу, попередньої підготовки даних дозволяє одержати значно кращі результати прогнозування, а використання програмного забезпечення SAS Energy Forecasting – урахувати галузеві особливості досліджуваної компанії чи галузі національної економіки.

**Приклад 3.** У роботі запропоновано багатомодельний підхід, застосування якого у системі підтримки прийняття інвестиційних рішень на ринку криптовалют забезпечило отримання прогнозів високої якості та вибір оптимальної торгівельної стратегії навіть в умовах невизначеності за необхідності швидкого прийняття рішень. Схема запропонованої системи підтримки прийняття рішень має структуру, представлену на рис. 10. Слід зазначити, що модуль управління збором даних з бірж орієнтований на використання API інтерфейсу та роботу переважно з біржами Bitfinex, Bitstamp, Huobi і т. ін.; у модулі зберігання даних з використанням баз даних переважно застосовуються системи управління базами даних MySQL, PostgreSQL та аналогічні; у модулі прогнозного моделювання передбачено використання таких аналітичних систем як SAS, R, Python, особливістю блоку прогнозного моделювання є наявність модуля тестової перевірки ідей та гіпотез, що викликано необхідністю проведення значної кількості чисельних експериментів з метою апробації запропонованих методик прогнозного моделювання; у блоці прогнозного моделювання також реалізовано модуль вибору стратегії торгівлі криптовалютою; модуль представлення результатів надає можливість користувачам додаткового генерувати звіти про якість прогнозів на ретроспективних даних, а також про прибутковість поточних угод.

Використаний у дослідженні набір даних про угоди щодо купівлі-продажу криптовалюти складається з 957510 рядків, кожний з яких містить агреговану інформацію за кожні 15 секунд роботи біржі за період с 24.06.2018 по 18.12.2018 г. Агрегована інформація складається з таких елементів, як часова мітка в форматі день-місяць-рік-година-хвилина-секунда (DateTime), планова ціна пропозиції 1 біткоіна

(Bid), загальний оборот на біржі за 15 секунд (Aggregated\_tick\_volume), кількість угод на біржі за кожні 15 секунд її роботи (Trades\_count).

На етапі попередньої обробки даних було досліджено фактори, що впливають на кількість та періодичність утворення пропусків у часових рядах біржових даних щодо торгів біткоїн. Чисельні експерименти показали, що наявності на годинному фреймі 7,23% часових інтервалів містять від 80 до 100% пропусків, і 4,99% годинних інтервалів – від 50 до 80% пропусків. Однак, на добових інтервалах відсоток пропусків складає 5,74% і 4,92%, відповідно.

Для прогнозування обсягів торгів криптовалюти в рамках відповідного часового інтервалу на основі часових рядів, що містять пропуски даних, пропонується наступний підхід:

Крок 1. обчислити кількість хвилин за той проміжок часу для якого є дані щодо прогнозованого фінансового інструменту за формулою (5)

$$Indicator60 = \sum_{t=1}^{60} I(t), \quad (5)$$

де  $I(t) = 1$ , якщо для відповідного  $t$ -го хвилинного фрейму значення ряду заповнено, в іншому випадку  $I(t) = 0$ .

Крок 2. обчислити сумарне значення досліджуваного показника для годинного часового фрейму на основі наявних хвилинних фреймів (6):

$$Volume60 = \sum_{t=1}^{60} Volume(t), \quad (6)$$

де  $Volume(t)$  – значення показника хвилинного фрейму  $t$ .

Крок 3. обчислити уточнене значення досліджуваного показника для відповідного часового фрейму (7):

$$Volume\_bias\_60 = Volume60 \cdot \frac{60}{Indicator60} \quad (7)$$

Для інших часових фреймів використовується аналогічний підхід. Лише базис для розрахунків використовується рівний не 60, а фактичному значенню. Наприклад, для чотирьохгодинних фреймів, він дорівнює 240.

В таблиці 6 наведено порівняння результатів прогнозування обсягів торгів валютної пари BTC-USD після застосування різних методів експоненційного згладжування для заповнення пропусків у часових рядах. В якості критерію для порівняння результатів прогнозного моделювання, використано показник середньоквадратичного відхилення (RMSE) на годинному та добовому фреймах.

Таблиця 6 – Порівняння результатів прогнозування обсягів торгів валютної пари BTC-USD за застосування різних методів експоненційного згладжування

Модель	Значення показника середньоквадратичного відхилення (RMSE)	
	на 15-ти хвилинному фреймі	за добового фрейму
Експоненційного згладжування з сезонністю	1,16	24,32
Вінтерса з адитивною складовою	1,17	17,01

Демпфуючого тренду	1,09	15,65
Подвійного експоненційного згладжування (модель Брауна)	1,35	18,57
Експоненційного згладжування з лінійним трендом (модель Хольта)	1,11	16,45
Експоненційна з мультиплікативною сезонністю	1,18	19,75
Просте експоненційне згладжування	1,09	16,71
Вінтерса з мультиплікативною складовою	1,21	21,66

Як видно з таблиці 6, найкращий результат одержано за застосування для заповнення пропусків у часових рядах даних моделей експоненційного згладжування з демпфуючим трендом.

