

Державний вищий навчальний заклад
«Українська академія банківської справи Національного банку України»

АВТОМАТИЗАЦІЯ ПРОЦЕСУ ПРОГНОЗУВАННЯ БІРЖОВОЇ ІНФОРМАЦІЇ
НА ОСНОВІ НЕЙРОМЕРЕЖЕВОГО МОДЕЛЮВАННЯ

Конкурсна робота Всеукраїнського
конкурсу студентських наукових робіт
з природничих, технічних та гуманітарних наук у 2011/2012 навчальному році
за напрямом «Інформатика, обчислювальна техніка та автоматизація»

Автор

студент групи МЕК-12
спеціальності 8.03050201
«Економічна кібернетика»
Наливайко Віталій Юрійович

Науковий керівник

асистент кафедри
економічної кібернетики
Сергієнко Олексій
Володимирович

ЗМІСТ

ВСТУП	3
1 ОБСТЕЖЕННЯ ОБ'ЄКТА АВТОМАТИЗАЦІЇ ТА ФОРМУВАННЯ ВИМОГ ДО СИСТЕМИ	4
1.1 Характеристика задачі, що потребує автоматизації.....	4
1.2 Аналіз стану автоматизації в даній галузі та на об'єкті впровадження ...	6
1.3 Формування вимог до системи	8
2 РОЗРОБКА ПРОЕКТУ АВТОМАТИЗОВАНОЇ СИСТЕМИ	10
2.1 Архітектура та технології вирішення задачі	10
2.2 Склад функціональної частини	11
2.3 Склад підсистем забезпечення функціональної частини.....	13
3 РЕАЛІЗАЦІЯ ПРОТОТИПУ АВТОМАТИЗОВАНОЇ СИСТЕМИ.....	15
3.1 Структура та особливості реалізації інформаційного забезпечення	15
3.2 Структура та особливості реалізації алгоритмічного забезпечення.....	16
3.3 Інтерфейс користувача та інструкція по використанню	18
3.3.1 Робота з інтерфейсом на етапі завантаження даних	19
3.3.2 Модуль підготовки даних	20
3.3.3 Модуль прогнозування.....	25
ВИСНОВКИ.....	29
ПЕРЕЛІК ПОСИЛЕНЬ	30
ДОДАТКИ.....	31

ВСТУП

Прогнозування – одна із самих необхідних, але при цьому і самих складних задач аналізу. Проблеми при її рішенні обумовлені багатьма причинами – недостатня якість і кількість вихідних даних, мінливість середовища, в якому протікає процес, вплив суб'єктивних факторів. Але саме якісний прогноз є ключем до рішення бізнес – задач.

Нейромережеве моделювання знаходить все нові успішні застосування в практиці управління та прийняття рішень, в тому числі – у фінансових і торгових сферах. Теорія нелінійних адаптивних систем, яка лежить в основі нейромережевого моделювання, довела свою корисність при розробці прогнозів в багатьох галузях економіки та фінансів.

Метою даної роботи є розробка прототипу автоматизованої системи прогнозування біржової інформації на основі кластерного аналізу та нейромережевого моделювання засобами інтелектуального аналізу даних Microsoft SQL Server 2008 Analysis Services.

Об'єктом дослідження в даній роботі є процес прогнозування біржової інформації.

Предметом дослідження є методи та програмні засоби підтримки процесу прогнозування біржової інформації щодо забезпечення більш ефективних результатів у процесі прийняття майбутніх управлінських рішень.

1 ОБСТЕЖЕННЯ ОБ'ЄКТА АВТОМАТИЗАЦІЇ ТА ФОРМУВАННЯ ВИМОГ ДО СИСТЕМИ

1.1 Характеристика задачі, що потребує автоматизації

Прогнозування фінансових часових рядів – необхідний елемент будь-якої інвестиційної діяльності. Сама ідея інвестицій – вкладання фінансових коштів із метою отримання прибутку в майбутньому – ґрунтується на ідеї прогнозування майбутнього. Відповідно, прогнозування фінансових часових рядів лежить в основі діяльності всієї індустрії інвестицій – всіх бірж і небіржових систем торгівлі цінними паперами [1].

Наведемо декілька цифр, які ілюструють масштаб цієї індустрії для прогнозування. Щоденний обіг ринку акцій тільки в США перевищує 10 млрд. доларів. Депозитарій DTC (Depository Trust Company) в США, де зареєстровано цінних паперів на суму 11 трлн. доларів (із загального обсягу 18 трлн. доларів), реєструє в день операцій, приблизно на 250 млрд. доларів. Ще більш активно йде торгівля на світовому валютному ринку FOREX. Його добовий обіг перевищує 1000 млрд. доларів. Це, приблизно 1/50 всього сукупного капіталу людства.

Відомо, що 99% всіх цих угод – спекулятивні. Всі вони засновані на прогнозах зміни курсу їх учасниками. При цьому передбачення учасників кожної угоди протилежні один одному. Так що обсяг спекулятивних операцій характеризує рівень відмінностей в прогнозах учасників ринку, тобто реально – рівень непередбачуваності фінансових часових рядів. Таким чином на перший план виходить рішення задачі вибору ефективних методів їх прогнозування, оскільки, той хто володіє кращими математичними методами вилучення закономірностей із зашумлених хаотичних рядів, може сподіватися на більшу норму прибутку – за рахунок своїх менш оснащених опонентів.

В останнє десятиліття спостерігається стійке зростання популярності технічного аналізу – набору емпіричних правил, заснованих на різномірних

індикаторах поведінки ринку. Технічний аналіз зосереджується на індивідуальній поведінці даного фінансового інструмента, поза його зв'язками з іншими цінними паперами.

Згідно з основною доктриною технічного аналізу, передбачення повинні бути засновані на алгоритмі, тобто їх можна і навіть необхідно доручити комп'ютеру. За людиною залишається лише створення цього алгоритму, для чого в її розпорядженні є численні програмні продукти, що полегшують розробку і подальший супровід комп'ютерних стратегій на базі інструментарію технічного аналізу.

Слідуючи цій логіці, чому б не використовувати комп'ютер і на етапі розробки стратегії, причому не в якості асистента, який розраховує відомі ринкові індикатори і тестуючого задані стратегії, а для витягання оптимальних індикаторів і знаходження оптимальних стратегій за знайденими індикаторами. Такий підхід – із залученням технології нейронних мереж – завойовує з початку 90-х років XIX століття все більше прихильників оскільки має низку незаперечних переваг.

Нейромережевий аналіз, на відміну від технічного, не припускає ніяких обмежень на характер вхідної інформації. Це можуть бути як індикатори даного часового ряду, так і відомості про поведінку інших ринкових інструментів. Недарма нейромережі активно використовують саме інституційні інвестори (наприклад, великі пенсійні фонди), що працюють з великими портфелями, для яких особливо важливі кореляції між різними ринками.

На відміну від технічного аналізу, заснованого на загальних рекомендаціях, нейромережі здатні знаходити оптимальні для даного інструмента індикатори і будувати за ними оптимальну, знову ж для даного ряду, стратегію передбачення. Більш того, ці стратегії можуть бути адаптивними, змінюючись разом з ринком, що особливо важливо для молодих, активно розвиваючихся ринків [4].

