

**Міністерство освіти і науки України**  
**Національний університет «Чернігівський колегіум» ім. Т. Г. Шевченка**  
**Природничо-математичний факультет**  
**Кафедра математики та економіки**

## **Кваліфікаційна робота**

Освітній ступень – магістр

на тему: «**Методи прогнозування та їх застосування**»

Виконала студентка 2-го року навчання,  
Напрямку підготовки  
111 Математика

Сергієнко Яна Михайлівна

Керівник: доцент Балюнов О. О.

Кваліфікаційна робота захищена  
з оцінкою « \_\_\_\_\_ »

« \_\_\_\_ » \_\_\_\_\_ 2021 р.

## Зміст

Вступ	2
Розділ 1. Теоретичні основи прогнозування.	
1.1. Історія прогнозування	5
1.2. Етапи прогнозування	5
1.3. Підходи до прогнозування	13
Висновки до розділу 1	24
РОЗДІЛ 2. Вибір та опис методів статистично-ймовірнісного прогнозування	
2.1. Мережа Байєса	25
2.2 Прогнозування за допомогою нейронних мереж	27
2.3. Метод експоненційного згладжування	30
2.3.1. Методи вибору $\alpha$	32
2.3.2. Вибір початкового наближення	33
2.4. Метод групового врахування аргументів	33
2.4.1. Основні ідеї методу	33
2.4.2. Зовнішні критерії оптимальності	34
2.4.3. Опис алгоритмів селекції	35
Висновки до розділу 2	37
Розділ 3. Порівняльний аналіз результатів моделювання, отриманих нейромережними та економетричними методами	39
Висновки до 3 розділу	50
Висновки по виконаній роботі	52
Список використаних джерел	54

## Вступ

Прогнозування відіграє важливу роль у різних областях господарства. Оскільки умови технологічних процесів і умови ведення бізнесу швидко змінюються в часі, необхідно постійно прогнозувати ці зміни для успішної реалізації технічних розв'язків або здійснення ділових операцій.

*Актуальність теми* обумовлена наступними обставинами. В наш час розроблено багато методів прогнозування, кінцевим завданням яких є пророкування майбутніх подій з тим чи іншим ступенем надійності з метою використання цього прогнозу при прийнятті рішення.

Формально є два підходи до прогнозування - якісний і кількісний. Методи якісного прогнозування, такі як метод експертного оцінювання, особливо важливі, коли статистичні дані за минулі періоди часу недоступні або ненадійні. Усі якісні методи вкрай суб'єктивні й піддані високій помилці прогнозу.

Кількісні методи прогнозування засновані на істотному використанні інформації за минулі періоди часу. При дослідженні тенденції процесу за минулий час вдається з'ясувати основні взаємозв'язки між величинами й дати більш надійний прогноз на майбутнє.

Методи прогнозування, засновані на аналізі часових рядів, відносяться до класу кількісних методів прогнозування. Прогнозування часових рядів передбачає визначення прогнозного значення змінної переважно на основі минулих і поточних значень цієї ж змінної. Якщо визначаються значимі фактори й функціональна або стохастична залежність відгуку від цих факторів із застосуванням множинного регресійного аналізу, то, як правило, говорять про аналіз багатомірних часових рядів.

У зв'язку із широким поширенням програмних засобів прийняття рішень особливу важливість представляє розробка методів автоматизованого прогнозування для використання результатів прогнозу в автоматизованих системах прийняття рішень.

Автоматизоване прогнозування з використанням обчислювальної техніки припускає функціонування прогнозуючих моделей і алгоритмів з мінімальною

участю людини, автоматичний вибір моделі й параметрів моделі для прогнозування конкретних показників, що чисельно виражаються за допомогою часового ряду. На сьогоднішній день існує багато моделей прогнозування часових рядів: регресійні й авторегресивні моделі, нейромережеві моделі, моделі експонентного згладжування, моделі на базі ланцюгів Маркова, класифікаційні моделі тощо. Найбільш популярними й широко використовуваними є класи авторегресійних і нейромережевих моделей. Істотним недоліком авторегресійного класу є велика кількість вільних параметрів, ідентифікація яких неоднозначна й ресурсоємка.

Одним із сучасних методів, використовуваних для прогнозування, є штучні нейронні мережі. Теорія штучних нейронних мереж була створена й розвинена такими вченими, як Ф. Розенблат, М. Мінський, С. Гросберг, Т. Кохонен. Сьогодні питанням теорії й практичного використання штучних нейронних мереж присвячені роботи таких учених як А.І. Галушкин, А.Н. Горбань, В.В. Золотарьов, С.Г. Короткий, А.А. Єжов, В.М. Неделько тощо.

*Об'єкт дослідження:* процеси використання методів прогнозування.

*Предмет дослідження:* методи прогнозування

*Мета роботи:* дослідити особливості методів прогнозування.

*Завдання дослідження:*

1. Вивчити теоретичні основи прогнозування;
2. Розглянути методи статистично-ймовірнісного прогнозування
3. Виконати порівняльний аналіз результатів моделювання, отриманих нейромережними та економетричними методами

*Практична цінність:* запропоновані у роботі методи прогнозування можуть бути використані на практиці, як дієвий спосіб покращення прийняття рішень при діяльності підприємства.

*Структура роботи:* магістерська робота складається із вступу, трьох розділів, висновків та списку використаних джерел.

Список використаних джерел містить 41 найменування.

## **Розділ 1. Теоретичні основи прогнозування.**

### **1.1. Історія прогнозування**

Сучасні підходи до прогнозування розробили ще в дев'ятнадцятому столітті, яскравим підтвердженням цього може бути метод регресійного аналізу [1]. Однак, серед практик, що часто використовуються в прогнозуванні є, поряд з тими, чий корінь сягають століття тому, і такі, як техніки Бокса-Дженкінса, нейронні мережі, що використовуються для цілей знаходження прогнозних значень.

З розвитком, а також зростанням рівня складності апаратів прогнозування, появою і широким поширенням комп'ютерів, які були оснащені спеціальними програмами для отримання прогнозів, прогнозуванню стали приділяти набагато більше уваги, як до процесу здатного зменшувати ступінь невизначеності при прийнятті рішень.

Техніка для цілей прогнозування продовжує свій розвиток і все нові методи і техніки прогнозування розробляються для різних цілей в умовах сучасної доступності інформації для отримання більш точних значень прогнозу [2]. Оскільки на сьогоднішній день не існує методу прогнозування, який повністю був очищений від можливості виникнення помилок, то розвиток даної галузі зупинився, однак оскільки такий метод ще не розроблений, то людина, яка займається прогнозуванням, може лише пом'якшити, якщо це є можливим, наслідки від виникнення помилок у прогнозі [3 с.43].

### **1.2. Етапи прогнозування**

Співвідношення використання кількісних і якісних методів у прогнозуванні значно змінилося останніми роками, у зв'язку з поширенням комп'ютерів, до виникнення серйозного математичного апарату прогнозування оцінка, часто інтуїтивного характеру, була єдиним можливим варіантом отримати будь-які припущення про характер спостережуваних подій у майбутньому. Завдяки працям, пов'язаним з кількісними методами прогнозування Макрідакіса, стало зрозуміло, що використання тих прогнозів, що ґрунтуються лише на якісних оцінках, не можуть бути настільки ж точними, як ті, що мають математичне обґрунтування отриманих значень [4 с.382]. Також у довгостроковій перспективі використання якісних методів прогнозування виявляється суттєво дорожчим, ніж

використання програмного забезпечення, що ґрунтується на кількісних підходах [5].

Тим не менш, точність у прогнозуванні не може бути досягнута лише шляхом використання програм відповідного призначення [6]. У такій ситуації ми отримаємо лише інформацію, яка без належного знання і розуміння, може виявитися абсолютно марною.

Програмне забезпечення та комп'ютери стали абсолютно невід'ємною частиною будь-якої організації незалежно від її розмірів чи спрямованості, а також приналежності до приватного бізнесу чи державного, тому що кожна компанія потребує процедури планування [7]. Причому важливо розуміти, що прогнози використовуються компанією практично у всіх стратегічно важливих її відділах, таких як фінансовий, маркетинг, рекрутинг, логістика та багатьох інших для прийняття обґрунтованих рішень в умовах невизначеності.

Існують різні класифікації за типами прогнозів, які можуть розглядати їх залежно від тимчасової перспективи прогнозування, так і за їхньою позицією, яку вони займають у макро- або мікро-просторі, також як уже згадувалося, може бути зроблено поділ шляхом віднесення прогнозів до кількісних або до якісних типів прогнозів, від того які цілі стоять будуть використані різні підходи до прогнозування [8].

На початковому етапі вибору методу потрібно визначити ступінь деталізації, тобто зрозуміти який з макро- чи мікро-прогнозів нам потрібен, наступним кроком буде визначення тимчасової перспективи, на яку має бути розрахований прогноз і заключним етапом стане детермінація фактора використання кількісних або якісних оцінок, або визначення їх співвідношення після ухвалення кінцевого рішення [9]. Можливість аналізу методів прогнозу при виборі одного з них для цілей прогнозування має дозволити спростити процедуру прийняття рішення [10, с.11]. Основним вимогою буде неможлива складність математичного апарату, а точний і зрозумілий результат прогнозу, який можна інтерпретувати для ухвалення рішень.[11, с.57].

Процес прогнозування є екстраполяцією попередніх спостережень для отримання деякого уявлення про те, як ситуація виглядатиме в майбутньому.

Розуміння цього факту призводить до того, що апарат прогнозування аналізує і піддає обробці ті дані, які відбулися і не відрізняються від умов майбутнього, проте це не завжди вірно, оскільки розрахунок показників ефективності заснованих лише на оцінках буде не точним, оскільки можлива зміна даних показників, тому процес прогнозування може бути розбитий на кілька складових:

- Збір даних
- Редукція або ущільнення даних
- Побудова моделі та її оцінка
- Екстраполяція обраної моделі
- Оцінка отриманого прогнозу

Перший етап передбачає отримання вірних даних, а також проведення необхідної перевірки достовірності. [4 с.382]. Даний етап є найбільш важливим для всієї наступної процедури прогнозування, оскільки без належної перевірки коректності даних, ми можемо провести прогноз для невірних даних, отримавши в результаті результат несумісний з реальними показниками.

Другий етап може бути необов'язковим, проте дуже часто без його наявності коректно здійснити прогнозування неможливо. Можливі ситуації, коли стає зрозуміло, що вихідні дані отримані щодо прогнозу зібрані у недостатній чи надмірній кількості [12, с.359]. І той і інший варіант не дозволяють побудувати точний прогноз, якщо у разі недостатньої кількості даних це представляється очевидним, то в другому випадку можливо зібрані дані, які не відносяться до досліджуваного предмета прогнозування, тим самим лише збільшуючи кількість помилок прогнозу та знижуючи його точність. Також на цьому етапі дані проходять перевірку на відповідність контексту. [13, с.24].

Третій етап може бути описаний, як побудова моделі прогнозу та її оцінка. Суть даного етапу зводиться до вибору конкретної моделі прогнозування, яка здійснюватиме за допомогою заданого алгоритму екстраполяцію попередніх спостережень. На даному етапі слід усвідомлювати, що чим простіше представлена модель прогнозування, тим легшою вона буде для розуміння її результатів, що є найважливішим фактором при прийнятті рішення. Це означає,

що простіші моделі представляють зрозумілі результати, які активно використовуються при прийнятті рішень [10, с.194].

Четвертий етап полягає в перенесенні обраної моделі прогнозування на майбутні періоди, що передбачає отримання конкретних значень прогнозу. Також на даному етапі відбувається моделювання прогнозу для даних, що вже спостерігаються, для оцінки точності прогнозу. Однак безпосередньо аналіз отриманих відхилень у ході проведення процедури прогнозування на події, що вже відбулися, буде розглянутий у п'ятому етапі [14].

П'ятий етап є заключним. На цьому етапі проводиться оцінка отриманих прогнозних значень, оцінка помилок прогнозу. Для оцінки помилки прогнозу використовуються різні техніки, які мають на увазі складання абсолютних значень прогнозу, які в залежності від техніки оцінки може залишатися як сумою відхилень при прогнозі, так і ділитися на кількість помилок, що спостерігаються, з метою пошуку середньої помилки за прогнозом. Існують також методи, що ґрунтуються на сумі квадратів помилок.

В основному статистичні показники використовують, щоб дати більш детальне уявлення про структуру даних у генеральній сукупності. Ціль даних процедур у статистиці зводиться до можливості описати великий масив даних за допомогою деяких найважливіших, ключових значень. Переважна більшість статистик описують данні шляхом усереднення значень спостережень [15 с.459]. Найпоширеніша процедура полягає у пошуку вибіркового середнього значення  $\bar{X}$ , яке визначається за формулою складання всіх наявних спостережень та поділених на кількість спостережень:

$$\bar{X} = \frac{\sum X}{N}$$

де  $\bar{X}$ -вибіркове середнє;  $\sum X$ -сума всіх значень вибірки; N-обсяг вибірки.



Поряд із використанням вибіркового середнього, при процесі визначення тенденції даних до угруповання навколо середнього значення використовують показник середньо квадратичного відхилення:

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

У цьому рівнянні ми маємо суму квадратів різниць між спостереженнями та їх середнім значенням [16].

Середнє значення та середньоквадратичне відхилення вважаються одними з найважливіших характеристик при необхідності опису набору даних сукупності. Їхня основна перевага полягає в тому, що вони досить прості до обчислень і надають змістовні характеристики даних спостережень. Також поряд з використанням середньоквадратичного відхилення та знаходження вибіркового середнього для визначення центрального значення даних застосовується процедура пошуку медіани. Медіаною називають значення, яке ділить вибірку на дві частини, причому в одній вони будуть меншими, а в іншій більше значення медіани. Розмах використовується для того, щоб приблизно оцінити дисперсію вибірки, щоб обчислити розмах потрібно від найбільшого значення вибірки відняти найменше.

Для вивчення присутності лінійної залежності в наборі даних між двома величинами, а також визначення сили даної залежності може бути використаний коефіцієнт кореляції [17]. Даний коефіцієнт вимірюється від -1 до 1. Мінімальне значення коефіцієнта кореляції рівне -1 показує, що величини, що розглядаються, мають досконалу негативну залежність, в даному випадку збільшення однієї спричинить зменшення іншої, з іншого боку коефіцієнт рівний +1 свідчить про наявність досконалої позитивної залежності між двома величинами [18, с.198]. Якщо коефіцієнт кореляції дорівнює 0, то лінійна залежність відсутня.

Коефіцієнт кореляції можна знайти за такою формулою:

$$r_{xy} = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}}$$

При використанні коефіцієнта кореляції слід враховувати кілька критично важливих моментів. Перше на що потрібно звернути увагу, що коефіцієнт кореляції показує дані про наявність залежності між двома величинами, однак він не може бути використаний як засіб при визначенні причинно-наслідкового зв'язку між факторами. Такими чином, якщо між величинами присутня кореляція, це означає вплив одного фактора на інший, оскільки зміна двох величин може бути викликана зміною третього фактора, який не включений в аналіз коефіцієнта кореляції [19 с. 34].