Поряд із математичним моделюванням були застосовані інструменти технічного аналізу, використання яких також дозволяє аналізувати та передбачати загальні тенденції на ринку біткоїнів зокрема: осциляторами та індикаторами, такими як MACD, RSI, SMA та ін. У чисельному експерименті було використано індикатор Order Indicator. Ще однією частиною чисельного експерименту є дослідження ефективності застосування математичних моделей за запропонованого варіанту адаптивного методу прогнозування. Дослідження виконане на чотирьохгодинному часовому фреймі, за показниками якості прогнозного моделювання. Кращою виявилась математична модель нейронної мережі (двошарова мережа прямого розповсюдження). Перший прихований шар формується на основі вхідних змінних  $Amount_k$  з двох елементів  $H_{11}$  і  $H_{12}$ , а на основі змінних  $Volume_k$  - другий прихований шар – також з двох елементів  $H_{21}$  та  $H_{22}$ , в той час як змінні  $Price(t-1), \dots, Price(t-12)$  напряму впливають на цільову змінну. Особливістю цієї мережі є те, що в кожному із прихованих шарів використовуються по два прихованих елементи. Для визначення оптимальної кількості прихованих елементів у прихованому шарі у загальному випадку можуть використовуватися різні підходи: простий емпіричний перебір; кластерний аналіз, виконуваний для всього набору прецедентів з метою виявлення оптимальної кількості кластерів, у які можуть бути об'єднані прецеденти; метод головних компонентів.

У нейронній мережі для визначення оптимальної кількості прихованих елементів у прихованому шарі використано метод головних компонентів та експертні оцінки фахівців. В якості функції активації використано гіперболічний тангенс. Тобто, математична модель нейронної мережі має вигляд (8):

$$g_0^{-1}(E(Price(t))) = \omega_0 + \omega_1 \cdot H_{11} + \omega_2 \cdot H_{12} + \omega_3 \cdot H_{21} + \omega_4 \cdot H_{22} + \omega_5 \cdot Price(t-1) + \dots + \omega_{16} \cdot Price(t-12) \quad (8)$$

$$\begin{aligned} \text{де } H_{11} &= g_{11}(\omega_{01} + \omega_{11} \cdot Amount_{16} + \dots + \omega_{71} \cdot Amount_{96}), \quad H_{12} = g_{12}(\omega_{02} + \omega_{12} \cdot Amount_{16} + \dots + \omega_{72} \cdot Amount_{96}), \\ H_{21} &= g_{21}(\omega_{03} + \omega_{13} \cdot Volume_{16} + \dots + \omega_{73} \cdot Volume_{96}), \quad H_{22} = g_{22}(\omega_{04} + \omega_{14} \cdot Volume_{16} + \dots + \omega_{74} \cdot Volume_{96}) \end{aligned}$$

За результатами розрахунку показників якості моделей, значення похибки RMSE рівне 31, а правильне визначення напрямку тренду ціни відбулось у 63% випадків.



За дотримання даної стратегії доходність угод з «купівлі-продажу» склала 4900 дол США. У розрахунку доходності враховувався чистий прибуток після вирахування комісії біржі за угоди. Всього, під час експерименту було виконано 139 торговельних операцій, з яких 57 – на підвищення курсу біткоїну, а 82 - на його зниження, середня доходність по кожній угоді склала 35 дол США.

Отже, на підставі проведеного експерименту запропоновано універсальну методику для формування інвестиційної політики на ринку криптовалют, яка реалізується у чотири етапи.

Етап 1. На основі наявних статистичних історичних даних щодо реалізації обраних стратегій протягом місяця розрахувати:

- вектор доходності кожної зі стратегій протягом місяця;
- коваріаційну матрицю зміни доходності всіх стратегій.

Етап 2. Для подальшого аналізу обирати стратегії з позитивну доходністю.

Етап 3. Використовуючи методи оптимізації на основі вектору доходностей стратегій та коваріаційної матриці сформувані відповідну інвестиційну політику. При цьому застосувати лінійні обмеження експертного характеру, зокрема, частка кожної стратегії не має перевищувати порогове значення (наприклад, 30%) та при цьому бути більшою за нуль, щоб забезпечити диверсифікацію інвестиційного портфелю.

Етап 4. Після формування інвестиційної політики, наявні активи розділяються у визначеній пропорції для кожної з рекомендованих стратегій та протягом місяця використовуються для біржових операцій. По закінченню місяця усі наведені на кроках 1 – 3 операції повторюються та формується рекомендація щодо оптимальної інвестиційної політики на наступний місяць.