Нейромережеве моделювання в чистому вигляді базується лише на даних, не привертаючи ніяких апріорних міркувань. У цьому його сила й одночасно – «ахіллесова п'ята». Наявних даних може не вистачити для навчання, розмірність потенційних входів може виявитися занадто високою.

Все це свідчить про перспективність нейромережевого трейдингу, принаймні на «коротких» часових проміжках. Більш того, в силу подібності фінансових часових рядів, норма прибутку за одиницю часу буде тим вище, чим менше характерний час трейдингу. Таким чином, автоматичні нейромережеві трейдери виявляються більш ефективними при торгівлі в реальному часі, де якраз найбільш помітні їх переваги над звичайними брокерами: невтомність, несхильність до емоцій, потенційно більш висока швидкість реагування. Навчена нейронна мережа, під'єднана до електронної системи торгів, може приймати рішення ще до того, як брокер-людина встигне розпізнати зміни графіка котирувань на своєму терміналі.

1.2 Аналіз стану автоматизації в даній галузі та на об'єкті впровадження

На даний момент на ринку присутній ряд готових систем, розроблених найбільшими фірмами-виробниками в області аналітичного та статистичного програмного забезпечення. Серед них є певна усталена градація, виходячи з такого параметра, як розмір, функціональність продукту.

За результатами останнього дослідження компанії Reuters, аналітичні пакети піддані наступній градації:

а) Пакети «для новачків» містять мінімум функціонала, налаштування якого, як правило, залишаються всередині системи і користувач не має доступу до них. Цей клас систем є скоріше ознайомчим, ніж придатним для справжньої роботи.

б) «Легкі» пакети характеризуються обмеженим числом методів, але користувач системи має доступ до більшості налаштувань. Не менш істотним

є й такий параметр як візуалізація, так як вона починає зароджуватися саме в цьому класі (Brain Maker Pro, Neuro).

в) «Середні» системи, які відрізняються від «легких» пакетів більш значним набором методів (Neuro Shell Trader, Deductor, PolyAnalyst) [3,5,6].

г) «Важкі» системи, як правило, це системи дослідного типу, так як для успішного їх застосування необхідні специфічні знання не тільки прикладної області, а й математичних методів, реалізованих у системі. До цього класу належать: Statistica, MatLab. Дані пакети більш складні у використанні, оскільки вони є професійно спрямованими, а їх використання вимагає спеціальних знань від користувачів. Дані пакети найчастіше використовуються в наукових дослідках та з метою навчання [3].

Наведемо діаграму, яка демонструє, що «важкі» системи практично не використовуються в реальних завданнях (рис. 1.2). Що пояснюється високими вимогами до обчислювальних потужностей. Технологічно процес виглядає так: на важких системах розробляється той або інший підхід, який надалі набуває комерційного образу у вигляді більш легких пакетів.

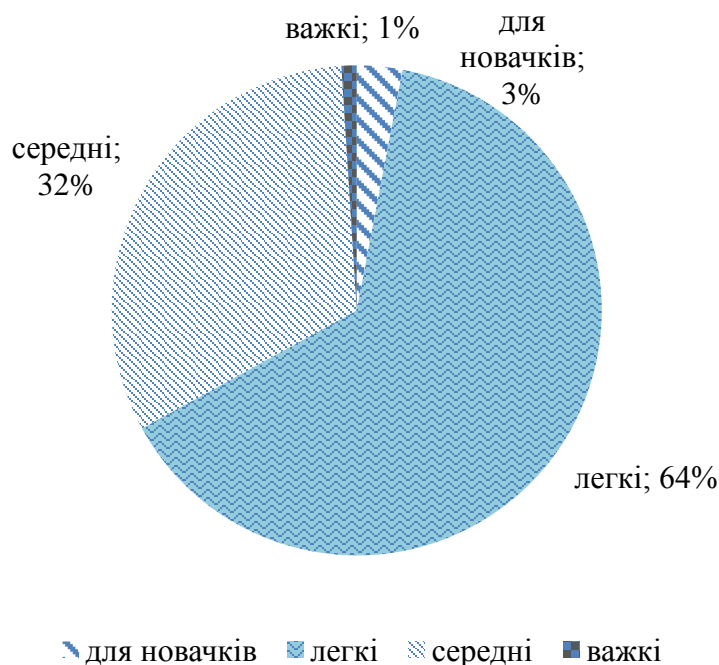


Рисунок 1.2 – Розподіл ринку на класовому рівні

Домінування «легких» пакетів обумовлено тим, що, по-перше, користувачі, не являються великими фахівцями в галузі математичного налаштування методів, по-друге, ці пакети є проміжною частиною рішення, і від них вимагається максимальна продуктивність.

1.3 Формування вимог до системи

Головною метою розробки аналітичної платформи є автоматизація процесу прогнозування економічної інформації (біржових котирувань) на основі застосувань алгоритмів кластеризації і нейронних мереж, реалізованих у службах Microsoft SQL Server 2008 Analysis Services (SSAS).

Вибір даних для навчання мережі та їх обробка є найскладнішим етапом рішення задачі методами нейромережевого моделювання. Набір даних для навчання повинен задовольняти декільком критеріям:

- репрезентативність, тобто дані повинні ілюструвати справжній стан речей в предметній області;
- несуперечність, так як суперечливі дані у навчальній вибірці призведуть до поганої якості навчання мережі;
- обсяг, так як існує думка, що кількість записів у вибірці повинна перевищувати кількість зв'язків між нейронами в мережі на кілька порядків;
- корелювання, так як існує думка, що чим вище ступінь кореляції між входами і виходами нейронної мережі, тим якісніше буде виконано прогноз.

Так, як найважливішим етапом роботи з програмою нейромережевого моделювання є очищення даних, що гарантує якісний прогноз, то розроблювана система повинна вміти відкидати первісні дані, які мають погрішності (екстремальні викиди або пусті значення) [4].

Після завантаження вхідних даних користувач повинен мати змогу отримати деяку статистику, яка дасть первісне представлення про завантажені дані. Також має бути передбачена можливість графічного представлення числових рядів у вигляді двовимірних діаграм.

Після завершення підготовки даних, система самостійно повинна створювати таблиці, що відповідають основному логічному переходу на етапі підготовки даних.

Розроблювана автоматизована система відноситься до класу «легких» систем, так як містить лише одну методологію, але при цьому дає достатній спектр налаштувань, закладених у ній методів.

Очевидно, що цільова аудиторія включатиме в себе малі банки і приватних інвесторів. Також проект може зацікавити тих користувачів, які займаються вивченням ринку.

Даний програмний продукт повинен мати інтуїтивно зрозумілий інтерфейс і враховувати той факт, що кінцевий користувач не має поглиблених знань в області кластеризації і нейронних мереж.

Аналітична платформа повинна мати наступні переваги і властивості, які дозволять зайняти їй місце «під сонцем»:

- простота у використанні, бо саме простота роботи з настільки складними методами інтелектуального аналізу даних є одночасно і властивістю і перевагою розроблюваного програмного продукту;

- швидкість отримання результатів, так як в даній прикладній області немає второваних стежок, то рішення знаходиться ітеративно, і висока швидкість отримання прогнозів, безумовно, сподобається користувачам;

- зрозумілість процесу, тому що розуміння користувачем технологічного процесу є перевагою над звичайними нейропакетами, так як користувач позиціонував нейронну мережу як «чорний ящик», тепер же він зрозуміє як працювати з цим ящиком;

- «Know-How» – раніше подібне поєднання настільки ефективних методів Data Mining не існувало, а отже, користувач отримує в свої руки саму останню розробку.