По-друге, коефіцієнт кореляції спрямований визначати наявність лінійної залежності, тоді як розрахунок коефіцієнта показує, що лінійна залежність дорівнює 0 чи перебуває у межах низької кореляції, то невірно буде вважати, що між змінними немає абсолютно ніякої залежності, оскільки аналізовані величини, можуть бути схильні до нелінійної залежності [20].

Етап який являє собою збір даних, перевірку їх на достовірність, є одним з найскладніших з усіх етапів прогнозування, оскільки, якщо буде допущена помилка на даному етапі, то помітити її присутність буде вкрай складно при наступних обчисленнях, тому невірні дані на вході дають на виході невірні прогнозні значення, які мають релевантності стосовно прогнозованої величині. Точність прогнозу багато в чому визначається точністю даних, які будуть використані при побудові моделі прогнозування [21]. Для того, щоб оцінити чи будуть корисні дані при пошуку рішення для якоїсь конкретної задачі можна скористатися перевіркою за такими чотирма критеріями:

а) Точність та достовірність вихідних даних. Дотримання даного критерію передбачає використання даних, отриманих з джерел достовірної інформації, а також вимагає, щоб дані відповідали контексту об'єкта, що досліджується.

б) Значення даних. Дані відображатимуть аналізовані обставини.

в) Узгодженість даних. В даному випадку мається на увазі, що за обставин зміни даних для об'єкта, для якого вони були зібрані, повинні бути внесені відповідні коригування, які дозволять зберегти узгодженість нових даних з структурою, що історично склалася.

г) Прив'язка до часу. Даний критерій дозволяє перевірити дані на їхню хронологічну відповідність. Таким чином дані, які задовольняють даному критерію, є найкращими для проведення прогнозу, також тут необхідно зазначити, що даних може бути занадто мало, це означає, що недостатньо історичної передісторії, проте також важливо розуміти, що використання надто великої кількості даних накопичених за час може пошкодити точність прогнозу через можливу низьку релевантність у контексті прогнозованого об'єкта [22].

Основними під час проведення прогнозів можуть вважатися дві категорії даних. Перші є набором даних, які були зібрані в якийсь конкретний момент часу, це можуть бути дані за різні проміжки часу: години, тижні, роки, декади і так далі. Друга категорія даних показує дані, які були зібрані з плином часу. Перший тип даних називають крос-секційним, їхнє основне завдання полягає в тому, щоб з'ясувати взаємозв'язки всередині досліджуваної сукупності, з метою екстраполяції отриманих результатів на генеральну сукупність. Дані, які були зібрані з часом називаються часовими рядами, зазвичай для цих даних існують однакові інтервали в часі, через які збираються дані про ці об'єкти.

Кількісна модель прогнозування використовується при прогнозуванні часових рядів. Для позначення величини в конкретний момент часу, прогнозного значення та помилки прогнозу використовуються такі показники [23]:

$Y_t$  - значення часового ряду в момент  $t$ .

$\hat{Y}_t$  - значення прогнозу на момент  $t$ .

$e_t = Y_t - \hat{Y}_t$  - похибка або помилка прогнозу.

Оцінюючи помилки прогнозу використовується кілька методів, більшість яких полягає в усередненні деяких функцій помилок і фактичних значень. Для обчислення помилки прогнозу використовується різниця між фактичним значенням та значенням прогнозу, отриманим для цього моменту часу.

Найбільш поширений метод обчислення помилки прогнозу полягає у складанні абсолютних значень похибки прогнозу та ділення на кількість спостережень. Цей метод називається середнє абсолютне відхилення (Mean Absolute Deviation, MAD) [3 с .43]. Використання даної техніки оцінки помилки прогнозу можливе у разі, якщо фахівець, який здійснює вимірювання помилки

прогнозу, намагається отримати результат в тих же одиницях, в яких знаходиться вихідний ряд. Помилка середнього абсолютного відхилення вимірюється за такою формулою:

$$MAD = \frac{1}{n} \sum |Y - \hat{Y}|$$

Наступний спосіб оцінки помилки прогнозування – середньоквадратична помилка (Mean Squared Error, MSE) [24], який полягає у зведенні у квадрат кожної помилки прогнозу та подальшого підсумовування всіх помилок у квадраті, сума яких ділиться на кількість спостережень. Дана техніка в силу своєї особливості зведення помилок у квадрат є необхідною для того, щоб звернути увагу фахівця на великі похибки, допущені моделлю при прогнозуванні. Середньоквадратична помилка прогнозу обчислюється за такою формулою:

$$MSE = \frac{1}{n} \sum (Y - \hat{Y})^2$$

Однак обчислення абсолютних величин похибок не завжди є кращим, оскільки розрахунок середньої абсолютної помилки у відсотках дозволяє оцінити величину розбіжності прогнозного значення та фактичними даними у відсотковому відношенні. Середня абсолютна помилка у відсотках (Mean Absolute Percentage Error, MAPE) [25, с.96] обчислюється шляхом розрахунку абсолютної величини помилки в кожний конкретний момент і діленням на фактичне значення, що спостерігається в даний період часу, сума даних дій за всіма позиціями прогнозу підсумовується і ділиться на кількість спостережень прогнозу. Істотною перевагою даного підходу є те, що якщо вихідний ряд містить великі значення, то в результаті ми отримаємо оцінку прогнозу у відсотках, значення якого не перевищуватиме трьох знаків. Розрахунок цього показника оцінки помилки прогнозу, середньої абсолютної процентної помилки проводиться за формулою:

$$MAPE = \frac{1}{n} \sum \frac{|Y - \hat{Y}|}{Y}$$

Метод середньої процентної помилки (Mean Percentage Error, MPE) [6] допомагає визначити, чи існує зміщення в прогнозних значеннях, чи є отриманий

прогноз постійно заниженим або завищеним. Обчислення середньої відсоткової помилки відбувається шляхом виявлення помилки прогнозу в кожен момент часу, з наступним розподілом знайденої похибки на фактичне значення, що характеризує даний період, подальше підсумовування отриманих результатів попередніх дій і поділ на кількість спостережень дозволяє оцінити помилку прогнозу методів середньої процентної помилки. Для аналізу даного показника потрібно розуміти, що отримане занадто велике позитивне значення у відсотках означає, що метод є постійно недооцінюючим, тобто прогнозні значення менше фактичних. Якщо значення прогнозу є великим негативним, то це означає, що метод прогнозування, для якого проводиться оцінка помилок прогнозу послідовно переоцінює. Формула, яка описує процес знаходження середньої процентної помилки:

$$MPE = \frac{1}{n} \sum \frac{(Y - \hat{Y})}{Y}$$

Вибір того чи іншого методу прогнозування часто ґрунтується на оцінці помилок прогнозу, які виходять при обчисленні таких показників як MAD, MSE, MAPE, MPE, правильно обраний метод даватиме найменше відхилення від фактичних значень, тобто буде найбільш точним у прогнозуванні майбутніх показників.

### 1. 3. Підходи до прогнозування

До групи простих методів прогнозування належать моделі усереднень, згладжувань, а також наївні методики. При прогнозуванні з допомогою наївних методів будується модель, яка передбачає, що дані майбутніх періодів найкраще описуються останніми подіями аналогічного характеру. Методи усереднення дозволяють виконати процедуру прогнозування з огляду на середню динаміку показників за минулі спостереження. Методи, що згладжують, дозволяють побудувати прогноз із застосуванням техніки усереднення минулих даних, з урахуванням вагових коефіцієнтів, що зменшуються.

Наївне прогнозування для конкретного періоду є фактичним значенням попереднього періоду. Такий вид прогнозу називається незмінним прогнозом, так

як поточному значенню показника дається 100% вага при розрахунку прогнозу. У такому разі прогноз розраховується за такою формулою:

$$\hat{Y}_{t+1} = Y_t - \text{проста модель}$$

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) - \text{модель з урахуванням тренда}$$

$$\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}} - \text{модель швидкості змін}$$

$$\hat{Y}_{t+1} = Y_{t-3} - \text{модель сезонної залежності}$$

Якщо потрібно постійне оновлення прогнозів, а найменування для яких потрібно здійснити цю процедуру обчислюються тисячами, то не завжди у фахівця з прогнозування є можливість визначити прогнозні значення для кожного конкретного найменування. Виходом з цієї ситуації може стати використання короткострокового планування, методикою, яка заснована на усередненні та згладжуванні даних у рядах [26, с.137]. В основі даних методів лежить припущення про те, що різного роду пікові значення носять випадковий характер. Це означає, що прогнозування даним методом передбачає побудову якоїсь гладкої кривої, продовження якої показуватиме прогнозні значення. У цій формулі показаний приклад усереднення даних, на основі яких будується прогноз:

$$\hat{Y}_{t+1} = \frac{1}{t} \sum Y_t$$

Метод, який використовує техніку простих середніх може бути корисний у випадках, коли процеси, які є генераторами тимчасових даних, показують стабільність, а оточення, в якому знаходяться дані ряди є незмінним. Ковзаючі середні на відміну від методу простих середніх враховує останні спостереження, тоді як простих середніх задіяний весь тимчасовий ряд. Суть моделі ковзаючих середніх полягає у постійному оновленні моделі прогнозування шляхом заміщення найбільш старих спостережень на нові [27]. У простому вигляді формула ковзаючого середнього обчислюється:

$$\hat{Y}_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-k+1})}{k}$$

$k$  – число членів у ковзному середньому.

Важливим аспектом застосування даного методу є те, що кожному спостереженню з  $k$  спостережень у формулі надається однакові вагові коефіцієнти. Швидкість реакції на зміни безпосередньо залежатиме від кількості аналізованих спостережень, чим їх буде більше, тим швидкість буде меншою, проте при незмінному навколишньому середовищі, в якій знаходяться досліджувані спостереження, точність прогнозу буде вищою, ніж при використанні меншої кількості спостережень [15 с.459].

Техніка прогнозування, в основі якої лежать подвійні ковзаючі середні, підходить для даних, які мають лінійний тренд. Суть даного підходу полягає у обчисленні ряду за допомогою ковзаючих середніх. Виходячи з отриманих значень ковзаючих середніх проводиться усереднення, отримане значення і буде прогнозом [28 с.195]. Спосіб розрахунку подвійного ковзаючого середнього показаний у формулах:

$$M_t = \hat{Y}_{t+1} = \frac{(Y_t + Y_{t-1} + Y_{t-2} + \dots + Y_{t-k+1})}{k}$$

Після розрахунку ковзаючого середнього кожного періоду, що у прогнозі, обчислюється вторинне ковзаюче середнє:

$$M'_t = \frac{(M_t + M_{t-1} + M_{t-2} + \dots + M_{t-k+1})}{k}$$

Далі необхідно розрахувати різницю між подвійним ковзаючим середнім і простим ковзаючим середнім:

$$a_t = 2M_t - M'_t$$

Також для побудови моделі прогнозування подвійних ковзаючих середніх необхідно розрахувати коригувальний фактор, суть якого схожа з коефіцієнтом нахилу прямої:

$$b_t = \frac{2}{k-1} (M_t - M'_t)$$

Зрештою рівняння подвійних ковзаючих середніх виглядає так:

$$\hat{Y}_{t+1} = a_t + b_t$$

У порівнянні з методами, які використовують техніки прогнозування із застосуванням ковзаючих середніх, в яких використовуються найбільш актуальна

інформація доступна на даний момент часу, то методи експоненційного згладжування використовують зважене ковзаюче усереднення всіх даних з попередніх спостережень. Дані, які найбільше підходять для експоненційного згладжування не мають чіткого тренду. Експоненційне згладжування передбачає не тільки постійне оновлення моделі, як модель ковзаючих середніх, а й надає найбільші коефіцієнти останнім спостереженням як найбільш репрезентативні [29 с.264]. Коефіцієнт ваги для останнього спостереження визначається за формулою  $\alpha$ , для попереднього за формулою  $(1-\alpha)$ , наступне за ним  $(1-\alpha)^2$ , причому  $0 < \alpha < 1$ . Формула, за якою розраховується прогноз моделі експоненційного згладжування виглядає наступним чином:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t$$

$\hat{Y}_{t+1}$  – прогнозоване значення для наступного періоду

$\alpha$  – постійна згладжування ( $0 < \alpha < 1$ )

$Y_t$  – величини в поточний період

$\hat{Y}_t$  – попередній згладжуваний прогноз цієї величини

Прогноз побудований із застосуванням експоненційного згладжування являє собою старий прогноз, який враховує помилку прогнозу з коефіцієнтом  $\alpha$  [2]. Значення постійної згладжування залежить від того яке спостереження надає найбільший вплив на значення прогнозу, якщо воно наближається до 1, це означає, що помилка останнього прогнозу істотно впливатиме. У той же час мінімальне значення свідчитиме про те, що значення нового прогнозу буде наближено до значення попереднього [30 с.1518]. Постійна згладжування буде ключовим значенням при побудові прогнозу, оскільки якщо ситуація вимагає отримання прогнозу для даних без серйозних відхилень, і щоб вони були згладжені, потрібно брати мінімальне значення  $\alpha$ . Велике значення може бути застосовано у разі, якщо потрібно отримати швидку реакцію на зміни. Для визначення значення  $\alpha$  проводиться оцінка помилки прогнозів побудованих з різними показниками постійного згладжування  $\{0,1;0,2;\dots;0,9\}$  для того показника  $\alpha$ , який показав найменший результат при оцінці помилок прогнозу і буде використаний при подальших обчисленнях. [29 с.264].



Лінійна регресія представляє пряму, що складена з урахуванням спостережень двох параметрів і на основі якої можна припустити майбутні значення однієї з величин, знаючи значення іншої величини [31]. Для того, щоб визначити місце розташування прямої на графіку, на якому позначено точками спостереження, буде використана процедура критерію найменших квадратів, оскільки дозволяє досягти кращого наближення по відношенню до всіх спостережень, нанесених на графік. Рівняння прямого наближення має вигляд:

$$\hat{Y} = b_0 + b_1 X$$

Перший параметр  $b_0$  є вільним членом рівняння, другий  $b_1$  показує тангенс кута нахилу. Метод найменших квадратів дозволяє підібрати значення коефіцієнтів такими, щоб сума всіх відхилень була мінімальною [17]. Коефіцієнти для рівняння лінійної регресії розраховуються так:

$$b_1 = \frac{\sum(X - \bar{X}) \sum(Y - \bar{Y})}{\sum(X - \bar{X})^2}$$
$$b_0 = \frac{\sum Y}{n}$$

Різниця, отримані при відніманні фактичного значення від прогнозу, називаються відхиленнями. Відхиленнями називають перпендикуляри, опущені з точок спостережень на пряму регресії. Для того, щоб оцінити ступінь відхилення вихідних точок від отриманої прямої регресії, використовується процедура оцінки стандартної помилки, яка обчислюється за формулою:

$$S_{y*x} = \sqrt{\frac{\sum(Y - \hat{Y})^2}{n - 2}}$$

Оцінка стандартної помилки показує рівень відмінності між фактичними значеннями та прогнозованими [22]. Для більших за обсягом даних, даний показник слід очікувати на рівні 67% помилок прогнозу по модулю будуть менше ніж оцінка стандартної помилки і близько 95% помилок по модулю не будуть більше ніж  $2S_{y*x}$ . Якщо отримана невелика стандартна помилка, це означає, що пряма регресії добре описує дані, оскільки всі точки спостережень знаходяться неподалік прямої.