Ефективність запропонованої методики була перевірена експериментальним шляхом. Вона була застосована під час торгівлі криптовалютою біткоїн в парі з доларами США (BTC-USD) на біржі Bitfinex у період з 01.07.2018 р. по 01.04.2019 р. за 12 стратегіями, які вирізнялись використовуваними індикаторами, їх комбінаціями, індексами, сигналами з бірж. Використання запропонованої методики дозволило після восьми місяців застосування отримати майже 43 центи прибутку на 1 доллар, вкладений у біткоїни (табл. 7).

Таблиця 7 – Аналіз доходності портфеля криптовалют

Рік	Місяць	Доходність за інвестиційним портфелем, %	Приріст капіталу на 1 доллар, вкладений у біткоїн, долл
2018	Серпень	-7,47	-0,07
	Вересень	17,56	0,09
	Жовтень	-0,56	0,08
	Листопад	-3,06	0,05
	Грудень	11,18	0,17
2019	Січень	14,01	0,40

Лютий	5,67	0,41
Березень	1,50	0,43

Програмна реалізація запропонованої системи підтримки прийняття рішень виконана в середовищі SAS Enterprise Guide 7.1 із використанням мови програмування SAS Base та спеціалізованої мови оптимізаційного моделювання OPTMODEL, призначеної для побудови моделей та використанню відповідних алгоритмів рішення оптимізаційних задач.

**Приклад 4.** Застосування теорії подібності процесів у системі підтримки прийняття рішень для короткострокового прогнозування. Дана система може бути використана як доповнення до існуючих систем прогнозування поширення інфекційних захворювань, оскільки запропонована інформаційна технологія більше орієнтована на вирішення задач короткострокового прогнозування (зокрема поширення коронавірусної хвороби). У зв'язку з тим, що на даний час немає значної кількості накопичених даних, що описують динаміку поширення цієї хвороби, особливостей перебігу її серед різних категорій населення в регіонах, тощо, запропоновано використання такого технологічного ланцюга, який би дозволяв виявляти країни із схожим перебігом епідемії, процеси, подібні за динамікою, широтою охоплення, структурою захворівших, тощо. У розробці використано програмне забезпечення SAS Enterprise Miner 14.1 та SAS Enterprise Guide.

На момент розробки системи в медичних базах даних містилась інформація щодо захворюваності на коронавірусну хворобу у 210 країнах. Для побудови моделей з них було обрано 50 (за принципом найбільшої кількості зареєстрованих випадків захворювання).

Для виявлення схожих за ознаками (рівень смертності та індекс охорони здоров'я) країн було виконано кластерний аналіз статистичних даних. В якості вхідних даних використовувалися індекс смертності та індекс охорони здоров'я. Кластерний аналіз виконано в два етапи:

Етап 1. Із використанням методу Уорда було виконано перевірку гіпотез, щодо розбиття країн від 2 до 10 кластерів. На основі значення індексу ССС було з'ясовано, що оптимальна кількість кластерів дорівнює 6.

Етап 2. На основі методу кластерного аналізу  $k$ -середніх, із задаванням попередньої стандартизації вхідних змінних аналізу (з нульовим математичним сподіванням та одиничною дисперсією).

На основі кластерного аналізу було визначено країни-аналоги за рівнем смертності та станом медичної галузі.

Для подальшого дослідження були опрацьовані статистичні дані з Інтернет щодо захворюваності на коронавірусну хворобу у країнах, які входять до аналізованого кластеру та обчислено кумулятивні статистики із використанням розробленого автором програмного забезпечення. Підготовка даних для аналізу відбувалась в три етапи, а саме:

Етап 1. Введено змінну week (номер тижня). Відлік тижнів починається не від першого випадку, а від того дня коли було офіційно зареєстровано 100 випадків зараження.

Етап 2. Обчислено кумулятивне щотижневе значення підтверджених випадків зараження вірусом для кожної з країн, що увійшли до аналізованого кластеру.

Етап 3. Обчислено кумулятивне щотижневе значення підтверджених випадків смерті від вірусу для кожної з країн, що входять до аналізованого кластеру.

Порівняння часових рядів кількості захворівших на коронавірусну хворобу у досліджуваній країні і країнах-аналогах здійснюється автоматизовано за допомогою SAS STAT у наступні кроки (результати представлені у таблиці 8).

Крок 1. На основі вхідного цільового та часового ряду країни-аналога будується матриця відстаней.

Крок 2. Від елемента (1,1) до елемента (m,n) матриці відстаней будується оптимальний шлях з урахуванням обмежень на стиснення та розширення.

Крок 3. Обчислюються статистичні характеристики побудованого оптимального шляху (табл. 8).