2 РОЗРОБКА ПРОЕКТУ АВТОМАТИЗОВАНОЇ СИСТЕМИ

2.1 Архітектура та технології вирішення задачі

Будь-яка інформаційна система повинна мати мінімум три основні функціональні частини – модулі зберігання даних, їх обробки і інтерфейсу з користувачем. Кожна з яких може бути реалізована незалежно від двох інших.

Виходячи із вимог до розроблюваної автоматизованої системи найбільш доцільним є використання клієнт-серверної архітектури, що дозволить три основні функціональні частини розподілити по двох фізичних модулях. Зазвичай програмне забезпечення, що відповідає за зберігання даних, розташовується на сервері (наприклад, сервері бази даних), інтерфейс з користувачем – на стороні клієнта, а обробку даних доводиться розподіляти між клієнтською і серверною частинами [7].

Таким чином автоматизована система включатиме в себе з клієнтську частину (аналітичну платформу) і серверну частину у вигляді двох сервісів MS SQL Server 2008:

- а) DB Engine (збереження даних);
- б) Analysis Services (методи DataMining).

Архітектура автоматизованої системи зображена на рисунку А.1:

У серверну частину системи винесені, головним чином, функції доступу до даних, а всі або більша частина прикладних обчислень виконуються клієнтською частиною. Це означає, що сервер тільки відбирає потрібні дані і пересилає їх на комп'ютер конкретного робочого місця для обробки. Якщо результати обробки повинні бути збережені в загальній базі даних, то вони пересилаються тому серверу, який і виконує ці функції [2].

Дана автоматизована система заснована на технології клієнт-сервер, використовуючи модель «товстого клієнту» (рисунок А.2):

Основна технологія, на якій базується робота аналітичної платформи – служба інтелектуального аналізу даних Microsoft SQL Server 2008 Analysis Services [8].

2.2 Склад функціональної частини

Функціональну частину автоматизованої системи прийнято поділяти на підсистеми програм для реалізації конкретних функцій системи. Автоматизована система складається з трьох модулів:

- Модуль завантаження даних.
- Модуль підготовки даних.
- Модуль прогнозування.

Схема взаємодії модулів зображена на рисунку А.3:

Модуль завантаження даних призначений для завантаження і зберігання біржової інформації, із засобів звернення до DB Engine через SQL-запити.

Вхідною інформацією для аналітичної платформи є файли формату «CSV», сформовані програмним продуктом Meta Trader 4. За кілька років свого існування, торгова платформа Meta Trader 4 завоювала величезну популярність серед брокерів і трейдерів. Даний продукт використовують понад 200 брокерських компаній і банків зі всього світу.

Так як вихідна інформація зберігається в розрізненних CSV-файлах, то основними завданнями даного модуля є:

- а) створення з'єднання з SQL Server;
- б) завантаження в БД вихідних даних (біржовий інформації);
- в) відображення інформації в графічному і текстовому вигляді.

Модуль підготовки даних реалізує функції очищення, доповнення і т.п. за допомогою звернення до DB Engine через SQL-запити.

У цьому модулі відбувається спочатку перехід від таблиці DOHLCV до таблиці DV, а потім перехід до остаточної таблиці, яка буде представляти

вихідну інформацію для модуля прогнозування. Тому, основні завдання модуля:

- а) перетворення формату DOHLCV до формату DV (Date; Value);
- б) очищення, доповнення, заповнення прогалін і прирощення DV;
- в) обчислення кореляції;
- г) нормалізація даних;
- г) створення таблиць вхідних даних для Data Mining.

Модуль прогнозування призначений для створення моделі інтелектуального аналізу даних на базі алгоритмів кластеризації і нейронних мереж та отримання прогнозу. Модуль пов'язаний з DB Engine, за допомогою SQL-запитів, і Analysis Services, за допомогою DMX-запитів.

Даний модуль на основі створеної узагальненої таблиці реалізує кластерний аналіз. Після того як будуть отримані правила, що визначають сегментацію даних, для кожного кластера створюється нейронна мережа.

Після цього нейронні мережі і правила сегментації зберігаються і коли необхідно отримати прогноз, з цього модуля запитуються вхідні дані (формат CSV). До цих даних застосовується правило сегментації, відповідно до того, до якого сегменту віднесена ця точка, вибирається відповідна нейронна мережа.

Розглянемо більш детально бізнес-логіку роботи АС. Для вирішення задач функціонального моделювання, тобто для опису існуючих процесів або процесів, які ми намагаємося отримати в ідеалі, широко використовуються методологія структурного аналізу і проектування (Structured Analysis and Design Technique – SADT).

SADT-методологія передбачає побудову ієрархічної системи діаграм – одиничних описів фрагментів системи. При цьому спочатку виконується опис системи в цілому, та її взаємодія з навколишнім світом (контекстна діаграма), після чого виконується функціональна декомпозиція – система розбивається на підсистеми і кожна підсистема описується окремо (діаграми декомпозиції) [9].

Контекстна діаграма процесу Прогнозування біржової інформації, за допомогою прототипу АС, а також його декомпозиція представлені у додатку Б.

У процесі проектування інформаційної системи досить важливим етапом є документування механізмів передачі і обробки інформації в модельованій системі, для вирішення даної задачі на практиці широко використовуються так звані діаграми потоків даних (Data Flow Diagram). Дана нотація є діаграмою потоків даних. Діаграми DFD зазвичай будуються для наочного відображення поточної роботи системи документообігу організації. Найчастіше діаграми DFD застосовують як доповнення моделі бізнес-процесів, виконаної в IDEF0 [9].

Декомпозиція процесів Кластерний аналіз та Нейромережеве моделювання до DFD діаграм приведена в додатку В.

На етапі прогнозування подається набір конкретних значень входів, який проходить через сформовані правила, отримані на етапі кластеризації. Так визначається, до якого кластеру відноситься цей набір конкретних входів (точка, для якої здійснюється прогноз). Після визначення кластеру цій точці співвідноситься одна з натренованих нейронних мереж, за допомогою якої й буде зроблено прогноз. Одночасно з цим здійснюється прогноз для цієї ж точки і через нейронну мережу, яка побудована і натренована на всій навчальній вибірці.

Очевидно, що кластерний аналіз може в деяких випадках виявитися шкідливим, тому прогноз і робиться не тільки на кластері, але і на всій вибірці для порівняння результатів.

Розглянемо декомпозицію процесу Прогнозування до DFD діаграми (див. додаток Г).

2.3 Склад підсистем забезпечення функціональної частини

Що стосується програмного забезпечення функціональної частини системи, то для налаштування робочого місця користувача та забезпечення

коректної роботи аналітичної платформи необхідно створити умови для того щоб можна було встановити Microsoft SQL Server 2008 Analysis Services.