Використання стандартної помилки прогнозу, дозволяє створити інтервальний прогноз, який, по-перше, дозволить уникнути невизначеності, яка може виникнути через те, що в точковому прогнозі спостереження не лежать на прямій регресії, а знаходяться в межах квадратного відхилення. По-друге, при побудові точкового прогнозу можливі неточності, викликані відхиленням прямої регресії для вибіркової сукупності від прямої генеральної сукупності [32].

Стандартна помилка прогнозу дає необхідну міру варіації прогнозованого значення, що дозволяє уникнути вищезазначених невизначеностей точкового прогнозу [31]. Стандартна помилка прогнозу обчислюється за такою формулою:

$$S_f = S_{y*x} * \sqrt{\frac{(X - \bar{X})^2}{\sum(X - \bar{X})^2} + 1 + \frac{1}{n}}$$

Якщо модель, яка була побудована із застосуванням лінійної регресії вірна, то інтервали для прогнозованого значення  $Y$  будуть перебувати в межах інтервалу  $Y^{\wedge} \pm tS_f$ , де в даному випадку  $t$  означає розподіл Стюдента з кількістю ступенів свободи рівному  $n-2$ . Якщо обсяг вибірки досить великий, тобто більше або дорівнює 30 спостережень, значення прогнозу будуть перебувати в інтервалі  $Y^{\wedge} \pm 2S_f$ . Також поряд із стандартною помилкою прогнозу використовується коефіцієнт детермінації, який дорівнює квадрату коефіцієнта кореляції, цей показник вимірюється в межах  $0 < r^2 < 1$ . Коефіцієнт детермінації рівний нулю показує, що зміна однієї величини ніяк не пов'язана зі змінами іншої величини. У випадку коли коефіцієнт дорівнює 1, то це означає, що побудована модель описує 100% випадків зміни  $Y$  від  $X$  [33].

Проста лінійна регресія досліджувала наявність взаємозв'язку між залежною та незалежною змінними, якщо дана залежність між величинами виявлена, то знайти значення залежної змінної не складає труднощів. Проте лінійна регресія показує залежність між двома чинниками, тоді як в умовах розширення процесу глобалізації дослідження залежності між двома змінними буває недостатньо, оскільки багато чинників залежить від цілого спектра незалежних величин. Модель багатовимірної регресії дозволяє враховувати кілька змінних, загалом багатовимірна регресія схожа своїми поняттями з лінійною регресією, проте є й

відмінності, які зумовлені тим, що в моделі використовується більше однієї змінної [17].

У випадку, якщо за допомогою моделі лінійної регресії ми отримали коефіцієнт детермінації 75%, що означає, що дана модель не описує чверть мінливості об'єкта, що досліджується, то нам необхідно знайти ще одну змінну, яка б дозволила збільшити відсоток описаної мінливості. Важливим аспектом пошуку полягає знаходження такої змінної, яка не була б тісно пов'язана з уже обраною раніше, оскільки кореляція між змінними може призвести до мультиколінеарності. Це означало б, що вибрані змінні при високому коефіцієнті кореляції між собою описують ту ж мінливість, яку описувала і одна змінна.

У багатовимірній регресії математичне очікування для залежної змінної може бути виражене функцією:

$$\mu_y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Цей вираз також використовується для багатовимірної регресії генеральної сукупності. У простій лінійній регресії спостереження позначаються парою змінних  $(X_i, Y_i)$ , у той час як у багатовимірній регресії спостереження описуються поєднанням значення залежної змінної та значенням незалежної змінної, де  $i$ -е спостереження  $j$ -ї незалежної змінної позначатиметься символом  $X_{ij}$ .

Залежна змінна, яка приймає позначення  $Y$  – це випадкова величина, яка має зв'язок із незалежними змінними, описану наступним співвідношенням:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

У цій формулі прийняті такі позначення:

$\varepsilon$  – це помилки, які відповідають значенню відхилення залежної змінної від справжнього значення. Передбачається, що помилки є незалежними і мають нормальний розподіл.

Коефіцієнти регресії  $\beta_0, \beta_1, \dots, \beta_k$ , невідомі.

Рівняння регресії буде набувати вигляду:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k$$

У даному рівнянні є коефіцієнти  $b_0, b_1, \dots, b_k$ . Значення коефіцієнта  $b_0$ , як і раніше, означає вільний член функції регресії [31]. Коефіцієнти  $b_1$  і  $b_2$  називають

чистими або приватними коефіцієнтами регресії, кожен з яких вимірюватиме середню зміну для величини  $Y$  при одноразовій зміні незалежної змінної, що кореспондує з нею.

Інтервальний прогноз для майбутнього значення  $\hat{Y}$  можна отримати шляхом знаходження інтервалу для великої вибірки і де всі  $X$  - незалежні змінні, в такому випадку інтервал для отримання стандартної помилки прогнозу можна обчислити за такою формулою:

$$(\hat{Y} - t_{\alpha/2} S_{yx}, \hat{Y} + t_{\alpha/2} S_{yx})$$

Для того, щоб рівняння регресії якнайкраще описувало взаємозв'язок між незалежними факторами та досліджуваними, необхідно для початку визначити ті незалежні фактори, які враховуватимуться при побудові моделі прогнозування. На цьому етапі потрібно включення всіх змінних, які покращуватимуть точність прогнозування. Також потрібно враховувати витрати, які будуть понесені з включенням кожної нової змінної в модель, тому потрібно знайти компроміс між двома умовами, що суперечать:

- Використання більшого набору незалежних факторів, які не описуватимуть мінливість, описану іншими вже включеними до прогнозу факторами.
- Співвідношення витрат та збільшеної точності прогнозу за рахунок включення до неї нових змінних. Оскільки збирання та обробка інформації для незалежних чинників становить витрати, вони повинні бути співвідносні з точністю прогнозу.

Після того, як зібрали усі необхідні потенційні незалежні змінні, потрібно зробити детальніший відбір і прибрати ту частину з них, яка може не відповідати контексту досліджуваної проблеми [34]. Незалежний фактор може бути виключений зі списку претендуючих на потрапляння безпосередньо в процес прогнозування через відсутність істотного ставлення до пошуку рішення для цього завдання. Якщо існує проблема такого роду, коли одна незалежна змінна описує ту ж мінливість залежної змінної, що й попередня, то це викличе мультиколінеарність, тобто наявність лінійної залежності між незалежними змінними, що призведе до погіршення точності прогнозу. Також неприпустимі

випадки потрапляння у модель тих незалежних чинників, які не мають реального взаємозв'язку з досліджуваним об'єктом. Іноді збирання та обробка великої кількості даних може бути занадто витратною, тому в даному випадку теж можлива відмова від пошуку, тих даних, які в кілька разів збільшують вартість прогнозу, але лише трохи зменшують помилки, що виникають у процесі прогнозування. Наступний крок у пошуку оптимального рівняння регресії також скорочуватиме кількість незалежних змінних, існує кілька способів, за допомогою яких можна видалити нерелевантні фактори з моделі, проте жодна з цих процедур не може гарантувати отримання найкращого рівняння, також тут варто розуміти, що кожен із способів, який скорочуватиме кількість незалежних змінних може використовувати зовсім різні підходи і тому на виході очікується отримання абсолютно різних найкращих рівнянь для кожного з методів.

Ця процедура передбачає набір всіх можливих рівнянь регресії, які включають потенційні незалежні чинники. Проводиться аналіз рівняння, яке не містить жодної незалежної змінної та досліджуються всі можливі комбінації, що містять незалежні змінні [21]. У цій процедурі можуть бути використані різні критерії, проте найголовнішим є коефіцієнт детермінації, що описує, скільки випадків описує побудована модель прогнозування. Спочатку будується рівняння регресії для вихідної залежної змінної, фактору, який буде схильний до прогнозування, і всіх можливих незалежних факторів, оскільки кожна з незалежних змінних має два результати: результат, коли вона присутня, тобто позитивний і негативний, коли вона відсутня в рівнянні. Тоді кількість можливих рівнянь буде представлена у загальному вигляді  $2^K$ , де  $K$  – кількість незалежних змінних, що беруть участь у моделі. Далі потрібно згрупувати всі рівняння відповідно до кількості параметрів, які необхідно оцінити [20].

Далі потрібно вибрати найкраще поєднання змінних, тобто те, що дасть найбільший коефіцієнт детермінації з тих рівнянь, що містять аналогічну кількість параметрів.

Четвертий етап вимагає суб'єктивної оцінки. Оскільки критерієм вибору буде необов'язково коефіцієнт детермінації, а найбільш просте рівняння, яке в той же

час не сильно поступатиметься максимальному значенню коефіцієнта детермінації з кращих рівнянь регресії, що зустрічаються в угруповання.

Процедура покрокової регресії є поступовим додаванням у модель прогнозування, тобто у рівняння регресії незалежних факторів, по одному на кожному окремому етапі. Для проведення покрокової регресії потрібно виконання наступних алгоритмів:

– На першому етапі покрокової регресії розглядатимуться усі прості моделі. Незалежний чинник, який пояснює значну частку мінливості досліджуваного залежного чинника і перша змінна, включена до рівняння регресії.

– Наступна змінна, яка буде додана до списку незалежних факторів рівняння регресії, буде та, яка додає найбільш значну частку в регресійну суму квадратів. Для визначення значущості вкладу використовується F-статистика, що показує значення, яке має бути перевищено для визнання змінної для включення.

– Після додавання до рівняння другої змінної вже включених у модель змінних результат перевіряється за допомогою F-тесту. Якщо результат тест покаже величину нижче, ніж значення для виключення, то змінна видаляється з регресії.

– Етапи 2 і 3 повторюються поки незалежні змінні, що вводяться в модель, не перестануть бути значущими для моделі, в момент, коли такі змінні вичерпані, процедура покрокової регресії вважається завершеною.

Поки не буде отримано підтвердження, що обрана модель регресії адекватно може надати дані, регресійний аналіз не може вважатися завершеним [24]. Причому отримання відомостей про модель прогнозування, про її можливість правильно оцінювати дані повинні бути перевірено до того, як ця модель буде включена до методики прийняття рішень в умовах невизначеності.

Дослідження залишків прогнозу є важливим етапом у процесі визначення репрезентативності моделі. Якщо модель має справу з даними тимчасових рядів, необхідно провести відповідну перевірку залишків на відсутність автокореляції між ними. Висновки чи рішення, основою яких стали результати прогнозування рівнянням, яке не задовольняє основні регресійні припущення може бути

абсолютно неправильними [31]. Останні спостереження в наборі даних називаються викидами, тобто такі значення, які значно відхиляються від динаміки інших даних. У процесі згладжування або припасування викиди не можуть досить просто бути визначені, проте вони мають значний вплив при прийнятті рішення про вибір відповідного рівняння регресії. Також є необхідність вивчення всіх викидів з метою визначення необхідності їхньої присутності в наборі даних або їх потрібно виключити з нього. У разі коли прийнято рішення про залишення залишків у наборі даних необхідно визначити ступінь їх впливу на рівняння регресії.

Ступінь впливу  $i$ -ї точки даних на положення рівняння може бути виміряно за допомогою важеля  $h_{ii}$ . Значення цього показника залежить від незалежних чинників і не залежить від досліджуваного об'єкта [20]. Для рівняння регресії до складу якої входить лише одна незалежна змінна ми маємо таку формулу для важеля  $h_{ii}$ :

$$h_{ii} = \frac{1}{n} + \frac{(X - \bar{X})^2}{\sum(X - \bar{X})^2}$$

Для  $k$  незалежних факторів вираз для  $i$ -го важеля виглядає складніше, однак значення має знаходитися в межах  $0 < h_{ii} < 1$ , середнє значення важеля дорівнюватиме  $\bar{h} = (k+1)/n$ .

У випадку, коли  $i$ -та точка є більшою за значенням, тобто близьким до 1, то  $Y$  буде повністю визначатися фактором  $Y_i$ , при цьому вплив інших даних буде мінімальним [33].

Визначення залишків відбувається за допомогою різниці фактичного значення досліджуваного фактора та спрогнозованого. Великі залишки означають велику неточність у прогнозах порівняно з подіями, що фактично спостерігаються.

Метод стандартизації залишків ґрунтується на припущенні, що оцінки для стандартних відхилень дорівнюють:

$$S_e = S_{y \cdot x} \cdot s \sqrt{1 - h_{ii}}$$

У цьому рівнянні  $S_{y^*x's} = \sqrt{MSE}$  – стандартна помилка оцінки [32]. У цьому випадку стандартизований залишок дорівнює:

$$\frac{e_i}{S_{Ei}} = \frac{e_i}{S_{y^*x's} \sqrt{1 - h_{ii}}}$$

Стандартизований залишок можна вважати великим, а відповідне значення залежної змінної крайнім, коли:

$$\left| \frac{e_i}{S_{Ei}} \right| > 2$$

Значення  $Y$ , які відповідають точкам спостережень з великими стандартизованими залишками, можуть істотно змінити розташування прямої регресії [35]. Таким чином, сучасне різноманіття різних технік прогнозування дозволяє знайти оптимальний метод прогнозування кожної конкретної ситуації.

### **Висновки до розділу 1**

Сьогодні основним завданням, яке можна вирішити шляхом прийняття рішень, орієнтованих на майбутнє, є зменшення неминучої невизначеності. З цієї причини аналіз може бути використаний як один із інструментів доведення раціональності короткострокових чи довгострокових рішень, а також як спосіб оцінки якості правильності рішень, прийнятих у минулому, або як спосіб передбачити очікувані результати.

Сьогодні існує багато статистичних і ймовірнісних методів розв'язання задач прогнозування, серед яких найчастіше використовуються моделі на основі авторегресії, експоненціального згладжування, нейронних мереж і байєсівських мереж.

Істотним недоліком авторегресійних моделей є велика кількість вільних параметрів, які необхідно ідентифікувати. Недоліком немережевої моделі є непрозорість моделювання та складність мережевого навчання. Проте кожен метод має значні переваги, які не можна ігнорувати при застосуванні в конкретних ситуаціях.

Сьогодні проблеми підвищення якості прогнозування існуючих моделей та створення адаптивних (комбінованих) моделей на основі існуючих моделей, безсумнівно, залишаються актуальними.



## **РОЗДІЛ 2. Вибір та опис методів статистично-ймовірнісного прогнозування**

### **2.1. Мережа Байєса**

Останнім часом байєсівські мережі набули популярності в різних типах методів моделювання процесів і стали самостійною галуззю в області аналізу даних. Байєсівські мережі широко використовуються для обробки часових рядів (статистичні данні) та інших можливих типах представлення даних, таких як експертні оцінки, лінгвістичні змінні тощо [36].

Варто зазначити, що байєсівські мережі застосовуються при вирішенні проблем медичної діагностики. У разі недостовірності та неповноті інформації зроблено великий прорив у уточненні діагностики різних захворювань. Хоча б через цю характеристику він заслуговує значної уваги для дослідження. Мережі також популярні у різних типах систем класифікації даних, системах автоматичного розпізнавання мовлення, маркетингу та бізнесу та багатьох інших сферах людської діяльності. Таким чином, мережа допомагає визначити причинно-наслідковий зв'язок між подіями і допомагає визначити ймовірність настання події, коли вона отримує нову інформацію про зміну стану будь-якого вузла (змінної) мережі. Однак успіх використання цього методу моделювання та створення статистичних висновків значною мірою залежить від досвіду експерта у вмінні чіткого вираження умов проблеми, вибору залежних змінних процесу для повного опису його статистики, правильного використання та формування вибірок, а також вміння правильно сформулювати висновок за допомогою побудованої мережі [37].