Таблиця 8 – Статистичні характеристики результатів аналізу на ступінь подібності часових рядів

Показник	Країна				
	Білорусь	Угорщина	Нігерія	Румунія	Сербія
Процент пропусків в даних, %	2	0	5	0	0
Процент прямих співставлень шляху, %	72.00	83.33	86.96	87.50	83.33
Процент стиснень, %	12.00	8.333	4.348	8.333	8.333
Процент розширень, %	16.00	8.333	8.696	4.167	8.333
Загальний процент перетворень даних, %	28.00	16.67	13.04	12.50	16.67
Значення функціоналу вартості Cost	0.168	4,08	0.397	0.148	0.64
Міра подібності	59,88	74,21	89,64	96,79	88,16

Як можна побачити з таблиці 8, застосування міри подібності до аналізу нелінійних нестационарних процесів дозволяє відображати ступінь подібності (схожості) часових рядів, навіть у випадках, коли їх довжини відрізняються. На відміну від стандартної кореляції Пірсона, яка відображає наявність лінійного зв'язку, міра подібності відображає ще й нелінійний. Міра подібності обчислюється на основі розробленої методики, що описана у третьому розділі дисертації.

Очевидним є факт, що відсутність тестів впливає на об'єктивну статистичну картину урахування кількості хворих та померлих. Тому, в роботі пропонується використовувати підхід уточнення даних показників, розрахованих на основі даних країн-аналогів із використанням вагових коефіцієнтів, розрахованих для різних груп кластерів. Скориговані значення основних показників захворюваності в подальшому будуть використані для прогнозування поширення захворюваності серед населення в країні, що досліджується. Для прогнозування побудовано більше 700 моделей,

зокрема, із застосуванням методів експоненційного згладжування (просте, подвійне, з лінійним трендом, з демпфуючим трендом) для прогнозування кількості хворих (померлих) в країні (табл. 9).

Таблиця 9 – Порівняння значення статистичної характеристики МАРЕ результатів прогнозування, отримане за різних методів експоненційного згладжування

Назва країни	Кількість етапів однокрокового прогнозування (для порівняння)	Експоненційне згладжування			
		просте	подвійне	з лінійним трендом	з демпфуючим трендом
Україна	17	11,78	1,95	1,87	1,9
Румунія	18	9,96	1,45	1,45	1,47
Білорусь	16	3,37	0,65	0,65	0,64
Угорщина	18	8,83	3,64	3,64	3,68
Нігерія	18	9,79	1,49	1,61	1,59
Сербія	18	5,91	1,48	1,48	1,48
Казахстан	17	13,72	7,09	7,25	7,04
Італія	21	1,92	0,55	0,55	0,54
США	20	8,38	0,93	0,93	0,93

Як можна побачити з табл. 9, в більшості випадків кращі результати прогнозування отримано за використання методу подвійного експоненційного згладжування. Застосування теорії подібності процесів для вирішення задачі прогнозування поширення коронавірусної хвороби дало можливість не лише отримати результати високої якості, а й опрацювати інформаційну невизначеність. Оскільки, як було підтверджено в ході дослідження, недостатнє охоплення населення тестуванням впливає на об'єктивну статистичну картину ураження кількості хворих та померлих внаслідок коронавірусної хвороби, тому, для уникнення спотворення вхідних даних, в роботі використано підхід уточнення значень показників із урахуванням значень аналогів-країн.

У додатках подані матеріали, що ілюструють практичні реалізації отриманих рішень, а також акти впровадження результатів дисертаційного дослідження.

## ВИСНОВКИ

На основі виконаних теоретичних та експериментальних досліджень вирішено важливу науково-прикладну проблему в галузі інформаційних технологій – підвищення ефективності прийняття рішень в управлінні розвитком нелінійних нестационарних процесів різного призначення методами та засобами сучасних інформаційних технологій. При цьому отримано такі нові результати:

1. Виконано аналіз сучасних методів моделювання та прогнозування розвитку процесів різної природи в умовах наявності нелінійності, нестационарності та

невизначеності.

2. Розроблено системну методологію побудови математичних моделей для опису та прогнозування процесів, що характеризують розвиток нелінійних нестационарних процесів в умовах невизначеності.

3. Удосконалено методи прогнозування нелінійних нестационарних процесів різної природи на основі використання ймовірнісних та комбінованих прогнозів, що дозволяє зменшити похибки при прогнозуванні на 5-30%.

4. Запропоновано метод оцінювання параметрів математичних моделей та їх ансамблів, що дозволило подолати зміщеність оцінок прогнозів та підвищити на 5-20% статистичні показники адекватності моделей.

5. Розроблено метод розкриття невизначеностей різних типів за допомогою ймовірнісно-статистичних методів з метою коректного розв'язання задач прогнозування розвитку обраних процесів.

6. Побудовано математичні моделі та ансамблі моделей обраних нелінійних нестационарних процесів для прогнозування їх розвитку з метою підтримки прийняття управлінських рішень в умовах інформаційної невизначеності.

7. Розроблено та виконано апробацію запропонованих підходів до моделювання та прогнозування, запропонувати підходи, що забезпечують адекватний опис причинно-наслідкових зав'язків різних груп чинників та визначення можливих варіантів розвитку досліджуваних нелінійних нестационарних процесів, що дозволило підвищити точність прогнозування на 10-20%.