Для коректної роботи розроблена аналітична платформа потребує Windows – платформу з налаштованим Microsoft SQL Server 2008 з можливістю виконання операцій бізнес-аналітики (Enterprise або Standard Edition), налаштований ADO-канал і програмна платформа .NET Framework версії 3.5 з пакетом оновлення SP1.

Коректне та комфортне функціонування розробленої аналітичної платформи забезпечить відповідне апаратне забезпечення – комплекс технічних засобів, який включає ЕОМ: зовнішні пристрої, термінали, абонентські пункти тощо, які необхідні для функціонування тієї чи іншої системи.

Розроблювана автоматизована система є засобом, який дозволяє користувачам з різним рівнем кваліфікації в області математики та статистики швидко і зручно використовувати засоби бізнес-аналітики пакету Microsoft SQL Server 2008 Analysis Services для аналізу і прогнозування біржових цін. Тому, апаратні вимоги для серверу визначаються лише програмним пакетом підтримки, тобто Microsoft SQL Server 2008 [8].

3 РЕАЛІЗАЦІЯ ПРОТОТИПУ АВТОМАТИЗОВАНОЇ СИСТЕМИ

3.1 Структура та особливості реалізації інформаційного забезпечення

Розглянемо більш детально структуру збереження інформації, необхідної для функціонування автоматизованої системи.

Структуру бази даних аналітичної платформи, а також структура і призначення кожної з таблиць БД представлені в додатку Д.

На рисунку 3.1 зображена загальна схема взаємодії модулів прототипу автоматизованої системи через таблиці БД.

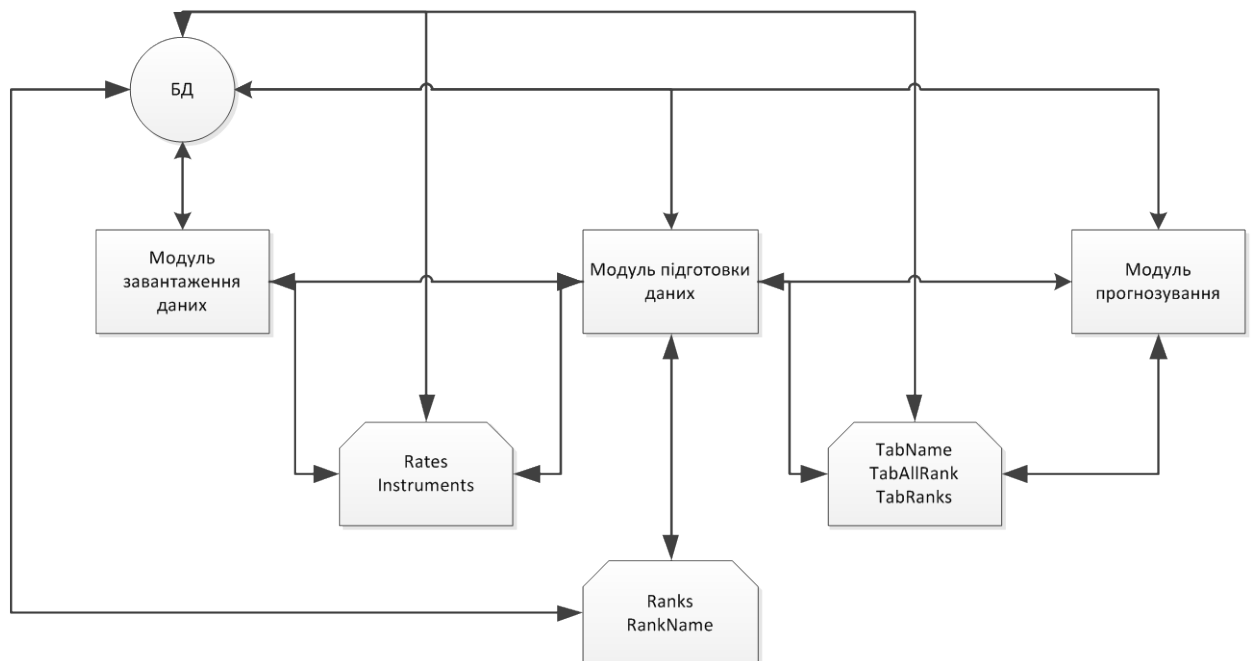


Рисунок 3.1 – Схема взаємодії модулів АС через таблиці БД

Таблиці Instruments і Rates є внутрішніми таблицями Модуля Завантаження і використовуються для інформаційного зв'язку з Модулем Завантаження.

Таблиці RankNames & Rank є внутрішніми таблицями Модуля Завантаження.

Таблиці TabName, TabAllRanks і TabRanks є таблицями зв'язку Модуля Завантаження та Модуля Прогнозування, так як вони спеціально створюються в цьому модулі і ніяких перетворень над ними не відбувається.

3.2 Структура та особливості реалізації алгоритмічного забезпечення

Робота системи буде представлена 3 етапами:

- кластерний аналіз;
- нейромережеве моделювання;
- прогнозування.

Вихідні дані, сформовані програмним продуктом MetaTrader 4.0, повинні бути завантажені в базу даних MS SQL Server. Далі вибирається період часу, за який буде проводитися аналіз, виконуються перетворення вихідних даних, формування даних у вихідну таблицю для Data Mining. На вихідній таблиці інструментів запускається алгоритм кластеризації, який розбиває вибірку на кластери. На кожному із кластерів і на загальній вибірці будуються нейронні мережі.

Результатом роботи алгоритму будуть прогнози майбутніх значень для вибраних, як прогнозованих, інструментів. Схема послідовності роботи алгоритмів зображена на рисунку 3.3:



Рисунок 3.3 – Послідовність роботи алгоритмів

На першому етапі (Кластерний аналіз) навчальна вибірка піддається кластерному аналізу. В результаті формується набір кластерів і описуючих їх правил.

Алгоритм кластеризації (Microsoft) представляє собою алгоритм сегментації, що надається службами Microsoft SQL Server 2008 Analysis Services (SSAS). Алгоритм використовує ітераційні методи для групування варіантів в набори даних у кластерах, що містять подібні характеристики, а саме: алгоритм К-середніх та алгоритм максимізації очікувань. За замовчуванням, моделі кластеризації Microsoft використовують алгоритм масштабованої максимізації очікування. Цей алгоритм використовується за замовчуванням, оскільки він володіє декількома перевагами в порівнянні з методом кластеризації К-середніх:

- не потребує більше одного перегляду бази даних;
- працює навіть при обмеженому об'ємі оперативної пам'яті;
- по продуктивності випереджає методи, що вимагають вибірки.

Графічна реалізація EM алгоритму представлена у додатку Е [10].

На другому етапі (Нейромережеве моделювання) для кожного кластера створюється і навчається нейронна мережа. Також, для порівняння результатів створюється і навчається нейронна мережа для всієї вибірки котирувань.

Алгоритм нейронної мережі (Microsoft) використовує мережу у вигляді багатошарового перцептрона, також відому під назвою мережа дельта-правила зі зворотним поширенням, до складу якої може входити до трьох шарів нейронів, або перцептрони. Такими шарами є вхідний шар, необов'язковий прихований шар і вихідний шар.

Блок-схема повного алгоритму навчання нейромережі матиме вигляд (див. додаток Є).