Можливість враховувати кількісні та якісні показники може бути основною перевагою байєсівських мереж. Також основними перевагами можуть бути можливість розгляду дискретних і безперервних змінних в моделі, враховування невизначеностей і майже відсутність обмежень на кількість змінних, і, звичайно, чіткість моделювання. На відміну від інших моделей, заснованих на рівняннях регресії, моделі мережі не вимагають повної інформації, тому опис мережі зазвичай не є складним. Також немає обмежень щодо закону розподілу змінних, що також забезпечує хорошу продуктивність як інструмент класифікації даних.

Як бонус можна додати, що байєсівські мережі дуже зручні для аналізу та звичайних користувачів, оскільки мають дуже гарне представлення у вигляді графів.

Байєсівська мережа показує логічну модифікацію та пояснення структури зв'язків між змінними в явній формі, що дозволяє врахувати попередній досвід експертів, і, звичайно, вона може пояснити їхні висновки більш чітко. Причина, чому мережі можуть успішно вирішувати практичні завдання, полягає в тому, що вони базуються на основних законах, аксіомах, твердженнях і положеннях теорії ймовірностей, які вивчалися і вдосконалювались протягом сотень років.

При вивченні можливих застосувань цього методу та його дослідженнях було виявлено багато недоліків, тому розглянемо їх більш детально нижче. При використанні байєсівської мережі часто буває важко знайти метод точного ймовірнісного висновку на основі навчальних вибірок: щоб обчислити ймовірність вершин вузла, необхідно використовувати загальнодоступну матрицю ймовірностей замість умовної ймовірності емпіричних значень. Основна мета полягає в тому, що це не дозволяє не розробляти структуру мережі заздалегідь. До всього цього ми не можемо ігнорувати труднощі вибору змінних для побудови моделі. Використовується алгоритм для конструктивного вибору змінних для створення прогнозної високопродуктивної байєсівської мережі.

Байєсівські мережі — це ймовірнісні мережі, які використовують байєсівські формули та їх узагальнення для вивчення необхідних даних про ймовірності мережі. По суті, це графічні моделі подій і процесів, засновані на поєднанні результатів досліджень теорії ймовірностей і теорії графів. Для побудови графічної моделі використовується поняття «модульності» - поділ великої складної системи на прості елементи. Потім, для того, щоб об'єднати елементи в один, ми використовуємо результати теорії ймовірностей, яка забезпечують хорошу продуктивність моделі для практичних завдань і навіть дозволяє поєднувати базу даних і графіки. Цей метод моделювання сприяє створенню структур даних для подальшої розробки, винаходить ефективні алгоритми обробки даних і прийняття рішень, а також дозволяє будувати моделі процесів навіть для набору сильно взаємодіючих змінних [36].

Якщо розглядати мережу байєса із математичної точки зору, то можна сказати що це модель подання існуючих і відсутніх імовірнісних залежностей. Якщо подія  $A$  є причиною  $B$ , то зв'язок  $A \rightarrow B$  називається причинним, тобто коли існує механізм, при якому значення  $A$  безпосередньо чи не безпосередньо впливає на значення  $B$ . Коли всі мережеві з'єднання є причинними, то таку мережу називають каузальною мережею.

Щоб полегшити формування байєсівських висновків щодо відомих конкуруючих гіпотез набору даних інтерпретації, необхідно описати необхідну послідовність дій (алгоритмів). Отже, для кожної гіпотези необхідно виконати наступне:

- парадоксальна або попередня інформація, що міститься в даних, буде записана як ймовірності;
- обчислити добуток ймовірностей;
- з метою отримання апостеріорної ймовірності для кожної гіпотези потрібно нормалізувати результати (при наявній інформації)
- обрати гіпотезу, яка має максимальну ймовірність.

Іноді статистичних даних достатньо для обчислення апіорної ймовірності. Однак у більшості випадків це неможливо через суб'єктивну складність отримання статистичних даних, а апіорні значення можуть бути представлені в інших формах [37; 38].

## 2.2 Прогнозування за допомогою нейронних мереж

Прогнозна конструкція, заснована на імовірнісних методах і суб'єктивних знаннях аналітиків, набирає значного оберту і поширення. Зупинимось на недоліках імовірнісних методів:

- відсутність концептуальної моделі структури та системи зв'язку реальних об'єктів;
- складність побудови моделей за умови, що дані мають тимчасове зміщення відносно один одного і зберігаються в різних часових рядах;
- зазвичай можлива недостатня точність прогнозу;
- отримане значення дуже чутливе до недостатньої інформації або її шуму;

– якість прогнозів сильно залежить від кваліфікації та досвіду експертів у певній предметній області.

Таким чином, недоліки даного методу призводять до необхідності розробки нових математичних моделей на основі методів штучного інтелекту, а також алгоритмів та спеціального програмного забезпечення для вирішення задачі підвищення надійності та точності прогнозів. Вони також можуть працювати з недостатньою інформацією або навіть значним шумом, що дозволяє отримати результат за короткий проміжок часу.

Нейронна мережа є одним із найпопулярніших методів вирішення питання прогнозування часових рядів математичних моделей. Метою прогнозування є зниження ризику прийняття рішень. Прогнози зазвичай помилкові, але помилка залежить від використовуваної системи прогнозування.

Збільшуючи наші ресурси прогнозування, ми можемо підвищити точність і навіть зменшити втрати, пов'язані з невизначеністю в процесі прийняття рішень. Не можна забувати, що жоден метод прогнозування не може повністю усунути ризик при прийнятті рішень, тому ми завжди повинні чітко визначати неточність прогнозу. Тому остаточне рішення має ґрунтуватися на отриманих точних прогнозах з урахуванням можливих помилок прогнозування.

Крім вищезгаданої байєсовської мережі, ще одним перспективним методом кількісного прогнозування є використання нейронних мереж. В останні роки розвиваються дослідження нейронних мереж.

Перша перевага – гнучка конструкція. У межах обраної архітектури мережі експертам потрібно лише налаштувати й регулювати кількість шарів навчання та нейронів, щоб глобально змінити результат. Ви також можете вибрати іншу функцію активації, щоб збільшити свободу дії. Така настройка дозволить вибрати найкращий стан мережі з обраною архітектурою та максимізувати точність навчання мережі для отримання найкращого прогнозу та мінімальної помилки. З інформацією про минулі події ці кроки та оптимізація мережі застосовуються навіть до найпростішої архітектури. Ви навіть можете змінити видимість зовнішніх параметрів.

Наступною перевагою є адаптивність, тобто експерт не залежить від математичної моделі поведінки часового ряду що досліджується, тому що фактичний процес побудови моделі нейронної мережі здійснюється під час навчання, і ніяких додаткових зусиль з боку експерта не потребується.

Основним недоліком є невизначеність нейронної мережі. Тобто неможливо знати логіку прийняття мережевих рішень. Така поведінка називається «чорним ящиком». Певним чином це працює, але логіка прийняття рішень прихована від людей. Для створення мережевої експертної системи був винайдений алгоритм вилучення знань з нейронної мережі, який допомагає розкласти і формалізувати вивчену нейронну мережу в список логічних правил [39]. Відомі вчені, які займалися дослідженнями в галузі штучних нейронних мереж це Бодянський Є.В., Руденко О.Г., Галушкін А.І., Барт Косько, Козадаєв А.С. та багато інших [40].

Розглянемо використання нейронних мереж для основних етапів створення та прогнозування. Підготовчий етап – це збір даних, у тому числі пошук якісних даних з доступних джерел, які містять повну та достатню інформацію в необхідних полях діяльності. Загалом, на цьому етапі потрібно враховувати наступні моменти:

- доцільність знайдених даних та їх своєчасність;
- методи відновлення розриву в спостереженнях (є два найбільш відомих методи відновлення відсутніх спостережень: перший полягає в тому, щоб просто виключити рядки або стовпці відсутніх значень у матриці даних. Однак зазвичай цей пропуск не дозволяє вам побудувати часовий ряд. Інший варіант — припустити, що втрачені дані можуть бути відновлені);
- очищення та фільтрація даних (очищення даних має включати видалення небажаних ефектів шуму, щоб отримати кращі результати спостереження за навчанням мережі. Крім того, під час очищення даних потрібно порівняти введені значення з очікуваним діапазоном варіації спостереження).

Аналіз вхідних даних безпосередньо залежить від проблеми, яку необхідно вирішити. Зібрані для аналізу спостереження являють собою набір багатовимірних векторів, які описують досліджуваний процес.

Наступним кроком буде перетворення даних і його призначення. У більшості випадків перетворення необхідно для:

- класифікації вхідних даних;
- перетворення з лінійних задач в нелінійні;
- проведення дослідження в межах певного діапазону спостережень.

Найпоширенішими і найпростішими методами є логарифмування та використання відсоткових різниць, які можна легко обчислити у вигляді таблиці. Більш складними і менш застосовними є статистичні перетворення, такі як вимірювання нахилу лінії регресії в напрямку тенденції потоку даних, або більш складні на основі перетворення Фур'є, іноді воно несе більше інформації, ніж вихідні дані.

Варто відзначити, що нейронна мережа погано працює при спостереженні великої кількості вхідних значень різних значень. Щоб мінімізувати цей негативний вплив, прийнято масштабувати вхідні дані в діапазоні  $[0 \dots 1]$  або  $[-1 \dots 1]$ . Однак функції та методи, які зазвичай використовуються для масштабування, не дозволяють отримати країв відрізків, близьких до вихідних значень. Тому результат зазвичай знаходиться лише в діапазоні  $[0,2 \dots 0,8]$  або  $[-0,8 \dots 0,8]$ .

Наступним кроком є вибір змінних, метою яких буде зменшення розмірності вхідних даних. Це тому, що менше вхідних змінних зазвичай гарантує вищу ефективність нейронної мережі.

Сьогодні існує три основні методи пошуку оптимальної кількості мережевих вхідних вузлів. Спочатку використовуються статистичні методи або методи нелінійної динаміки для вирішення проблеми частоти вибірки. Відомий метод аналізу чутливості для визначення важливості вхідних даних можна використовувати як інший метод навчання мережі. Остання процедура вважається дуже ефективною і заснована на ранжуванні вузлів і виключенні точок з найнижчим рейтингом. Це робиться наступним чином: спочатку виберіть по черзі набір підготовлених спостережень, потім тренуйте мережу на прийнятному рівні помилки, потім виконайте аналіз чутливості, і останній крок – виключити всі змінні нижче вибраного рівня.

### 2.3. Метод експоненційного згладжування

Ще в другій половині минулого століття почався активний розвиток методів експоненційного згладжування. Тоді вони потрібні були для вирішення питання прогнозування одночасного руху великої кількості різних товарів на складі. Саме тоді ці методи набули популярності через свою беззаперечну простоту та ефективність. Найвідомішими є звичайний, подвійний і потрійний методи згладжування, а також методи Холта і Холта-Вінтерса.

Якщо в часовому ряді очевидно немає сезонності та тенденції, ви, безсумнівно, можете використовувати простий метод експоненціального згладжування. Якщо в ряду є лінійна тенденція, слід використовувати подвійне експоненціальне згладжування або метод Холта. Якщо тенденція нелінійна, може допомогти потрійне згладжування. Якщо це ряд із сезонністю та тенденцією, незалежно від того, як вони поєднуються (множення чи додавання), слід використовувати метод Холта-Вінтерса.

Давайте детальніше дослідимо ідею розглянутих методів. Метод простого експоненційного згладжування передбачає лише згладжування часового ряду за допомогою ковзного середнього з експоненційними ваговими коефіцієнтами. Очевидно, назва методу вказує на те, що дані згладжуються за допомогою середньозваженого значення, а вага змінюється експоненціально. Порівняно з початком інтервалу дослідження, ковзне середнє може краще характеризувати значення в кінці процесу.

Формалізуємо метод простого експоненційного середнього: допустимо що у нас є ряд спостережень зі змінною  $y$ :  $y_1, y_2, \dots, y_n$ . Тоді просте експоненційне середнє можна порахувати за формулою:

$$S_n^1 = \sum_{j=0}^{n-1} \alpha(1-\alpha)^j y_{n-j}$$

де  $S_n^1$  – експоненційне середнє першого порядку для  $n$  – го періоду;  $\alpha$  – коефіцієнт згладжування:

$$0 < \alpha < 1, \alpha(1-\alpha)^j \geq 0, \sum_{j=0}^{n-1} \alpha(1-\alpha)^j \Rightarrow 1$$

Просте експоненційне середнє є прогнозом досліджуваного значення в періоді  $(n+1)$ , тобто:

$$\hat{y}_{n+1} = S_n^1.$$

Також для простого експоненційного середнього справедлива така формула рекурсивна формула:

$$S_n^1 = \alpha y_n + (1 - \alpha) S_{n-1}^1.$$

Ми побачили, що для використання формули нам потрібно визначити початкове значення. Це один з найбільших недоліків цього методу – відсутність формалізації при виборі значення параметра згладжування  $\alpha$ . Далі розглянемо деякі з найпоширеніших способів вибору параметрів згладжування.

### 2.3.1. Методи вибору $\alpha$

Якість параметру  $\alpha$  безпосередньо впливає на якість результатів прогнозування. Чим ближче значення  $\alpha$  до одиниці, тим більший вплив на прогноз. Якщо значення  $\alpha$  близьке до нуля, то старше значення в ряді матиме більший вплив, оскільки вага зменшується повільніше. Зазвичай значення цього параметра знаходиться в діапазоні від 0,1 до 0,3. Ця пропозиція була висунута вченим Брауном для того, щоб передбачити реальні проблеми, але його пропозиція не була нічим підкріплена [36].

Нижче наведено приклад ітераційного процесу вибору значень  $\alpha$  [41]:

- по-перше, ми обираємо критерії оцінки якості прогнозу. Це можуть бути СКП, СПП, АСПП та інші;
- вибираємо крок зміни для параметра  $\alpha$  наприклад з кроком 0.1. Отже, ми визначаємо, що підмножина значень дорівнює: [0; 0,1; 0,2; .....; 0,9; 1];
- вибираємо початкове наближення,
- обчислюємо кожне значення з побудованої підмножини простим методом експоненціального згладжування



- вибираємо один з критеріїв оцінки якості прогнозу та розраховуємо його значення: МАП, МАПП та ін.;
- оцінюємо отриманий набір значень і знаходимо результат, що відповідає мінімальному значенню стандарту.

### 2.3.2. Вибір початкового наближення

Особливу увагу слід звернути на експоненціальне усереднення різних порядків, яке відіграє важливу роль у простих, подвійних і потрійних експоненційних методах згладжування. Їх знаходять за рекурсивними формулами, але є проблема зі значенням початкового наближення. На сьогоднішній день не існує єдиного методу визначення початкових наближень, і зазвичай вони підбираються виходячи з конкретної предметної області дослідження.