8. Виконано аналіз адекватності побудованих моделей та формування висновків на їх основі щодо можливості їх застосування для розв'язання задач оцінювання і прогнозування розвитку нелінійних нестационарних процесів, зокрема, соціально-економічних та еколого-економічних.

9. На основі системного підходу побудовано та реалізовано інформаційну технологію для розв'язання задач моделювання та прогнозування нелінійних нестационарних процесів для її подальшого використання у системах підтримки прийняття рішень.

## **СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ**

Наукові праці, в яких опубліковані основні наукові результати дисертації

### ***Монографії***

1. Bidyuk P. Modelling Nonlinear Nonstationary Processes in Macroeconomy and Finances/ Bidyuk P., Prosyankina-Zharova T., Terentiev O. / *Advances in Computer Science for Engineering and Education. ICCSEE 2018. Advances in Intelligent Systems and Computing* / Ed. Hu Z., Petoukhov S., Dychka I., He M. (eds). Vol. 754. Cham: Springer, 2019. P.735-745. [http://doi.org/10.1007/978-3-319-91008-6\\_72](http://doi.org/10.1007/978-3-319-91008-6_72)

2. Trofymchuk O. M, Bidiuk P. I., Prosiankina-Zharova T. I., Terentiev O. M. Decision support systems for modelling, forecasting and risk estimation: monography. Riga: LAP LAMBERT Academic Publishing. 2019. 176 p.

**Статті у виданнях, індексованих у наукометричних базах**

3. Terentyev A. N., Bidiuk P. I., Mironova A. V., Medin N. Y. Comparison of data mining methods while credit rating of natural persons. *Journal of Automation and Information Sciences*, 2009. № 41(10). P. 71–80. (Scopus)

4. Bidiuk P. I., Davidenko V. I., Trofimenko D. V., Terentyev A. N. Comparative analysis of estimation methods of vertices correlation while Bayesian networks construction. *Journal of Automation and Information Sciences*, 2010. №42(11). P. 36–45. (Scopus)

5. Prosyankina-Zharova T. , Terentiev O. , Bidiuk P. , Makukha M. Features of SAS Enterprise Guide for probabilistic modeling system, macroeconomic analysis and forecasting. *Journal of Mathematics and System Science*. 2016. №6. P. 112-122. (Index of Copernicus)

6. Bidiuk P. I., Prosiankina-Zharova T. I., Terentiev O. M., Lakhno V. A., Zhmud O. V. Intellectual technologies and decision support systems for the control of the economic and financial processes. *Journal of Theoretical and Applied Information Technology*. 2019. Vol. 96. No. 1. P. 71-87. (Scopus Q3)

7. Terentiev O. M., Prosiankina-Zharova T. I., Lakhno V. A., Usatiuk Y. V. The features of the predictive computing modeling power system load in terms of reforming energy market. *Journal of Theoretical and Applied Information Technology*. 2020. Vol. 98. No. 2. P. 163-182. (Scopus Q3)

**Статті у наукових фахових виданнях України**

8. Терентьев А. Н., Бидюк П. И., Коршевич Л. А. Алгоритм вероятностного вывода в байесовских сетях. *Системні дослідження та інформаційні технології*. 2009. № 2. С. 107-111.

9. Бідюк П. І., Терентьев О. М., Коновалюк М. М. Байєсівські мережі в технологіях інтелектуального аналізу даних. *Наукові праці [Чорноморського державного університету імені Петра Могили комплексу "Києво-Могилянська академія"]*. Серія: *Комп'ютерні технології*. 2010. Т. 134. Вип. 121. С. 6-16.

10. Кашкаръов А. О., Терентьев О. М. Анализ випадковых процессов за допомогою швидкого перетворення Фур'є. *Праці Таврійського державного агротехнологічного університету*, 2010. Вип. 10, Том 10. С. 158-162.

11. Свиридюк Я. В., Дегтярьов А. О. Терентьев О. М. Застосування нечіткого-множинного методу для оцінювання ризику банкрутства корпорації. *Вісник Університету «Україна» (серія: «Інформатика, обчислювальна техніка та кібернетика»)*. 2010. № 8. С. 105-108.

12. Терентьев О. М., Бідюк П. І., Кузнецова Н. В. Система підтримки прийняття рішень для аналізу фінансових даних. *Наукові вісті КПП*. 2011. №1. С. 48-61.

13. Бідюк П. І., Коршевич Л. О., Терентьев О. М., Просянкін-Жарова Т. І. Аналіз інвестиційних і соціально-економічних процесів методами моделювання обмежених множин багатовимірних даних. *Наукові вісті КПП*. 2012. №2. С. 87–93.