Перед безпосередньою роботою із модулями кластерного аналізу та нейромережевого прогнозування необхідно привести вхідні дані до «нормального», цілісного вигляду. Для рішення даної задачі, в автоматизованій системі реалізовано очистка та нормалізація вхідного числового ряду.

Для реалізації функції очищення числового ряду використовується ковзна середня SMA (Simple Moving Average). Особливістю є той факт, що спочатку використовується два періоди SMA. Це обумовлено тим, що SMA має період, що визначає розрахункову базу, і, як наслідок, в рамках цього періоду він не може бути обчислений для перших значень. Тому будується протилежно спрямований йому другий SMA, з якого й беруться відсутні значення (формула 3.1):

$$SMA = \frac{\sum_{i=n}^n P_i}{n}, \quad (3.1)$$

де P_i – ціна на ринку (зазвичай Close, Open, High, Low, Median Price);

n – основний параметр, довжина згладжування або період ковзної середньої.

Для здійснення очищення числового ряду, необхідно вибрати період SMA і процентний коридор очищення, який визначить верхню і нижню границю (формули 3.2 – 3.3) [1]:

$$High = SMA \cdot (1 + \%) \quad (3.2)$$

$$Low = SMA \cdot (1 - \%) \quad (3.3)$$

Блок-схему алгоритму очищення числового ряду представлено в додатку Ж.

3.3 Інтерфейс користувача та інструкція по використанню

Для запуску програми необхідно двічі клацнути на файлі RatePrediction.exe, виконати з'єднання із базою і розпочати роботу із системою.

Після заповнення всіх необхідних полів необхідно натиснути на кнопку «Підключитися» (рис. 3.1).

Після вдалого підключення до бази даних в рядку стану RatePrediction, який знаходиться внизу, можна побачити повідомлення «З'єднання

встановлено» і відбудеться перехід на наступну по порядку вкладку системи – Інструменти.

Розглянемо інструкцію по використанню та інтерфейс аналітичної платформи на кожному з трьох етапів, відповідно до розроблених модулів системи: завантаження даних, підготовка даних та прогнозування.

Усі переходи між модулями (вкладками) і в середині одного модуля аналітичної системи здійснюються таким чином. Внизу основного діалогового вікна, знаходяться декілька навігаційних кнопок системи – «Вибрати», «Далі» і «Назад». У центральній області основного діалогового вікна аналітичної системи спеціалісту надається навігація між основними можливостями конкретних модулів (вкладок). Вибір потрібної опції конкретного модуля системи відбувається за допомогою «RadioButton» компонентів. Потрібно відзначити потрібну опцію, після чого натиснути кнопку «Вибрати». Для повернення до вибору опцій і навігації між модулями необхідно натиснути кнопку «Назад». Для навігації між модулями використовуються кнопки «Далі» або «Назад».

3.3.1 Робота з інтерфейсом на етапі завантаження даних

Вкладка «Інструменти» дозволяє завантажити в базу нові дані у форматі DONLCSV, отримані за допомогою програмного продукту MetaTrader з інтернету. Також на цій вкладці доступний перегляд раніше завантажених даних та їх видалення з бази (рис. 3.2):

Після вибору пункту на вкладці «Інструменти» і натиснення кнопки «Вибрати» відбудеться перехід до вкладки «Завантаження в БД» (рис. 3.3).

Для завантаження нових котирувань необхідно натиснути на кнопку «Завантажити». Далі вибрати потрібний .CSV файл і натиснути кнопку «Відкрити». Після цього, з'явиться діалогове вікно, де потрібно вказати назву інструменту (далі ця назва буде фігурувати при будь-яких маніпуляціях із завантаженим поруч).

Після завантаження файлів в лівому «CheckedListBox» (Інструменти) відображаються доступні для вибору інструменти. Вибираючи той чи інший

інструмент, у правому «ListBox» відображаються поля Date (+ Time), Open, High, Low, Close, Volume. Натискаючи на одне з котирувань, нижче групи кнопок можна побачити детальну інформацію, що включає в себе: дату, ціну відкриття, найбільшу ціну, найменшу ціну, ціну закриття, значення.

Кнопка «Зберегти в БД» зберігає завантажену біржову інформацію, взяту з файлу *.CSV у форматі DOHLCV в таблицю «Rates».

Кнопка «Видалити» видаляє обрані в «CheckedListBox» (Інструменти) інструменти.

Під полем «Інструменти» відображається інформація про завантажені дані: усього рядків; Max, Min Open; Max, Min High; Max, Min Low; Max, Min Close; Max, Min Volume.

Візуалізація нових котирувань проводиться на вкладці «Графіки». У лівому «CheckedListBox» (Інструмент) відображаються доступні інструменти (нові інструменти стають доступними для візуалізації тільки після завантаження в БД), графіки яких можуть бути побудовані.

Для можливості порівняння, робоче поле розділене на дві частини, в кожній з яких користувач може побудувати графік окремого інструменту (рис. 3.4):

Переглянути завантажені в БД інструменти можна на вкладці «Перегляд і видалення інструментів» (рис. 3.5). У полі «Інструменти торгової площадки» відображаються інструменти, які знаходяться в БД у даний момент:

При виборі одного з інструментів у вікні праворуч відображаються його значення у форматі DOHLCV.

Для вибору всіх наявних інструментів необхідно поставити галочку «Всі».

Для очищення бази даних необхідно виділити всі завантажені інструменти і натиснути кнопку «Видалити».

3.3.2 Модуль підготовки даних

За допомогою цього модуля вирішується ряд важливих завдань:

- здійснюється перехід від формату DOHLCV до формату DV;
- здійснюється очищення числових рядів;
- створюються збільшення;
- обчислюється кореляція між рядами DV, тим самим спрощується вибір найкращих входів;
- проводиться нормалізація рядів DV;
- створюється остаточна структура, яка використовується на етапах кластерного і нейромережевого аналізу.

Спочатку дані надходять у систему у вигляді CSV-файлів, які у свою чергу зберігають дані у форматі Date-Time; Open; High; Low; Close; Volume. Так як розроблювана система передбачає оперування поняттям числовий ряд, то було вирішено використовувати формат Date; Value (Close).

Даний перехід здійснюється на вкладці «Створення/Видалення колонок». Для переходу необхідно вибрати відповідний пункт на вкладці «Підготовка даних » і натиснути кнопку «Вибрати» (рис. 3.6).

У лівому CheckedListBox (Інструмент) відображаються доступні ряди, що зберігаються в БД в таблиці dbo.Rates (формат DOHLCV). Вибираючи той чи інший ряд, в ListBox (Числові ряди) відображаються поля Date (+ Time), Open, High, Low, Close, Volume.

Для перетворення завантажених в БД рядів у форматі DOHLCV до рядів формату DV необхідно вибрати один з інструментів, після чого нижче в декількох ComboBox компонентах буде доступний вибір діапазонів дати і часу. Також доступний вибір поля, яке займе місце Value в новому ряді формату DV.

Після натискання на кнопку «Зберегти ч. ряд» з'явиться діалогове вікно, в якому буде запропоновано вказати ім'я створюваного ряду. Після вибору імені нового ряду в форматі DV, дані з вибраного в CheckedListBox (Інструмент) ряду DOHLCV будуть відфільтровані відповідно до зазначеного діапазону дати і часу, далі з нього буде вибрано поле, яке займе місце Value в новому ряді. У результаті всіх операцій новий ряд у форматі DV буде

завантажений в БД в таблицю `dbo.Ranks` і відобразиться в нижньому `CheckedListBox`.