Методом вибору початкового наближення є можливість взяти середнє з кількох початкових або кінцевих значень ряду. Також можна використовувати приблизні значення як параметри:

$$S_0^1(y) = \frac{\sum_{i=1}^n y_i}{n}$$

## 2.4. Метод групового врахування аргументів

### 2.4.1. Основні ідеї методу

Одним із популярних сьогодні методів прогнозування є метод групової аргументації (МГВА), який базується на комбінації допоміжних функцій, системи правил для ускладнення моделі, критеріїв відбору та методу регуляризації за допомогою зовнішніх критеріїв. Спочатку вибирається модель-кандидат, а потім з неї вилучаються моделі, які не пройшли поріг відбору. Це алгоритм самоорганізації. Нижче ми покроково описуємо самоорганізуючу структуру алгоритму моделі:

- Системна обробка спостережень на основі обраних допоміжних функцій для зменшення кількості заявників;
- Згенерувати набір моделей-кандидатів;

– Розрахувати критерії відбору, які є зовнішніми доповненнями, та знайти найкращу модель складності.

Існує багато різних модифікацій у методі групування параметрів і вони відрізняються:

- кількістю виділених рядків;
- кількістю рівнянь;
- розрахунком залишків (існування чи неіснування).

Крім того, метод групового врахування параметрів можна розділити на однорядкові і багаторядкові відповідно до кількості вибраних рядків. Багаторядкові алгоритми зазвичай використовуються для моделювання ситуації, коли в задачі є невизначеність, оскільки регресійні моделі не можна використовувати — для них не можна побудувати відповідну модель процесу за межами інтерполяції.

#### 2.4.2. Зовнішні критерії оптимальності

Стандарти неупередженості та регулярності користуються популярністю при виборі моделей і використовуються як зовнішні доповнення в роботі. Загальним пунктом наступних стандартів є те, що вони визначаються на основі тестових зразків, а також дозволяють повторювати результати.

Реалізація всієї вибірки  $N$  поділяється на реалізацію навчальної вибірки  $N_A$ , яка використовується для навчання моделі, потім оцінюються параметри моделі і реалізація тестової вибірки  $N_B$ , яка використовується для перевірки моделі і обрання найкращої.

Потрібно використати критерій регулярності, щоб визначити стандартне відхилення моделі на тестовому зразку:

$$\Delta^2(B) = \frac{\sum_{t \in N_B} (y_t^M - y_t)^2}{N_B} \rightarrow \min$$

Якщо за набору постійних умов хороша апроксимація в минулому гарантує хорошу апроксимацію в найближчій перспективі, то критерій регулярності можна використовувати для короткострокового прогнозування. Слід зазначити, що результат, отриманий на більш новішій реалізації, має лише

невелике відхилення від вхідних даних. Хоча важливі змінні можуть бути втрачені в процесі відбору, їх вплив буде врахований через інші змінні.

Розглянемо більш детально стандарт неупереджених (або непослідовних) моделей. Основною ідеєю є те, що для одного об'єкту дослідження по різних вибіркам даних, отриманих від нього за інших рівних умов, мають бути отримані близькі моделі, за допомогою яких можна визначити поведінку об'єкту.

Стандартний запис критерію такий:

$$n_{cm} = \frac{1}{R_1 + R_2} \cdot \sum_{r=1}^{R_1+R_2} (z_r^* - z_r^{**})^2$$

де  $R_1, R_2$  – розміри першої та другої  $z_r^*, z_r^{**}$  - значення прогнозу першої та другої моделі відповідно до всіх вибраних точок.

### 2.4.3. Опис алгоритмів селекції

Далі розглянемо поліноміальні алгоритми методу групового врахування аргументу. Припустимо, що повний опис об'єкта дається наступними залежностями .

$$y = \varphi(x_1, x_2, \dots, x_n)$$

Замінімо цей вираз базисом часткових поліноміальних описів:

$$y_1 = f(x_1, x_2), y_2 = f(x_1, x_3), \dots, y_s = f(x_{n-1}, x_n), s = C_n^2.$$

Алгоритм комбінації одного рядка здійснює пошук усіх можливих моделей на заданому базисі та вибирає найкращу модель серед інших відповідно до заданих критеріїв відбору. При перебірці кількість параметрів, поступово збільшується від 1 до максимального числа  $n$  (кількості параметрів базового набору функцій). Отже, загальна схема комбінаторного алгоритму включає такі операції:

- за МНК визначаються коефіцієнти всіх часткових моделей при складності  $s = 1, s = 2, \dots, s = n$ ;
- розраховуються значення зовнішнього індивідуального або комбінованого критерія вибірки;

– з умов критерію мінімуму вибирається єдина модель оптимальної складності.

Можна сказати, що комбінаторний алгоритм МГВА базується на повній математичній індукції, оскільки не пропускає жоден з можливих варіантів моделі, що закладений на вихідному базисі.

Розглянемо, як реалізувати цю загальну схему в поліноміальному алгоритмі. Набір опорних параметрів тут є  $n$  членом деякого поліному. Наприклад, у випадку трьох змінних повний квадратний поліном має такий вигляд:

$$q = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_1^2 + a_5 x_2^2 + a_6 x_3^2 + a_7 x_1 x_2 + a_8 x_1 x_3 + a_9 x_2 x_3$$

Цей тип полінома є лінійним щодо коефіцієнтів, які визначаються за допомогою МНК використовувати.

За поданою схемою алгоритм працює так. Спочатку всі моделі визначаються при  $S = 1$ , тобто ті моделі, що складаються з одного параметра:

$$q_1 = a_0, q_2 = a_1 x_1, q_3 = a_2 x_2, \dots, q_{10} = a_9 x_2 x_3.$$

Кількість таких моделей –  $C_{10}^1$ . Після цього розглядаються всі можливі моделі при  $S = 2$ , що містять два параметри:

$$\begin{aligned} q_{11} &= a_0 + a_1 x_1, & q_{12} &= a_0 + a_2 x_2, \dots, \\ q_i &= a_0 + a_9 x_2 x_3, \dots, & q_j &= a_1 x_1 + a_2 x_2, \dots, \\ q_k &= a_1 x_1 + a_9 x_2 x_3, \dots, & q_{45} &= a_8 x_1 x_3 + a_9 x_2 x_3. \end{aligned}$$

Всього параметрів буде  $C_{10}^2 = 45$ . Спосіб побудови інших моделей аналогічний. Отже, загальна кількість  $p_n$  можливих часткових моделей, побудованих із  $n$  параметрів повного полінома методом пошуку певної комбінації дорівнює

$$p_n = \sum_{s=1}^n C_n^s = 2^n - 1.$$

Очевидно, що зі збільшенням  $n$  кількість моделей зростає дуже швидко – додавання одного параметра повоює час обчислення.

Вважається, що початкова комбінація параметрів з якого починається процедура багаторядкового вибору моделі процесу базується на так званому

нульовому ряді алгоритму. Наприклад, у класі алгебраїчних функцій найпоширенішою моделлю є поліном Колмагорова-Габора від  $k$  змінних:

$$q = a_0 + \sum_{i=1}^k a_i z_i + \sum_{i=1}^k \sum_{j=1}^k a_{ij} z_i z_j + \\ + \sum_{i=1}^k \sum_{j=1}^k \sum_{l=1}^k a_{ijl} z_i z_j z_l + \dots$$

Після перетворення всіх доданків, отримаємо лінійний поліном

$$q = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n.$$

Членами цього поліному є параметри, побудовані на нульовому ряді алгоритму.

Далі покажемо як працює багаторядковий алгоритм. У першому рядку на основі даних таблиці спостережень частковий опис будується з усіх попарних комбінацій вихідних параметрів, що наближують за МНК вихідну змінну  $q$

$$y_1 = f_1(x_1, x_2), y_2 = f_2(x_1, x_3), \dots, y_k = f_k(x_{n-1}, x_n).$$

У другому рядку отримані змінні приймаються як вхідні параметри і знову будуються всі часткові моделі від двох параметрів

$$z_1 = \phi_1(y_1, y_2), z_2 = \phi_2(y_1, y_3), \dots, z_l = \phi_l(z_{F_1-1}, z_{F_1}), l = C_{F_1}^2.$$

Із них за критерієм обирається  $F_2$ , що використовується в якості параметру для наступного ряду вибірки. Потрібно повторювати цей процес, поки стандартне значення не зменшиться. Саме так в багаторядних алгоритмах МГВА відтворюють схему масової селекції. Вони мають генератор опису що ускладнюється від рядка до рядка та порогові відбори кращих з них, тобто, як і генетичний алгоритм, реалізує гіпотезу відбору [36].

## Висновки до розділу 2.

Описано такі методи прогнозування часових рядів: метод експоненційного згладжування та метод групового врахування аргументів (МГВА).

Серед переваг відзначимо швидкість, простоту та математичну обґрунтованість результатів, отриманих експоненційним згладжуванням; немає необхідності вказувати структуру моделі про використанні методу МГВА;

отримання моделі у вигляді математичного рівняння в результаті застосування кожного з цих методів.

Недоліком МГУА відзначена ресурсоемність цього методу, що робить його непридатним для широкого кола завдань. Метод знаходження траєкторій вимагає повної реконструкції моделі при зміні будь-якого параметру. Однак у довгостроковій перспективі він може забезпечити достатньо якісні прогнози без необхідності проводити багато ресурсомістких обчислень.

Крім того, було досліджено використання нейронних мереж для яких основними перевагами є гнучкість та адаптивність, тобто незалежність експертів від математичної моделі поведінки часового ряду, що досліджується. Недоліком є те, що модель являє собою чорний ящик.

Також в роботі наведено опис імовірнісних методів прогнозування на основі байєсівських мереж. З кожним днем вони все частіше використовуються в інформаційних системах для обробки статистичних даних, представлених 68 часовими рядами і періодами часу, а також якісними даними, представленими експертними оцінками, лінгвістичними змінними, значеннями інтервалів тощо.

### **Розділ 3. Порівняльний аналіз результатів моделювання, отриманих нейромережними та економетричними методами**

Deductor – це аналітична платформа, основа для створення закінчених прикладних розв'язків у галузі аналізу даних. Реалізовані в Deductor технології дозволяють на базі єдиної архітектури пройти всі етапи побудови аналітичної системи: від консолідації даних до побудови моделей та візуалізації отриманих результатів.

До появи аналітичних платформ аналіз даних здійснювався здебільшого у статистичних пакетах. Їх використання вимагало високої кваліфікації користувача.

Більшість алгоритмів, реалізованих у статистичних пакетах, не дозволяло ефективно обробляти великі обсяги інформації. Для автоматизації операцій доводилося використовувати вбудовані мови програмування.

Наприкінці 80-х років відбулося стрімке зростання обсягів інформації, що накопичувалися на машинних носіях і збільшилися потреби бізнесу в застосуванні аналізу даних.

Відповіддю цьому стала поява нових парадигм у аналізі: сховища даних, машинне навчання, Data Mining, Knowledge Discovery in Databases. Це дозволило популяризувати аналіз даних, вивести його на промислову основу і вирішити величезну кількість бізнес-завдань з великим економічним ефектом.

Вінцем розвитку аналізу даних стали спеціалізовані програмні системи – аналітичні платформи, які повністю автоматизували всі етапи аналізу від консолідації даних до експлуатації моделей та інтерпретації результатів.

Перша версія Deductor побачила світ у 2000 р. і з того часу йде безперервний розвиток платформи. У 2007 р. випущено п'яту версію системи, 2009 р. – версія 5.2.

Прогнозування результату на певний час вперед, ґрунтуючись на даних за минулий час – завдання, що зустрічається досить часто (наприклад, перед більшістю торгових фірм стоїть завдання оптимізації складських запасів, для вирішення якої потрібно знати, чого і скільки має бути продано за тиждень, і

т.д.; завдання передбачення вартості акцій якогось підприємства за день і т.д. та інші подібні питання).

Deductor Studio пропонує для цього інструмент "Прогнозування".

Прогнозування з'являється у списку майстра обробки тільки після побудови якоїсь моделі прогнозу: нейромережі, лінійної регресії і т.д. Прогнозувати на кілька кроків уперед має сенс тільки при тимчасовому ряді (наприклад, якщо є дані по тижневим сумах продажів за певний період, можна спрогнозувати суму продажів на два тижні наперед). Оскільки при побудові моделі прогнозу необхідно враховувати багато факторів (залежність результату від даних день, два, три, чотири тому), то методика має особливості. Покажемо її на прикладі.

Дано. У аналітика є дані про місячну кількість проданого товару протягом кількох років. Йому необхідно сказати, яка кількість товару буде продана через тиждень, два тижні.

Після імпорту даних скористаємося діаграмою для їх перегляду. На ній видно, що дані містять аномалії та шуми, за якими важко зрозуміти тенденцію.

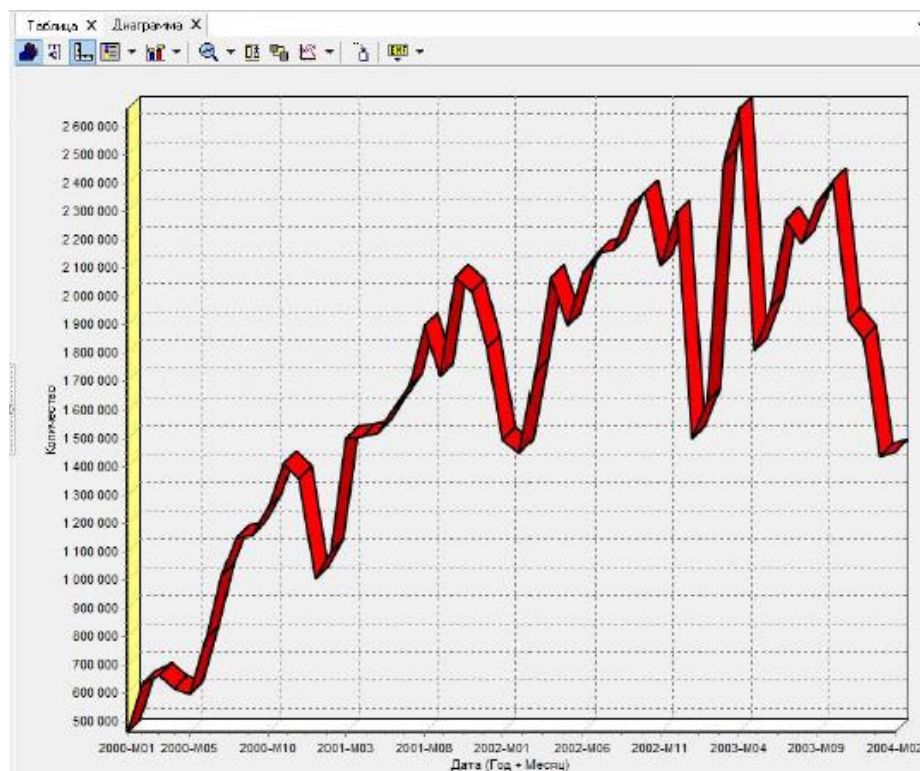


Рис. 1. Діаграма продажів



Перед прогнозуванням необхідно видалити аномалії та згладити дані. Зробити це можна за допомогою спектральної обробки. Запустимо майстер обробки (рис. 2), виберемо спектральну обробку та перейдемо на наступний крок майстра. Наступний крок відповідає за видалення аномалій вихідного набору. Виберемо поле для обробки «КОЛИЧЕСТВО» та вкажемо для нього віднімання шуму (ступінь віднімання – мала).