14. Терентьев О. М., Кириченко В. Е., Связінська Н. О., Просянкін-Жарова Т. І. Прогнозування фінансових ризиків з використанням наївного і доповненого деревом

класифікаторів на основі байєсівських мереж. *Наукові вісті НТУУ КПІ*. 2016. №2. С. 60-68. <https://doi.org/10.20535/1810-0546.2016.2.63882>

15. Терентьев О. М., Просьянкіна-Жарова Т. І., Савастьянов В. В. Використання засобів текстової аналітики як інструменту оптимізації підтримки прийняття рішень у задачах розробки планів соціально-економічного розвитку України. *Реєстрація, зберігання і обробка даних*. 2016. Т. 18. № 3. С. 75-86.

16. Терентьев О. М., Просьянкіна-Жарова Т. І., Савастьянов В. В. Застосування когнітивного та ймовірнісного моделювання у задачах формування сценаріїв розвитку соціально-економічних систем. *Наукові вісті НТУУ КПІ*. 2016. №5. С. 37-47. <https://doi.org/10.20535/1810-0546.2016.5.79876>

17. Бідюк П. І., Терентьев О. М., Просьянкіна-Жарова Т. І., Ефендієв В. В. Прогнозне моделювання нелінійних нестационарних процесів у рослинництві з використанням інструментів SAS Enterprise Miner. *Наукові вісті НТУУ КПІ*. 2017. №1. С. 24-36. <https://doi.org/10.20535/1810-0546.2017.1.87423>

18. Bidyuk P., Overmyer S., Prosyankina-Zharova T., Terentiev O. Methodology of Modeling and Forecasting Nonlinear Processes in Finances. *Наукові вісті НТУУ КПІ*. 2018. №1. С. 15 – 25. <https://doi.org/10.20535/1810-0546.2018.1.120361>

19. Bidyuk P., Prosyankina-Zharova T., Terentiev O., Medvedeva M. Adaptive modelling for forecasting economic and financial risks under uncertainty in terms of the economic crisis and social threats. *Технологічний аудит та резерви виробництва*. 2018. №4. С. 24-36. <https://doi.org/10.15587/2312-8372.2018.135483>

20. Bidyuk P. I., Korshevnyuk L. O., Gozhyi O. P., Kalinina I. O., Prosyankina-Zharova T. I., Terentiev O. M. Modeling and forecasting financial and economic processes with decision support system. *Наукові вісті КПІ*. 2019. № 5-6. С. 7-17. <https://doi.org/10.20535/kpi-sn.2019.5-6.176835>

#### ***Наукові праці, які засвідчують апробацію матеріалів дисертації***

21. Бідюк П.І., Коршевніук Л. О., Терентьев О. М. Підтримка розв'язання слабкоструктурованих задач в органах державної влади. *Системний аналіз і інформаційні технології (САІТ-2012)*. матеріали 14 міжнар. наук.-техн. конф. (Київ, 24 квітня 2012 р.) Київ: ННК “ІПСА” НТУУ “КПІ” 2012. С. 169-170.

22. Коршевніук Л.О., Терентьев О. М., Бідюк П.І. Методика побудови математичних моделей динамічних процесів. *Системний аналіз та інформаційні технології (САІТ 2013)*: матеріали 15-ї міжнар. наук.-техн. конф. (Київ, 27–31 травня 2013 р.). Київ: ННК “ІПСА” НТУУ “КПІ”. 2013. С. 288-289.

23. Prosyankina-Zharova T. I., Terentiev O. M., Bidyuk P. I., Gasanov A. Features of SAS Enterprise Guide for probabilistic modeling system, macroeconomic analysis and forecasting: *On control and optimization with industrial applications (COIA-2015)*, Proceedings of the fifth international conference (Baku, 27-29 August 2015), Baku: Institute of applied mathematics BSU, 2015. P. 365-368.

24. Терентьев О.М., Связінська Н.О., Просьянкіна-Жарова Т.І. Візуалізація типології регіонів України із використанням система SAS Enterprise Guide 7.1. *Геоінформаційні системи і комп'ютерні технології еколого-економічного моніторингу*: матеріали міжнар. наук.-техн. конф. (Дніпропетровськ, 13–15 квітня 2016 р. Днепропетровск: ГВУЗ «НГУ» МОН Украины, 2016. С. 21-25.

25. Терентьев О. М., Просьянкина-Жарова Т. І., Бідюк П. І., Связинська Н. О., Кириченко В. Е. Text mining analysis of agriculture internet sources using SAS software. *Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту (ISDMCI 2016)*: матеріали міжнар. конф.(м. Залізний порт, 24-28 трав. 2016 р.). Херсон: ХНТУ, 2016. С. 244-246.

26. Ivanova Y. V., Terentiev O. N., Korshevnyuk L. O., Prosyankina-Zharova T. I. Using modified logistic regression to increase customer response rate to marketing campaigns. *International conference on System Analysis and Information Technology (SAIT 2016)*. Proceedings of the 18-th conference (Kyiv, May 30 – June 2). Kyiv: Institute for Applied System Analysis, National Technical University of Ukraine “KPI”, 2016. С. 304-305.