Очищення і нормалізація числових рядів відбувається на вкладці «Створення очищеної колонки» (рис. 3.7):

Робота починається з вибору числового ряду після натискання кнопки «Вибрати ч. ряд ». Після натискання на кнопку відкривається діалогове вікно, в якому пропонується вибрати потрібний числовий ряд. Після успішного вибору необхідного числового ряду, його значення відобразяться в нижньому компоненті `DataGridView`. У цьому компоненті передбачені можливості примусової заміни конкретної позиції в числовому ряді або ж її повного видалення.

Далі необхідно вказати період `Simple Moving Average` та задати процентний коридор очищення.

Після натискання на кнопку «Виконати обробку» відбудеться автоматичне обчислення всіх точок, що вийшли за межі коридору. Точки, які вийшли за межу коридору виділяються кольором. Після чого експерт зможе самостійно оцінити важливість виділених точок, та зробити видалення або заміни значення в них.

Варто зауважити, що всі зміни числового ряду не будуть збережені до тих пір, поки не буде натиснута кнопка «Зберегти ч. ряд».

Для нормалізації числового ряду необхідно поставити відмітку у відповідному полі вкладки «Створення очищеної колонки». Далі нормалізація буде проводитися в складі процесу очищення ряду (рис. 3.7).

Можливість прогнозувати неабсолютні значення вкрай важлива, оскільки при схожій або навіть гіршій точності таких прогнозів загальна помилка на виході буде набагато меншою. Конкретний приклад. Припустимо, що ми хочемо за допомогою нейромережі прогнозувати вартість акції з похибкою 20%. Сьогодні акції коштують 95 дол., а завтра будуть коштувати 100. Що може бути спрогнозовано з більшою точністю абсолютна ціна акцій або зміна ціни? У першому випадку прогноз буде в

межах 80-120 дол., що явно представляє малий практичний інтерес. При прогнозі зміни ціни результат буде перебувати в інтервалі 4-6 дол., що відповідає абсолютній ціні 99-101 дол. і є більш прийнятним.

Прирощення числових рядів здійснюється на вкладці «Прирощення» (рис. 3.8).

На цій вкладці користувач має можливість перейти від абсолютних значень до приростів.

Вкладка має дві підвкладки: «Вихідний» і «Прирости» (рис. 3.9). Перша закладка відображає вихідний ряд, друга призначена для візуалізації результатів операції збільшення. Робота побудована за таким же принципом, що і на попередніх закладках.

За допомогою кнопки «Вибрати ч. ряд» вибирається потрібний ряд DV. Далі після натискання на кнопку «приростити», відбувається обчислення збільшень абсолютних значень котирувань вибраного ряду. Щоб зберегти зміни необхідно скористатися кнопкою «Зберегти ч. ряд».

Перегляд кореляції між раніше створеними числовими рядами DV можливий у вкладці «Перегляд кореляції двох колонок» (рис. 3.10). У ListBox «Перший числовий ряд» необхідно вибрати ряд, по відношенню до якого будуть обчислюватися коефіцієнти кореляції всіх інших рядів. Після вибору початкового ряду в DataGridView «Другий числовий ряд» будуть відображені обчислені коефіцієнти кореляції.

Реалізована можливість графічного виділення різних значень коефіцієнта кореляції по заданому граничному значенню.

Створити підсумкову таблицю для Data Mining з раніше сформованих числових рядів DV можливо у вкладці «Створення / Видалення таблиці даних » (рис. 3.11).

Дана вкладка призначена для об'єднання розрізнених числових рядів DV в кінцеву таблицю, яка буде подаватися на вхід алгоритмам кластеризації та нейронних мереж.

Робоче вікно поділено на дві області: ліва і права.

Ліва область – призначена для вибору ряду DV, ряд повинен бути названий і отримати одну з трьох можливих настройок, що використовуються для нейромережевого прогнозування (табл. 3.1):

Таблиця 3.1 – Опис типів полів

Тип	Опис
Вхід	Вхід нейронної мережі
Вихід	Вихід нейронної мережі. Якщо ряд позначається як вихід (виходом може бути помічений лише 1 ряд), то його поле Date береться для формованої таблиці.
Примітка	Стовпці, відмічені такою ознакою, не будуть використані під час роботи алгоритмів кластеризації нейромережевого моделювання.

Права область – призначена для перегляду вмісту створюваних або ж уже створених таблиць.

Для того щоб додати в таблицю раніше створений але зміщений ряд, необхідно вибрати даний ряд стандартним способом і далі скористатися кнопкою «Змістити». Додавання до таблиці зміщених рядів має істотне значення, так як вони найчастіше використовуються в якості виходів нейронної мережі, представляючи собою значення котирувань вибраного ряду «на один або кілька наступних врахованих днів». Таким чином, за допомогою зміщених рядів можливе навчання нейронної мережі та прогноз майбутніх котирувань за поточними значеннями інших котирувань.

Безпосереднє додавання ряду в таблицю відбувається наступним чином (рис. 3.11).

Вибираємо числовий ряд стандартним способом за допомогою кнопки «Вибрати ч. ряд». Обраний числовий ряд з'являється зліва знизу в DataGridView. Вибираємо одну з трьох можливих налаштувань для числового ряду. Якщо потрібно, вказуємо прирощення числового ряду.

За допомогою кнопки «Додати ряд =>» додаємо ряд в середнє праве поле, в якому будуть відображатися всі ряди, додані в таблицю.

Доданий числовий ряд з'являється праворуч знизу в DataGridView.

Перераховані вище операції повторюються необхідну кількість разів, щоб додати необхідну кількість рядів в таблицю.

Після того, як буде додано більше одного ряду, стане можливим використання кнопки «Обробити» в середньому правому полі (важливість використання даної кнопки пояснюється нижче).

Збереження таблиці здійснюється за допомогою кнопки «Зберегти як...» (буде необхідно задати ім'я нової таблиці), після збереження створена таблиця буде записана в таблицю `dbo.TabName` і додасться у верхнє праве поле.

Управління доданими рядами таблиць проводиться кнопками «Видалити», «Очистити» і «Обробити» в середньому правому полі. Видалення доданих таблиць проводиться кнопкою «Видалити» у верхньому правому полі.

Використання кнопки «Обробити» має велике значення для одержання прийнятних прогнозів. Розрізнені числові ряди, які додаються в таблицю, досить часто будуть мати різну кількість записів. Для коректного кластерного аналізу і нейромережевого прогнозування важливо, щоб таблиця з вихідними даними була «рівною», тобто кожен ряд у таблиці мав однакову кількість записів, з однаковими датами. Обробка доданих рядів вирішує цю проблему, урізаючи необхідні «надлишки».

3.3.3 Модуль прогнозування

На першому етапі прогнозування здійснюється:

- побудова моделі кластеризації;
- кореляційний аналіз даних в отриманих кластерах;
- візуалізація отриманих кластерів.