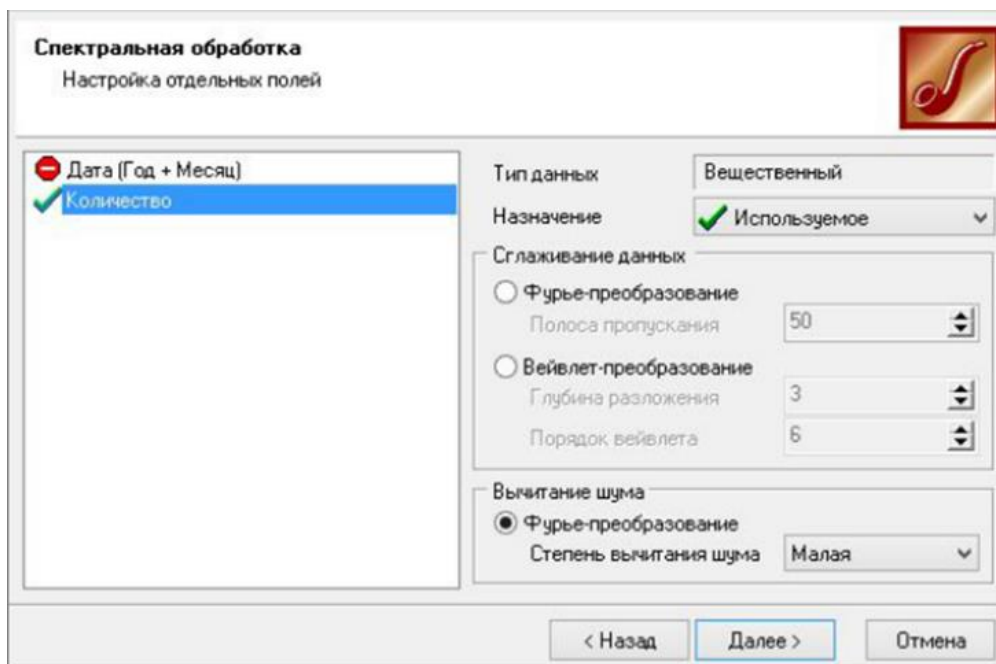


Рис 2. Вибір спектральної обробки

На наступному кроці запустимо обробку, натиснувши на «пуск» та подивимося отриманий результат (рис.3).

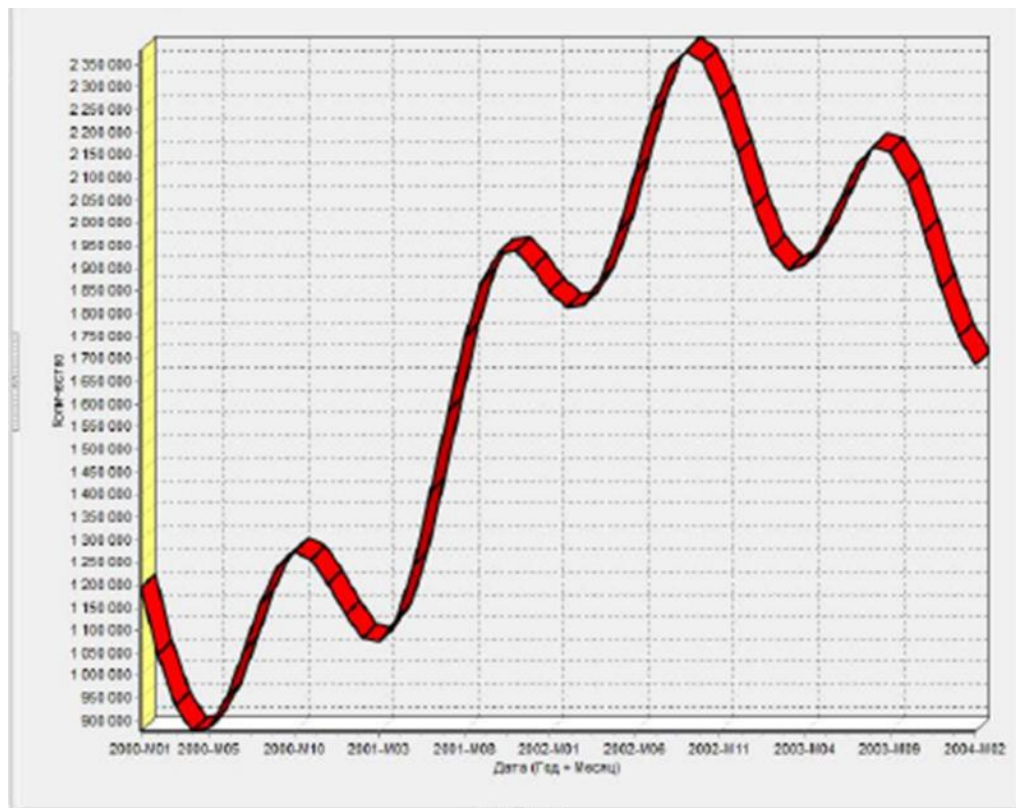


Рис 3. Результат спектральної обробки

Видно, що дані згладилися, аномалії та шуми зникли. Також видно тенденцію. Тепер постає питання, а як, власне, прогнозувати часовий ряд. У нашому випадку стовпець один. Будувати прогноз на майбутнє необхідно, ґрунтуючись на даних минулих періодів. Тобто. передбачається, що кількість продажів на наступний місяць залежить від кількості продажів за попередні місяці. Тобто. вхідними факторами для моделі можуть бути продажі за поточний місяць, продаж за місяць раніше і т.д., а результатом повинні бути продажі за наступний місяць. Тобто необхідно трансформувати дані до ковзаючого вікна.

Запустимо майстер обробки, виберемо як оброблювач ковзаюче вікно і перейдемо на наступний крок. Було вирішено будувати прогноз на тиждень вперед, ґрунтуючись на даних за 12, 11 місяців тому, два місяці тому та місяць тому. Тому необхідно, призначивши поле «КОЛИЧЕСТВО» використовуваним, вибрати глибину занурення 12. Тоді дані трансформуються до ковзаючого вікна так, що будуть доступні всі необхідні фактори для побудови прогнозу.

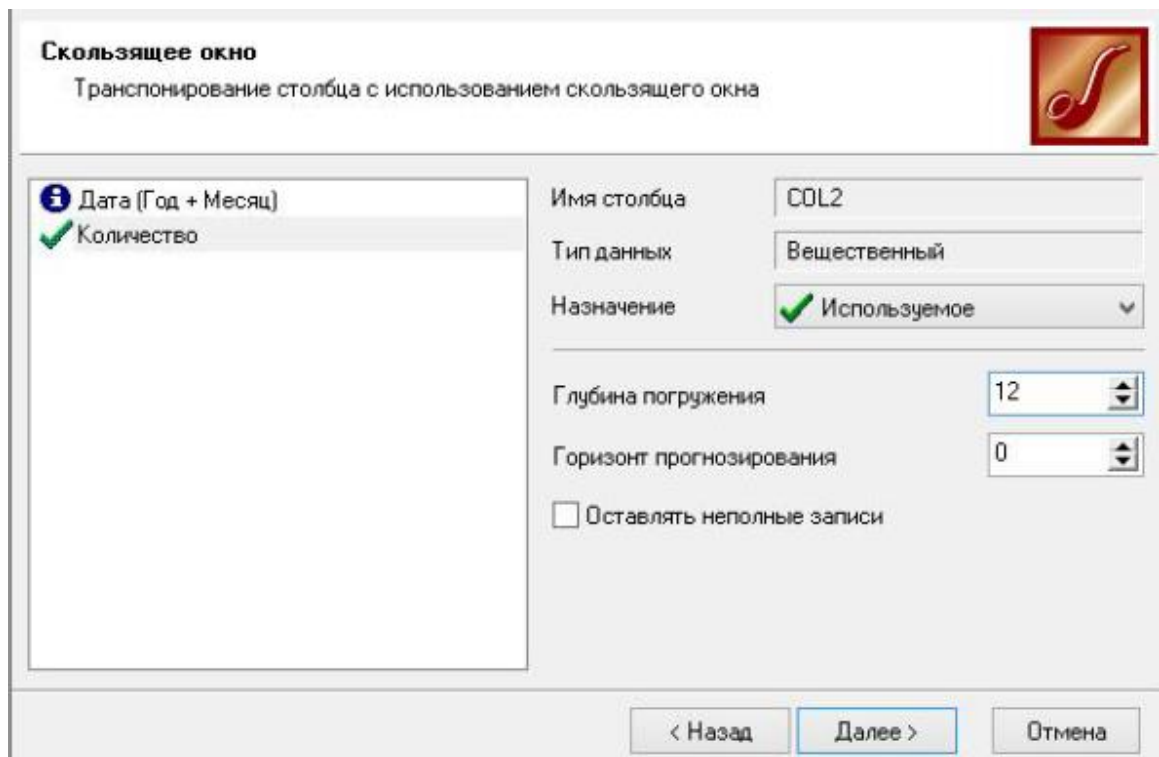


Рис 4. Ковзаюче вікно

Переглянути отримані дані можна у вигляді таблиці (рис. 5). Як видно, тепер як вхідні фактори можна використовувати «КОЛИЧЕСТВО - 12», «КОЛИЧЕСТВО - 11» - дані за кількістю 12 та 11 місяців тому (щодо прогнозованого місяця) та інші необхідні чинники. Як результат прогнозу буде вказано стовпець «КОЛИЧЕСТВО».

Перейдемо безпосередньо до самої побудови моделі прогнозу. Відкриємо майстер обробки та виберемо в ньому нейронну мережу (рис. 6). На другому кроці майстра, згідно з прийнятим раніше рішенням, встановимо як вхідні поля «КОЛИЧЕСТВО - 12», «КОЛИЧЕСТВО - 11», «КОЛИЧЕСТВО - 2» та «КОЛИЧЕСТВО - 1», а як вихідний - «КОЛИЧЕСТВО». Інші поля зробимо інформаційними.

The screenshot shows a window titled 'Таблица' (Table) with a grid of data. The columns are labeled 'Дата (Год + Месяц)' (Date (Year + Month)), 'Количество-12', 'Количество-11', 'Количество-10', 'Количество-9', 'Количество-8', and 'Количество-7'. The rows represent monthly data from 2001-01-01 to 2003-12-31. The values are numerical, representing quantities for each month.

Рис. 5. Таблица ковзающего окна

The dialog box is titled 'Настройка назначений столбцов' (Column Assignment Settings) and contains the instruction 'Задайте назначения исходных столбцов данных' (Specify the assignments of the original data columns). On the left, there is a list of columns: 'Количество-10', 'Количество-9', 'Количество-8', 'Количество-7', 'Количество-6', 'Количество-5', 'Количество-4', 'Количество-3', 'Количество-2', 'Количество-1', and 'Количество'. On the right, the configuration for the selected column 'COL2' is shown:

- Имя столбца: COL2
- Тип данных: Вещественный (Floating-point)
- Назначение: Выходное (Output)
- Вид данных: Непрерывный (Continuous)
- Статистика (Statistics):
  - Минимум: 1075589,03358277
  - Максимум: 2380312,1698561
  - Среднее: 1864843,52986291
  - Стандартное откл.: 354553,479266148

Buttons at the bottom include '< Назад' (Back), 'Далее >' (Next), and 'Отмена' (Cancel).

Рис. 6. Налаштування мастера спектральної обробки

На наступному кроці вкажемо розбиття тестової і навчальної множин (рис. 7).

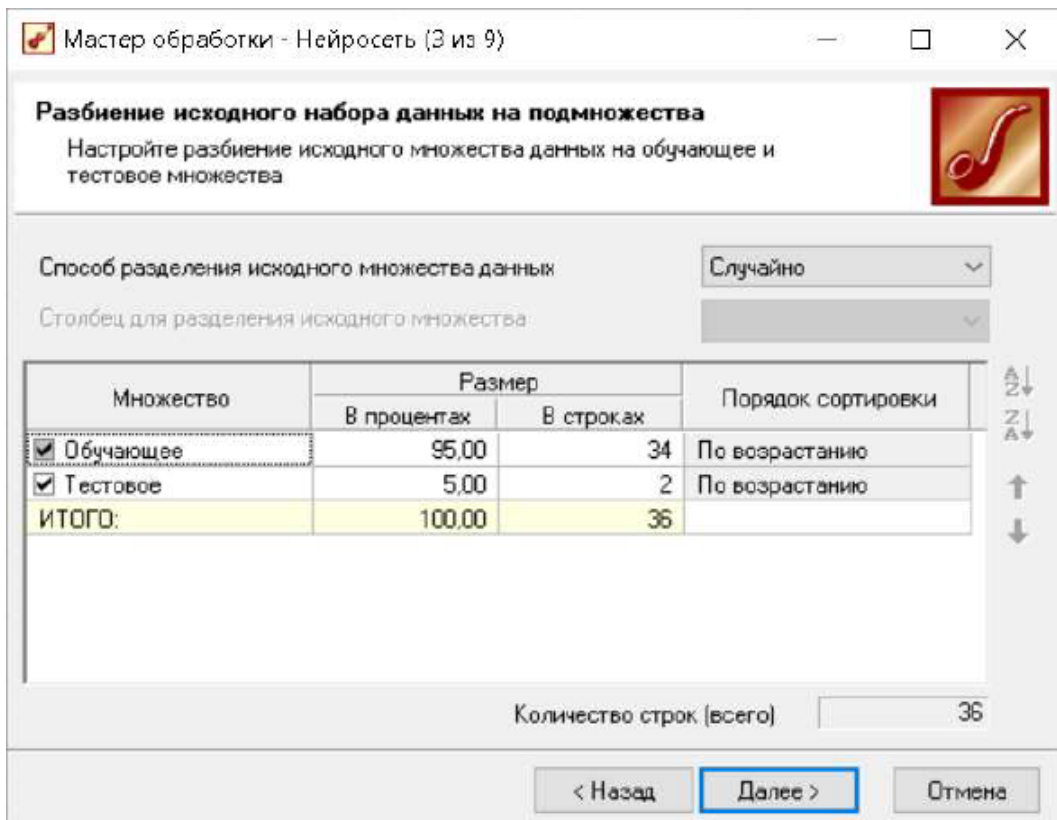


Рис 7. Розбиття набору

Перейдемо до наступного кроку, на якому відмітимо необхідну кількість шарів і нейронів в нейромережі (рис. 8).