27. Бідюк П. І., Терентьев О. М., Просьянкина-Жарова Т. І. Методи заповнення пропусків даних у задачах прогнозного моделювання соціально-економічних процесів. *Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту (ISDMCI 2016)*: матеріали міжнар. конф. (м. Залізний порт, 22-26 травня 2017 р.). Херсон: ПП Вишемирський В. С., 2017, С. 185-187.

28. Бідюк П. І. Терентьев О. М., Просьянкина-Жарова Т. І., Савастьянов В. В. Застосування інструментів SAS BASE для дослідження ефективності методів обробки пропусків у вибірках даних з метою підвищення якості прогнозування показників продовольчої безпеки країни. *International conference on System Analysis and Information Technology (SAIT 2017)*. Proceedings of the 19-th. conference (Kyiv, May 22 – 25. 2017).. Kyiv: National Technical University of Ukraine “KPI”, 2017. С. 253-255

29. Трофимчук О.М., Бідюк П. І., Просьянкина-Жарова Т. І., Терентьев О. М. Адаптивне моделювання і прогнозування у системах підтримки прийняття рішень сільськогосподарського призначення. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: колективна монографія / ред. С.О. Довгий*. К.: ТОВ “Видавництво Юстон”, 2017. С. 21-28.

30. Trofymchuk O. M. Forecasting nonstationary processes in demography, ecology, economy and finances/O. M. Trofymchuk, P. I. Bidyuk, T. I. Prosyankina-Zharova, O. M. Terentiev *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях: колективна монографія/ ред. С.О. Довгий*. К.: ТОВ “Видавництво Юстон”, 2018. С. 113-123.

31. Bidyuk P., Huskova V., Terentiev O.: Client solvency estimation using intellectual data analysis approach, *Advanced Information Systems and Technologies (AIST 2018)*, Proceedings of the 6th International Conference (Sumy, 16-18 May 2018), Sumy: Sumy State University, 2018 . С. 12-16.

32. Bidyuk, P. Terentiev O., Prosyankina-Zharova T. Dynamic processes forecasting and risk estimation under uncertainty using decision support systems. IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Proceedings of the Proceedings of the conference (Kyiv, 29 May-2 June 2017), Kyiv: Igor Sikorsky Kyiv Polytechnic Institute. 2017. P. 795-800. (Scopus) <https://doi.10.1109/UKRCON.2017.8100355>

33. Шолохов О. В., Терентьев О. М., Просьянкина-Жарова Т. І. Підвищення ефективності соціальних комунікацій на основі аналізу інтернет-джерел засобами text mining. *Прикладні системи та технології в інформаційному суспільстві:*



*матеріали* міжнар. наук.-практ. конф. (Київ, 30 вер. 2020 р.). К.: Київський нац. ун-т імені Тараса Шевченка, 2020. С. 23-44.

34. Trofymchuk O.M., Bidyuk P.I., Prosyankina-Zharova T.I., Terentiev O.M. Decision support system for modeling and forecasting ecological processes. *Сучасні інформаційні технології управління екологічною безпекою, природокористуванням, заходами в надзвичайних ситуаціях*: колективна монографія / ред. С.О. Довгий. К.: ТОВ «Видавництво Юстон», 2020. С. 23-44.

35. Korbicz J., Bidyuk P., Kuznietsova N., Terentiev O., Prosyankina-Zharova T. Multivariate distribution model for financial risks management. *The 9th International Conference "Information Control Systems & Technologies" (ICST2020)* (Odessa, Sept. 24–26, 2020). CEUR Workshop Proceedings, 2020. P. 416-429. (Scopus)

#### *Навчальні посібники*

36. Згуровський М. З., Бідюк П. І., Терент'єв О.М., Присянкін-Жарова Т. І. Байєсівські мережі в системах підтримки прийняття рішень: навч. посіб. К: ТОВ «Видавниче підприємство «Едельвейс», 2015. 300 с.

37. Домрачев В.Н., Костецкий Р.И. Терент'єв А.Н. SAS BASE: Основы программирования: наук. вид. К: Эдельвейс, 2014. 304 с.

38. Бідюк П. І., Терент'єв О. М., Присянкін-Жарова Т. І. Прикладна статистика: навч. пос. Вінниця : ПП «ГД Едельвейс і К», 2013. 288 с.

#### *Авторські свідоцтва*

39. Трофименко Д.В., Давиденко В. І. Комп'ютерна програма “Bayesian Network Master BNetMaster” Авторське свідоцтво № 34443 України. Заявл. 09.06.2010; № 34657. Опубл. 09.08.2010.

40. Д. А., Дегтярьов А. О. Комп'ютерна програма “Alpha” Авторське свідоцтво № 34191 України. Заявл. 21.05.2010; № 34419. Опубл. 21.07.2010.

41. Бідюк П.І., Кириченко В.Е., Связінська Н.О. Комп'ютерна програма “IMLBayesNet” Авторське свідоцтво № 64117 України. Заявл. 17.12.2015; № 64562. Опубл. 16.02.2016.