Вихідними даними для кластерного аналізу є таблиці, сформовані на етапі підготовки даних. Всі доступні таблиці відображаються в списку «Вибір таблиці прогнозування» на вкладці «Кластеризація» (рис. 3.12):

Вибір параметрів алгоритму здійснюється на вкладці «Параметри» (рис. 3.13) [8].

Після вибору таблиці і установки параметрів необхідно натиснути кнопку «Обчислити кластери» на вкладці «Кластеризація», після чого буде створена нова модель кластеризації, вона буде розгорнута і оброблена.

Отримані в результаті роботи алгоритму кластери відображаються в списку «Список кластерів».

Для перегляду результатів кореляційного аналізу необхідно обрати потрібний кластер зі списку. У правій формі «Кореляція Пірсона» відображається коефіцієнт кореляції по всій таблиці і кореляція за поточним кластером (рис. 3.12). Кореляція визначається коефіцієнтом Пірсона, застосовуваним для метричних величин.

У вікні «Опис кластера» відображаються правила розбиття вибраного кластера. У правій нижній частині форми відображається графік, на якому представлено графічне розбиття прогнозованого ряду на кластери (рис. 3.13).

Другий етап роботи з модулем прогнозування – це нейромережеве моделювання.

На основі збережених даних у пам'яті Analysis Services про отримані кластери та початкову навчальну вибірку створюються нейронні моделі. Кількість створених моделей буде рівною кількості кластерів, отриманих на етапі кластерного аналізу плюс одна модель без урахування розбиття на кластери (для всієї вибірки).

На вкладці «Нейромережеве моделювання» знаходяться дві підвкладки: «Нейромережеве моделювання» і «Параметри». На підвкладці «Нейромережеве моделювання» здійснюється безпосередній запуск алгоритму (рис. 3.14). Попередній вибір параметрів моделювання здійснюється на підвкладці «Параметри» (рис. 3.15). Так само на вкладці «Нейромережеве моделювання» є кнопка «Запам'ятати модель», при натисканні на яку відбувається збереження створених і навчених нейронних мереж, на етапі прогнозування можна буде їх використовувати (підкладка «Збережені моделі») в іншому ж випадку, прогноз буде будуватися на останній навченій мережі.

Після створення нейромережових моделей відбувається DMX-запит на їхнє навчання (для кожної моделі).

Оскільки дані в таблиці «Tab» не розкидані на кластери, то перед виведенням інформації відбувається створення тимчасової таблиці, яка містить у собі ключові дані для визначення того, які рядки в таблиці «dbo.Tab» відносяться до якого кластера, після чого у нас з'являється можливість написати запит в якому для навчання НС будуть надані дані тільки по одному кластеру.

Останній, третій етап роботи із даним модулем – це прогнозування.

Основним завданням даної підсистеми є отримання безпосередніх прогнозів, які можуть бути отримані на основі двох основних ідей:

- 1) НМ;
- 2) НМ + Кластеризація.

Підсистема прогнозування може генерувати одинарні та багаторядкові прогнози.

Для генерації одинарного прогнозу, тобто отримати прогноз на одне значення, ввівши необхідні для цього дані, необхідно скористуватися кнопкою «Спрогнозувати значення», яка знаходиться на вкладці «Одинарний прогноз» (рис. 3.16):

Після цього з'явиться діалогове вікно «Тіки», в яке потрібно ввести необхідні для прогнозу дані (тіки) у вказаному порядку (рис. 3.17).

Після введення всіх необхідних значень у полі «Спрогнозувати значення» відобразиться результат прогнозу двома методами (НМ, НМ + Кластеризація). Також у вкладці реалізована можливість графічного представлення результату прогнозування по відношенню до початкової вибірки вихідних значень, що використовується для тренування нейронної мережі і відображення кореляції в обраному кластері. Значення коефіцієнтів кореляції з'являються у відповідній формі після вибору якого-небудь з отриманих раніше кластерів у верхній лівій частині вікна. Графічне представлення результату прогнозування з'являється після натискання на

кнопку «Показати графік з новим значенням». Очищення графіка проводиться за допомогою кнопки «Очистити графік» (рис. 3.18):

Для отримання багаторядкового прогнозу необхідно переключитися на вкладку «Багаторядковий прогноз». Задання значень відбувається наступним чином. Для кожного разу задання нової вибірки необхідної для прогнозу значень слід використовувати кнопку «Додати». Після натискання кнопки у таблиці для введення з'являються нульові значення, а замість прогнозу у відповідній клітинці таблиці відображається «##». Для введення значень у клітинку таблиці необхідно клацнути на неї мишкою, після чого відкриється діалогове вікно, в якому необхідно ввести потрібне значення. Після задання всіх необхідних ячеек потрібно натиснути кнопку «Спрогнозувати», після чого в комірці прогнозу відобразиться результат прогнозу (рис. 3.19). Вибір між прогнозуванням методом «НМ» і методом «НМ + Кластеризація» здійснюється у відповідному «ListBox».

Іноді буває потрібно оцінити ступінь впливу різних параметрів нейронної мережі на результат прогнозу, для цього краще створити відразу кілька мереж з різними параметрами, для цього потрібно після запуску моделювання їх запам'ятати, скориставшись кнопкою «Запам'ятати модель» на етапі «Нейромережеве моделювання». Для можливості переключення між збереженими мережами була створена підвкладка «Збережені моделі». На цій вкладці зліва відображаються параметри мережі за яких вона була змодельована, справа ж – сама мережа (рис. 3.20).

ВИСНОВКИ

З кожним роком більш чітко прослідковується тенденція щодо збільшення інтересу використання нейронних мереж у математичному моделюванні, так як їх застосування відкриває нові можливості використання розрахунків у сферах, які раніше відносилися лише до області людського інтелекту. До однієї з таких задач можна віднести прогнозування майбутніх тенденцій зміни часових рядів. Враховуючи складність здійснення необхідних розрахунків, неабиякої актуальності набуває використання автоматизованих систем, при вирішенні задач такого характеру.

В процесі виконання роботи була досягнута її мета – створення прототипу автоматизованої системи прогнозування біржової інформації на основі нейромережевого моделювання.

Проаналізувавши стан вже існуючих інформаційних систем нейромережевого моделювання, слід відзначити, що для практичного використання в основному застосовуються легкі та середні автоматизовані системи, які мають невисокий рівень наочності представленої в результаті прогнозу інформації та дружності інтерфейсу. Прототип розроблюваної системи поєднує в собі підходи важких систем, але в той же час є досить легким у використанні і впровадженні.

Розроблений прототип автоматизованої системи є досить ефективним при торгівлі в реальному часі, оскільки автоматичні нейромережеві трейдери мають помітні переваги над звичайними брокерами, до яких можна віднести: невтомність, несхильність до емоцій, потенційно більш висока швидкість реагування. За допомогою даного прототипу автоматизованої системи можна приймати рішення ще до того, як людина-брокер встигне розпізнати зміни графіка котирувань на своєму терміналі.

Крім того, розроблений прототип автоматизованої системи є досить гнучким з точки зору можливості його вдосконалення і додання нових та оновлення вже існуючих функцій.