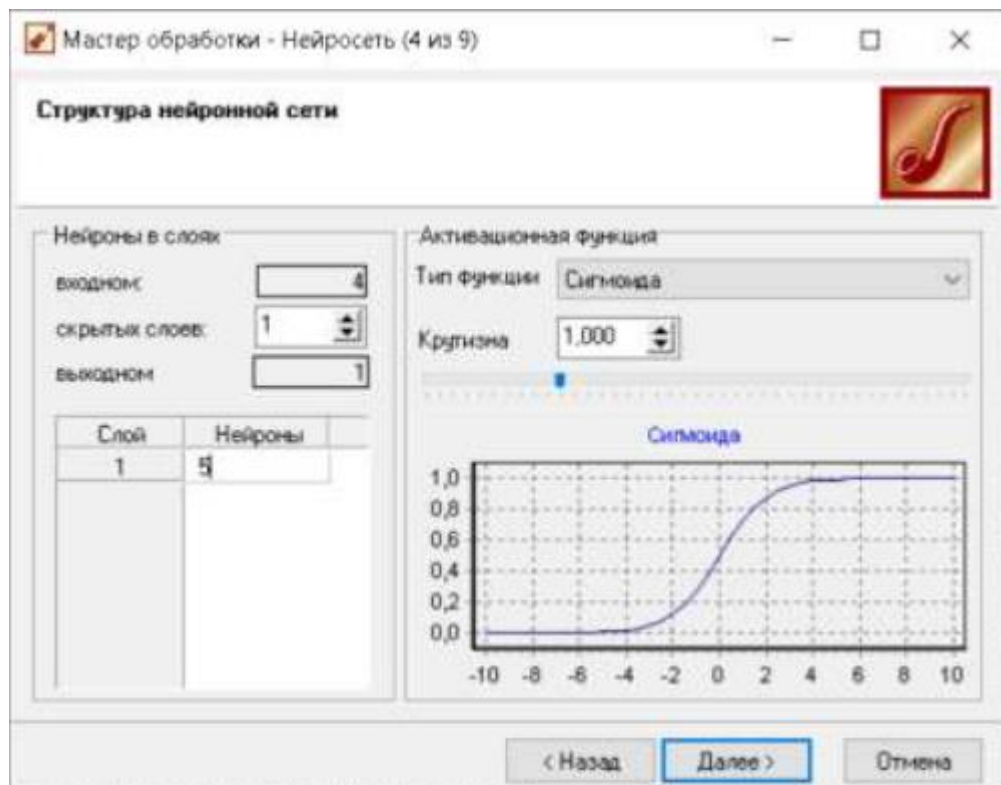


Рис. 8. Налаштування нейромережі

Далі виберемо алгоритм навчання нейромережі.

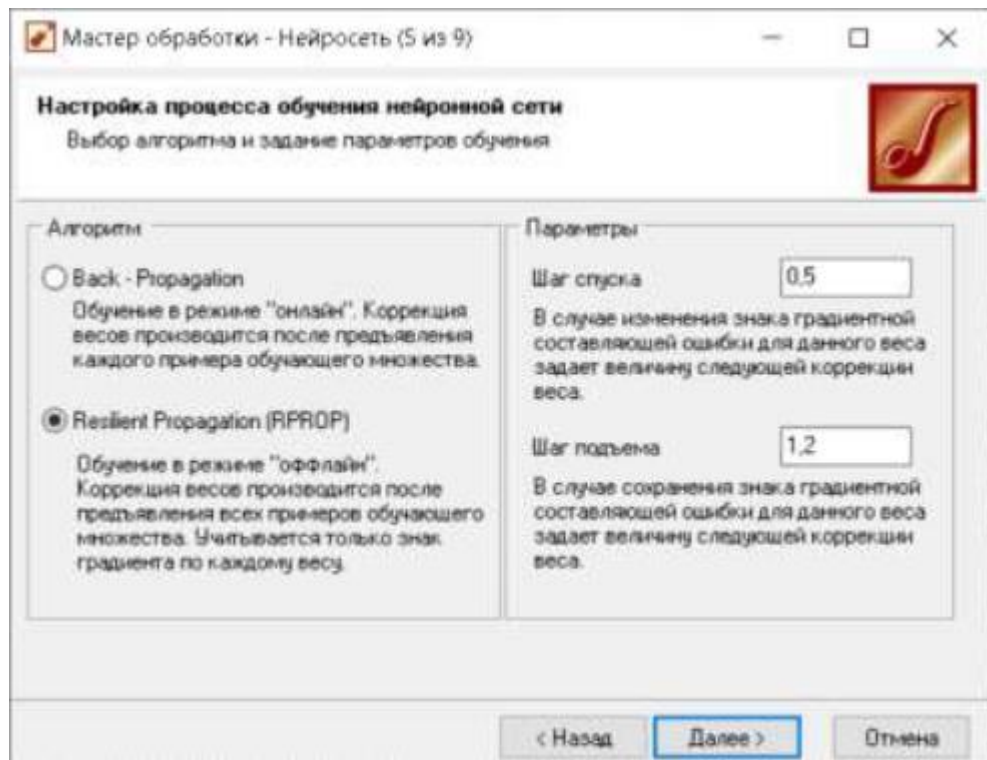


Рис. 9. Навчання нейромережі

Після побудови моделі для перегляду якості навчання представимо отримані дані у вигляді діаграми та діаграми розсіювання. У майстрі налаштування діаграми виберемо для відображення поля «КОЛИЧЕСТВО» та «КОЛИЧЕСТВО\_OUT» - реальне та спрогнозоване значення. Результатом будуть два графіка, що показані на рис 10.

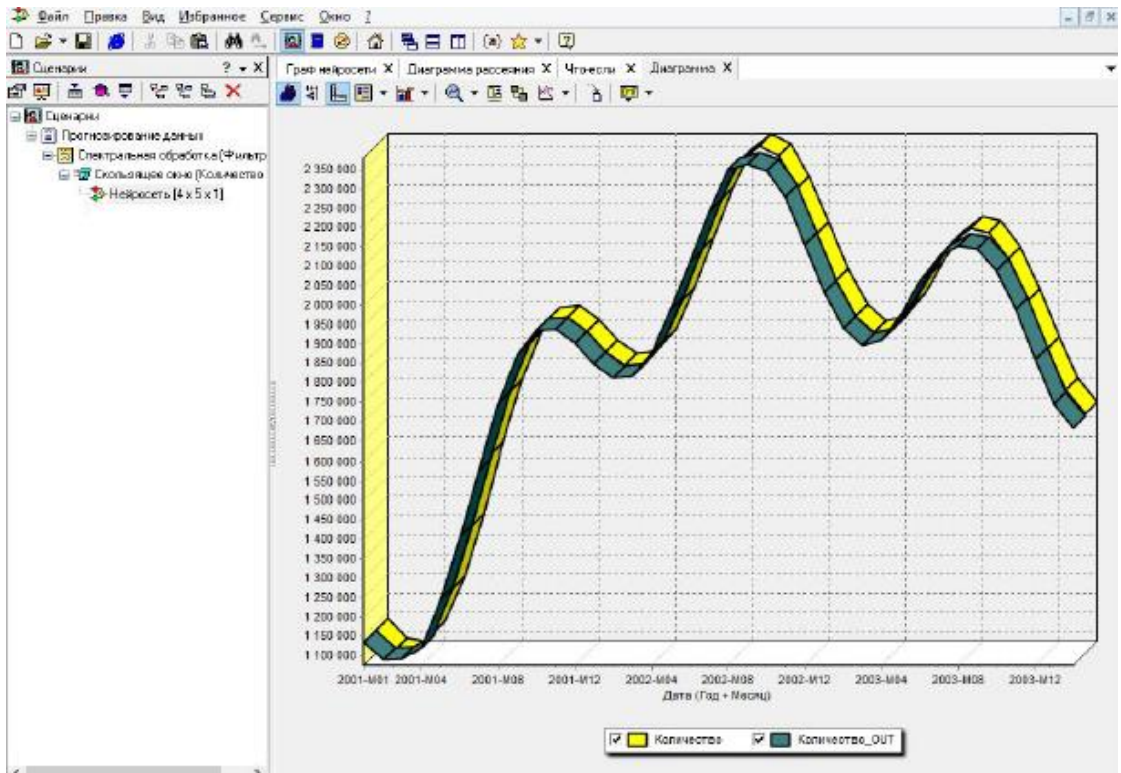


Рис 10. Порівняння еталонних даних з прогнозом  
 Діаграма розсіювання більш наочно показує якість навчання.

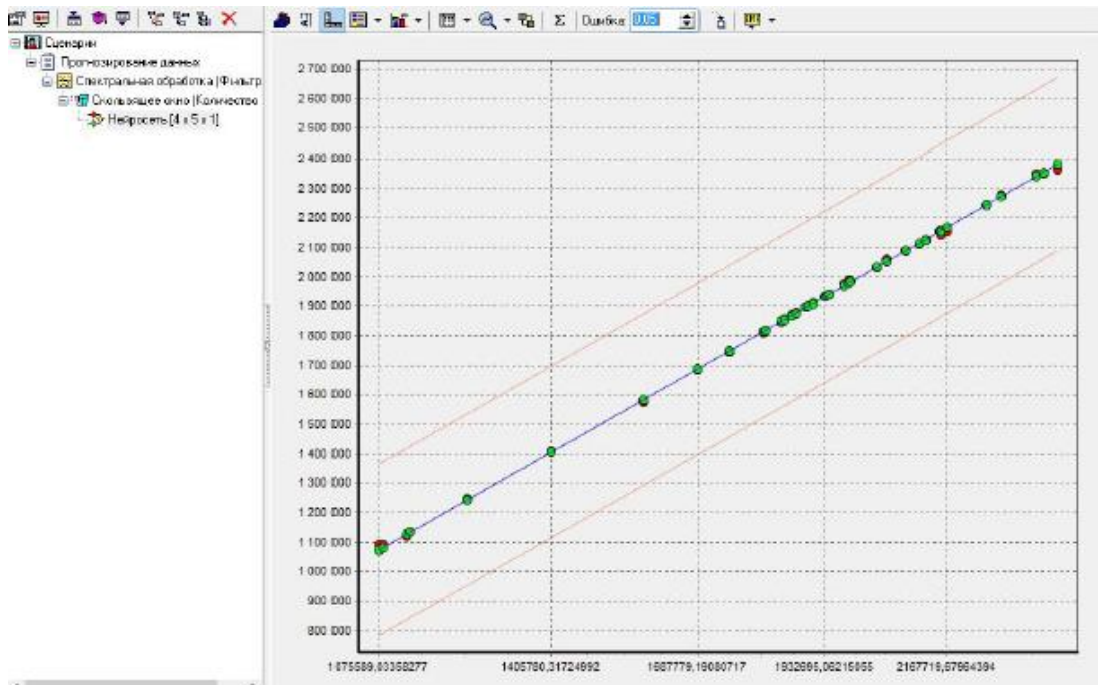


Рис. 11. Діаграма розсіювання

Нейросеть навчена, тепер залишилося найголовніше - побудувати необхідний прогноз. Для цього відкриваємо майстер обробки та вибираємо обробник, що з'явився тепер «Прогнозування» (Рис. 12).

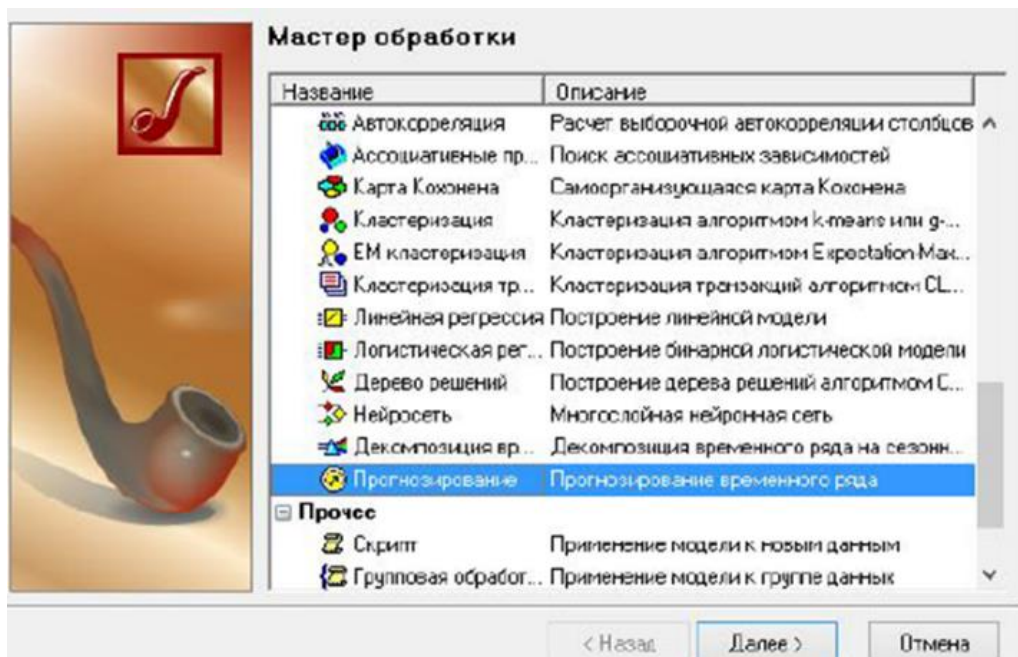


Рис 12.

На другому кроці майстра пропонується налаштувати зв'язки стовпців для прогнозування часового ряду – звідки брати дані для стовпця за чергового кроку прогнозу. Майстер сам вірно налаштував усі переходи, тому залишається лише вказати обрій прогнозу (на скільки вперед прогнозуватимемо) рівним трьом, а також, для наочності, необхідно додати до прогнозу вихідні дані, встановивши у майстрі відповідний прапорець.

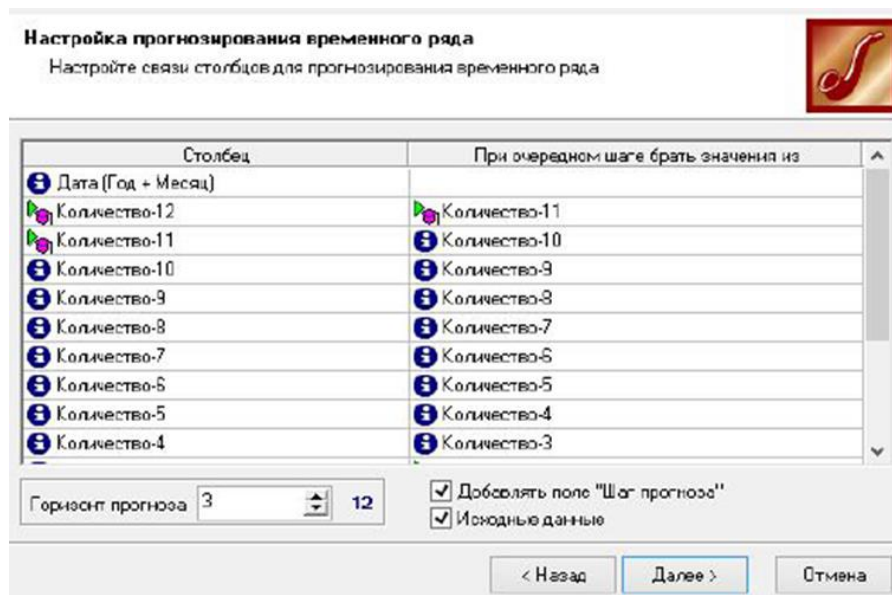


Рис. 13

Після цього необхідно як візуалізатор вибрати діаграму прогнозу, яка з'являється лише після прогнозування часового ряду. У майстрі налаштування



стовпців діаграми прогнозу необхідно вказати як відображуваний стовпець «КІЛЬКІСТЬ».