42. Щука Р.В., Іванов С.С., Терент'єв О.М., Бідюк П. І., Макуха М.П. Комп'ютерна програма “SAS Similar Trajectories” № 72358. Заявл. 07.03.2017. Опубл. 05.05.2017.

## АНОТАЦІЯ

**Терент'єв О. М. Моделі, методи та інформаційні технології прогнозування нелінійних нестационарних процесів в умовах невизначеності. – Кваліфікаційна праця на правах рукопису.**

Дисертація на здобуття наукового ступеня доктора технічних наук за фахом 05.13.06 «Інформаційні технології». – Інститут телекомунікацій і глобального інформаційного простору, Національна академія наук України, Київ, 2021.

Дисертаційну роботу присвячено вирішенню актуальної науково-прикладної проблеми розроблення і використання моделей та методів прогнозування нелінійних нестационарних процесів різної природи, призначених до використання у сучасних інформаційних системах підтримки прийняття рішень

У дисертації розроблено нову інформаційну технологію для реалізації та використання прогнозуючих моделей високого ступеня адекватності та якості, що

дозволяє прогнозувати нелінійні нестационарні процеси різних типів. Розроблена інформаційна технологія ґрунтується на принципах багатомодельного підходу, інтеграції структурованої та неструктурованої інформації, системному використанні методів інтелектуального аналізу даних, моделювання, прогнозування та прийняття рішень. Науковими результатами досліджень є інформаційна технологія для розв'язання задач прогнозування нелінійних нестационарних процесів, удосконалені методи прогнозування нелінійних нестационарних процесів різної природи на основі використання ймовірнісних та комбінованих прогнозів, метод оцінювання параметрів математичних моделей та їх ансамблів, метод розкриття невизначеностей різних типів за допомогою ймовірнісно-статистичних методів з метою коректного розв'язання задач прогнозування розвитку обраних процесів, підходи, що забезпечують адекватний опис причинно-наслідкових зав'язків різних груп чинників та визначення можливих варіантів розвитку досліджуваних нелінійних нестационарних процесів, побудовано та реалізовано інформаційну технологію для розв'язання задач моделювання та прогнозування нелінійних нестационарних процесів для її подальшого використання у системах підтримки прийняття рішень.

**Ключові слова:** нелінійні нестационарні процеси, інформаційна технологія, прогнозуючі моделі, часові ряди, методи подібності, інтелектуальний аналіз даних, багатомодельний підхід.

## ABSTRACT

**Terentiev O. M. Models, methods and information technologies of forecasting nonlinear nonstationary processes in conditions of uncertainty. – Qualification scientific work on the rights of the manuscript.**

The thesis for the degree of Doctor of Technical Sciences on specialty 05.13.06 «Information technologies». – Institute of Telecommunications and Global Information Sphere, National Academy of Sciences of Ukraine, Kyiv, 2021.

The thesis focuses on solving urgent scientific and applied problems related to development and usage of models and methods for forecasting nonlinear nonstationary processes of various origin to be further used in modern information decision support systems.

In the frames of the thesis new information technology was developed for implementation and application of high adequacy and quality forecasting mathematical models that allow for forecasting nonlinear nonstationary processes of various types and origin. The information technology developed is based upon the principles of multi-model approach, integration of structured and non-structured information, systemic applications of the methods for intellectual data analysis, modeling, forecasting and decision making. The principal scientific results include the following ones: information technology for solving the problem of modeling and forecasting nonlinear nonstationary processes; improved methods of forecasting the processes mentioned and of various origin on the basis of usage the probabilistic, statistical and combined forecasting techniques; the method for model parameters estimation as well as for their ensembles; the method for uncertainty identification and taking into consideration the uncertainty by making use of probabilistic and statistical approaches with the final purpose of correct solving the problem of

forecasting selected processes; the approaches that provide for adequate formal description of causal relationships between various groups of factors and determining possible future alternatives for development of nonlinear nonstationary processes in various areas of human activity; the information technology was developed and practically implemented for solving the problem of forecasting nonlinear nonstationary processes, and its further use in decision support systems. Multiple application examples based on statistical data collected are given illustrating practical usage of the methods, models and systemic methodology developed.

**Keywords:** nonlinear nonstationary processes, information technology, forecasting models, time series, similarity analysis methods, intellectual data analysis, multi-model approach.



Підписано до друку 25.03.2021 р. Формат 60×901 /16.  
Ум.-друк. арк. 1,9. Тираж 150 прим. Зам. № \_.

Видавець і виготовлювач

вул. \_\_, 11, м. \_\_, \_\_\_\_.

Свідоцтво про внесення суб'єкта видавничої справи до державного реєстру видавців,  
виготівників і розповсюджувачів  
видавничої продукції РВ № \_\_ від \_\_. \_\_. \_\_\_\_ р.