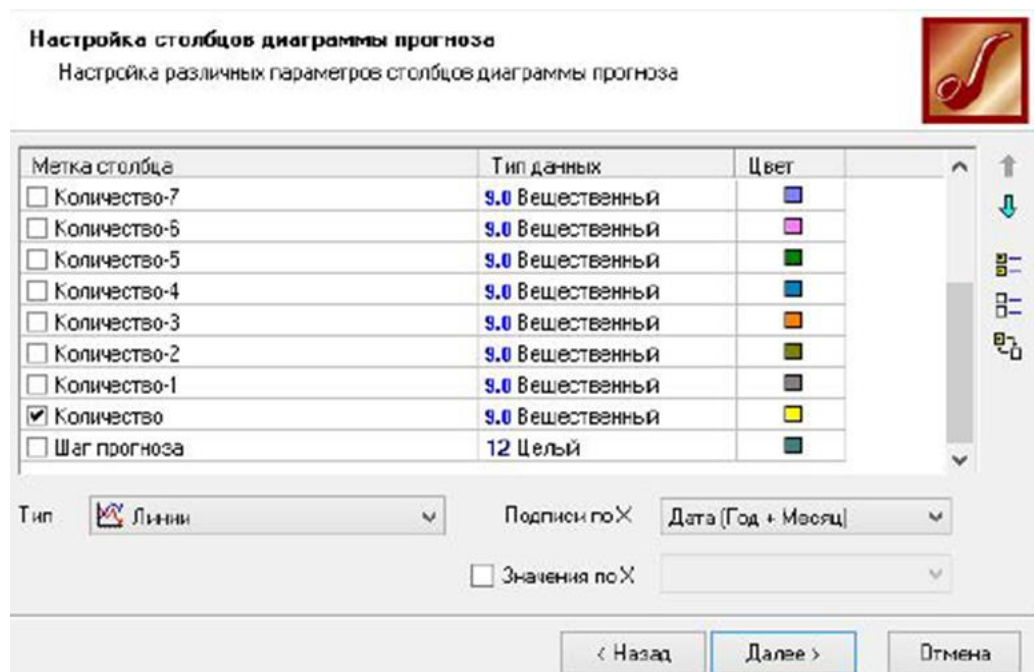


Рис. 14

Тепер аналітик може дати відповідь на питання, яку кількість товарів буде продано наступного місяця і навіть через два місяці. Масштабувавши результат і ввімкнувши мітки, можна побачити розрахункові значення на 3 місяці наперед (рис. 15).

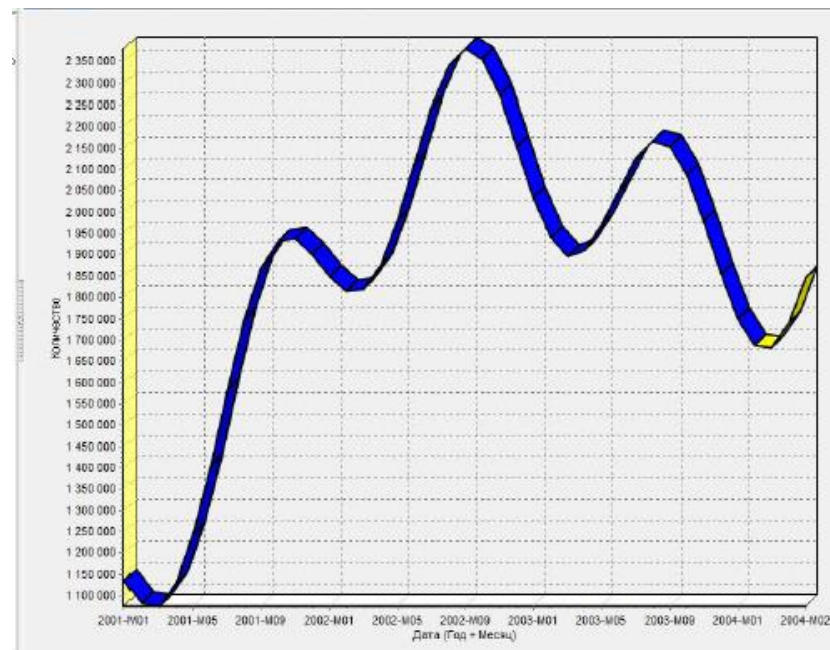


Рис. 15

Цей приклад показав, як за допомогою Deductor Studio прогнозувати часовий ряд.

При розв'язанні задачі були застосовані механізми очищення даних від шумів, аномалій, які забезпечили якість побудови моделі прогнозу і відповідно достовірний результат прогнозування кількості продажів на три місяці наперед. Також був продемонстрований принцип прогнозування часового ряду – імпорту, виявлення сезонності, очищення, згладжування, побудова моделі прогнозу та власне побудова прогнозу тимчасового ряду.

Після завершення аналізу дані можна експортувати. Так як дана версія є безкоштовною для освітніх цілей, то дані можна вивантажити або в текстовий файл, або в власний проект програми з розширенням «\*.ded».

Майстер експорту має точно такі ж налаштування, як і майстер імпорту. Більш того, якщо експорт даних зробити в текстовий файл, то далі дані можна скопіювати в файл табличного процесора Excel, і досить комфортно з ними працювати.

Даний приклад показав, як за допомогою Deductor Studio прогнозувати часовий ряд.

### **Висновки до 3 розділу**

В третьому розділі наведено порівняльний аналіз результатів моделювання, отриманих нейромережними та економетричними методами.

Прогнозування результату на певний час вперед, ґрунтуючись на даних за минулий час, - завдання, що зустрічається досить часто. Наприклад, перед більшістю торговельних підприємств стоїть завдання оптимізації складських запасів, для вирішення якої вимагається знати, що і скільки має бути продане через тиждень, завдання прогнозу вартості акцій якого-небудь підприємства через певний час тощо.

Для вирішення таких питань можна використати аналітичну платформу Deductor. Deductor Studio - це програма, призначена для аналізу інформації з різних джерел даних. Вона реалізує функції імпорту, обробки, візуалізації і експорту даних. Deductor Studio може функціонувати і без сховища даних.

В роботі розглянуто інструмент "Прогнозування" Deductor Studio.

При вирішенні задачі були застосовані механізми очищення даних від шумів, аномалій, які забезпечили якість побудови моделі прогнозу далі і відповідно достовірний результат самого прогнозування кількості продажів на три місяці вперед. Також було продемонстровано принцип прогнозування часового ряду - імпорт, виявлення сезонності, очищення, згладжування, побудова моделі прогнозу і власне побудова прогнозу часового ряду.

## **Висновки по виконаній роботі**

З урахуванням теоретичних досліджень провідних вітчизняних та іноземних учених, досвіду світової та національної практики визначено особливості моделей прогнозування.

В якості методів дослідження використані загальнонаукові методи пізнання явищ, зокрема, методи експертних оцінок, аналізу та синтезу.

В роботі виконано огляд найтипівіших процесів та популярних сьогодні статистичних та ймовірнісних методів прогнозування часових рядів. Показано актуальність та важливість якісного розв'язку загальної задачі прогнозування сьогодні. Проведений детальний аналіз наступних методів прогнозування: експоненційне згладжування, метод групового врахування аргументів, нейронні мережі та досліджено мережу Байєса. Виконано порівняльний аналіз.

Показано, що задача прогнозування часових рядів має велику актуальність для багатьох областей.

Установлено, що до теперішнього часу розроблено багато моделей для рішення задачі прогнозування часового ряду, серед яких найбільшу застосовність мають авторегресійні й нейромереві моделі.

Виявлені переваги й недоліки розглянутих моделей.

Визначено, що найбільш перспективним напрямком розвитку моделей прогнозування з метою підвищення точності є створення комбінованих моделей.

Використання нейронних мереж в усіх областях людської діяльності, у тому числі в області фінансових установах, рухається по наростаючій. З упевненістю можна сказати, що поява такого потужного і ефективного інструменту не переверне фінансовий ринок, і не "відмінить" традиційні математичні і економетричні методи технічного аналізу, або зробить непотрібною роботу висококласних експертів. Як новий ефективний засіб для вирішення самих різних завдань нейронні мережі приходять і

використовуються тими людьми, які їх розуміють, які в них мають потребу і яким вони допомагають вирішувати багато професійних проблем

За допомогою аналітичної платформи Deductor Studio виконано прогнозування часового ряду. При рішенні завдання були застосовані механізми очищення даних від шумів, аномалій, які забезпечили якість побудови моделі прогнозу далі і відповідно достовірний результат самого прогнозування кількості продажів на три місяці вперед. Також був продемонстрований принцип прогнозування тимчасового ряду - імпорт, виявлення сезонності, очищення, згладжування, побудова моделі прогнозу і власне побудова прогнозу часового ряду.

## Список використаних джерел

1. Васильев А. А. МЕТОДЫ И МОДЕЛИ ПРОГНОЗИРОВАНИЯ ОБЪЕМА ПРОДАЖ В МАРКЕТИНГЕ. – 2013.
2. Стрекалова Н. Д., Стрекалова Н. Д. Бизнес-планирование: для бакалавров и специалистов. – Издательский дом" Питер", 2012.
3. Guo P. One-shot decision theory: a fundamental alternative for decision under uncertainty //Human-centric decision-making models for social sciences. – Springer Berlin Heidelberg, 2014. – С. 33-55.
4. Dalrymple D. J. Sales forecasting practices: Results from a United States survey //International Journal of Forecasting. – 1987. – Т. 3. – №. 3. – С. 379-391.
5. Лапыгин Ю., Прохорова Н. Управление затратами на предприятии. Планирование и прогнозирование, анализ и минимизация затрат. – Litres, 2014.
6. Smets P. Decision under uncertainty //arXiv preprint arXiv:1304.1527. – 2013.
7. Вадзинский Р. Н., Вадзинский Р. Н. Статистические вычисления в среде Excel. – Издательский дом" Питер", 2013.
8. Дюков И. И., Дюков И. И. Стратегия развития бизнеса: практ. подход. – Издательский дом" Питер", 2013.
9. Granger C. W. J. Forecasting in business and economics. – Academic Press, 2014.
10. Бажанов Н. Н. ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ КАК МЕТОД ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ //Ответственный редактор: Сукиасян АА, к. э. н., ст. преп. – 2014. – С. 194.
11. Игнатъев В. М., Стрюкова А. Г. О БИЗНЕС-ПЛАНИРОВАНИИ В ОРГАНИЗАЦИИ //Ответственный редактор: Сукиасян Асатур Альбертович, кандидат экономических наук. – 2015. – С. 57.
12. Армстронг, Дж.С. Прогнозирование продаж / Дж. С. Армстронг // Маркетинг / под ред. М. Бейкера; пер. с англ. – СПб.: Питер, 2002. – С. 351–368.

13. Давнис, В.В. Современные методы анализа и прогнозирования в задачах обоснования маркетинговых решений / В.В. Давнис, В.И. Тинякова // *Маркетинг в России и за рубежом*. – 2006. – №2. – С. 16–26
14. Репин В., Елиферов В. Процессный подход к управлению. Моделирование бизнес-процессов. – Litres, 2013.
15. Taylor J. W. Short-term load forecasting with exponentially weighted methods // *Power Systems, IEEE Transactions on*. – 2012. – Т. 27. – №. 1. – С. 458-464
16. Ханк Д., Райтс А. Д., Уичерн Д. У. Бизнес-прогнозирование. – Вильямс, 2003.
17. Pankratz A. Forecasting with dynamic regression models. – John Wiley & Sons, 2012. – Т. 935.
18. Турунцева М. Прогнозирование в России: обзор основных моделей // *Экономическая политика*. – 2011. – №. 1. – С. 193-202.
19. Шевченко И. В. Некоторые модели анализа и прогнозирования временных рядов // *Системная информатика*. – 2013. – №. 2. – С. 23-40.
20. Chatterjee S., Hadi A. S. Regression analysis by example. – John Wiley & Sons, 2013.
21. Hyndman R. J., Athanasopoulos G. Forecasting: principles and practice. – OTexts, 2014.
22. Adamowski J. et al. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada // *Water Resources Research*. – 2012. – Т. 48. – №. 1.
23. Соколов Е., Измайлов Р. Экономико-математическая модель и инструментарий прогнозирования и оптимизации расходов торгового предприятия по видам рекламы. – Litres, 2014.
24. Федоренко А. И. Управление развитием компании: бизнес-планирование инвестиционных проектов // *Foresight*. – 2013.

25. Сондерс, Дж. Количественные методы в маркетинге / Дж. Сондерс // Маркетинг / под ред. М. Бейкера; пер. с англ. – СПб.: Питер, 2002. – С. 91–112. 10. Лукашин, Ю.П.
26. Егоров А. М. Алгоритм правильного прогнозирования продаж //Управление продажами. – 2012. – Т. 3. – С. 134-144.
27. Семёнычев В., Куркин Е., Семёнычев Е. Идентификация моделей жизненного цикла продукции на основе моделей авторегрессии–скользящего среднего и базисов Грёбнера. – Litres, 2014.
28. Sbrana G., Silvestrini A. Forecasting aggregate demand: analytical comparison of top-down and bottom-up approaches in a multivariate exponential smoothing framework //International Journal of Production Economics. – 2013. – Т. 146. – №. 1. – С. 185-198.
29. Corberán-Vallet A., Bermúdez J. D., Vercher E. Forecasting correlated time series with exponential smoothing models //International Journal of Forecasting. – 2011. – Т. 27. – №. 2. – С. 252-265.
30. De Livera A. M., Hyndman R. J., Snyder R. D. Forecasting time series with complex seasonal patterns using exponential smoothing //Journal of the American Statistical Association. – 2011. – Т. 106. – №. 496. – С. 1513-1527.
31. Турунцева М. и др. Некоторые подходы к прогнозированию экономических показателей //Научные труды. – 2013. – №. 89.
32. Stock J. H., Watson M. W. Introduction to Econometrics: Global Edition. – Pearson Education, 2012.
33. Armstrong J. S. Illusions in regression analysis //Available at SSRN 1969740. – 2011.
34. Green K. C., Armstrong J. S. Simple versus complex forecasting: The evidence //Journal of Business Research. – 2015.
35. Тихомиров Н. П., Тихомирова Т. М., Ушмаев О. С. Методы эконометрики и многомерного статистического анализа: учебник //М.: Экономика. – 2011



36. Ивахненко А.Г. Помехоустойчивость моделирования / А. Г. Ивахненко, В. С. Степашко - К.: Наук. думка, 1985. – 216 с.
37. Левченко Л.О. Обзор методів прогнозування фінансового стану підприємства на основі економетричних моделей / Л.О. Левченко, Д.С. Белова // Управління розвитком складних систем. - 2013. - Вип. 14. - С. 164-169.
38. Хайкин С. Нейронные сети: полный курс. / Хайкин С.- М.: ООО «И. Д. Вильямс», 2006. – 1104 с
39. Harvey A.C. Forecasting, Structural Time Series Models and the Kalman Filter / Harvey A.C. – Cambridge University Press , 1989.- 559 p.
40. Кулявець В. О. Прогнозування соціально-екрономічних процесів/ В.О. Кулявець . – К: Кондор, 2009. – 194 с
41. Ивахненко А.Г. Индуктивный метод самоорганизации моделей сложных систем / Ивахненко А.Г. – К: Наукова думка, 1982. – 296 с.