ПЕРЕЛІК ПОСИЛЕНЬ

1. Вітлінський В.В. Економічний ризик: ігрові моделі [Текст]: навч. посіб. / В.В. Вітлінський, П. І. Верченко, А.В. Сігал. – К. : КНЕУ, 2002. – 446 с. ISBN 966–574–318–X;
2. Мак-Леннан, Д. DataMining with Microsoft SQL Server 2008 [Текст] : учеб. пособие / Д. Мак-Леннан – М. : Вильямс, 2009. – 521 с. ISBN 946-374-306-9;
3. Research Software [Електронний ресурс], – Режим доступу: http://thomsonreuters.com/products_services/science/science_products/arz/research_software/, [Заголовок з екрану];
4. Предсказания финансовых временных рядов [Електронний ресурс], – Режим доступу: <http://articles.mql4.com/ru/542>, [Заголовок з екрану];
5. Обзор возможностей [Електронний ресурс], – Режим доступу: <http://www.neuroproject.ru/aboutproduct.php>, [Заголовок з екрану];
6. PolyAnalyst [Електронний ресурс], – Режим доступу: <http://megaruter.ru/polyanalyst.php>, [Заголовок з екрану];
7. Архитектура «клиент-сервер» [Електронний ресурс], – Режим доступу: http://www.mstu.edu.ru/study/materials/zelenkov/ch_7_1.html, [Заголовок з екрану];
8. Службы SQL Server Analysis Services — интеллектуальный анализ данных [Електронний ресурс], – Режим доступу: <http://technet.microsoft.com/ru-ru/library/bb510517.aspx>, [Заголовок з екрану];
9. Технология и организация практической деятельности в сфере бизнес-информатики [Електронний ресурс], – Режим доступу: <http://www.tstu.ru/education/elib/pdf/2010/duzenkova-a.pdf>, [Заголовок з екрану];
10. Base Group Labs. [Електронний ресурс], – Режим доступу: <http://www.basegroup.ru/library/analysis/clusterization/em/>, [Заголовок з екрану].

ДОДАТКИ

Додаток А

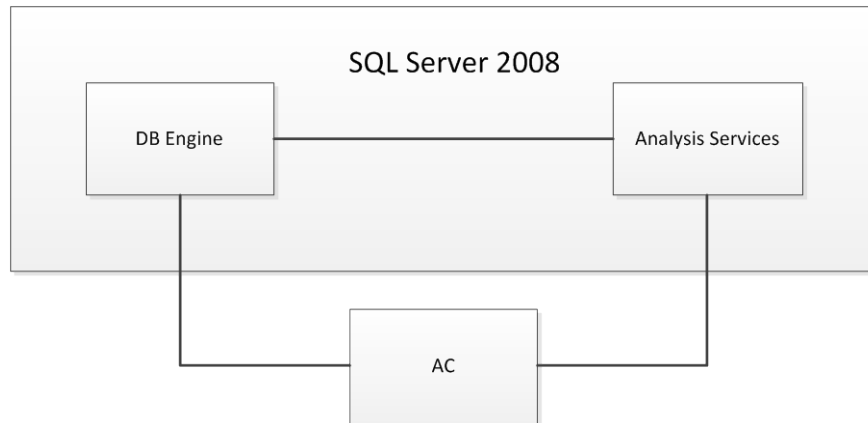


Рисунок А.1 – Архітектура автоматизованої системи

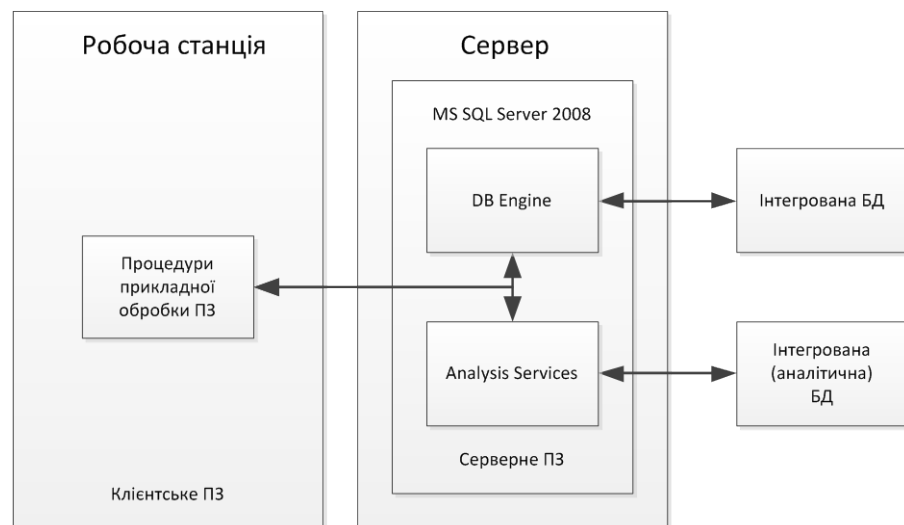


Рисунок А.2 – Модель «товстого клієнту»

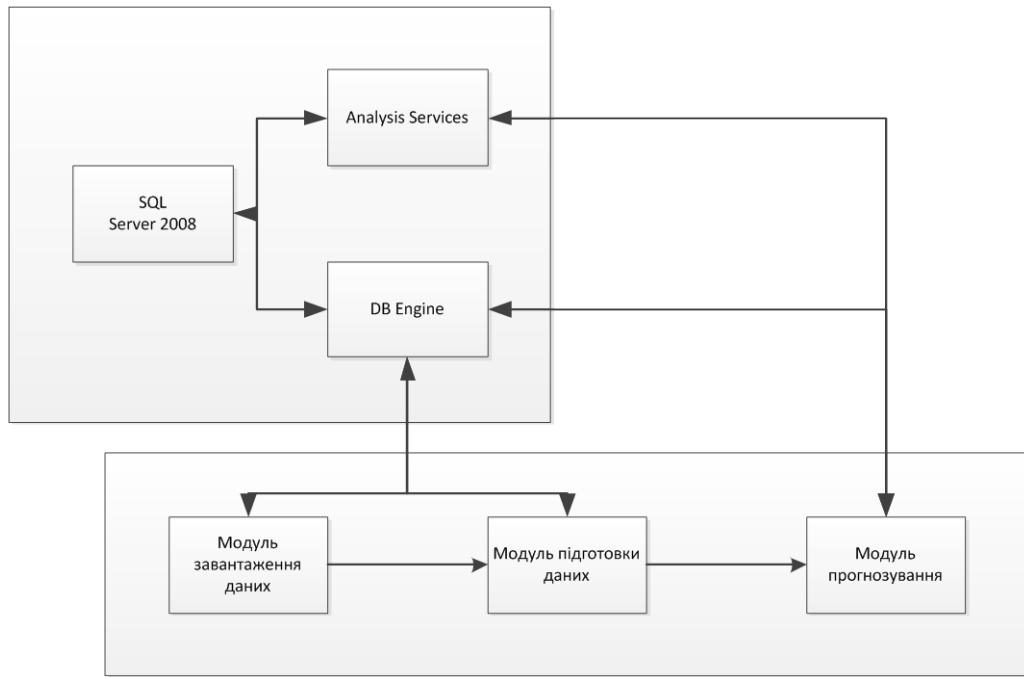


Рисунок А.3 – Схема взаємодії модулів

Додаток Б

Декомпозиція процесу Прогнозування біржової інформації

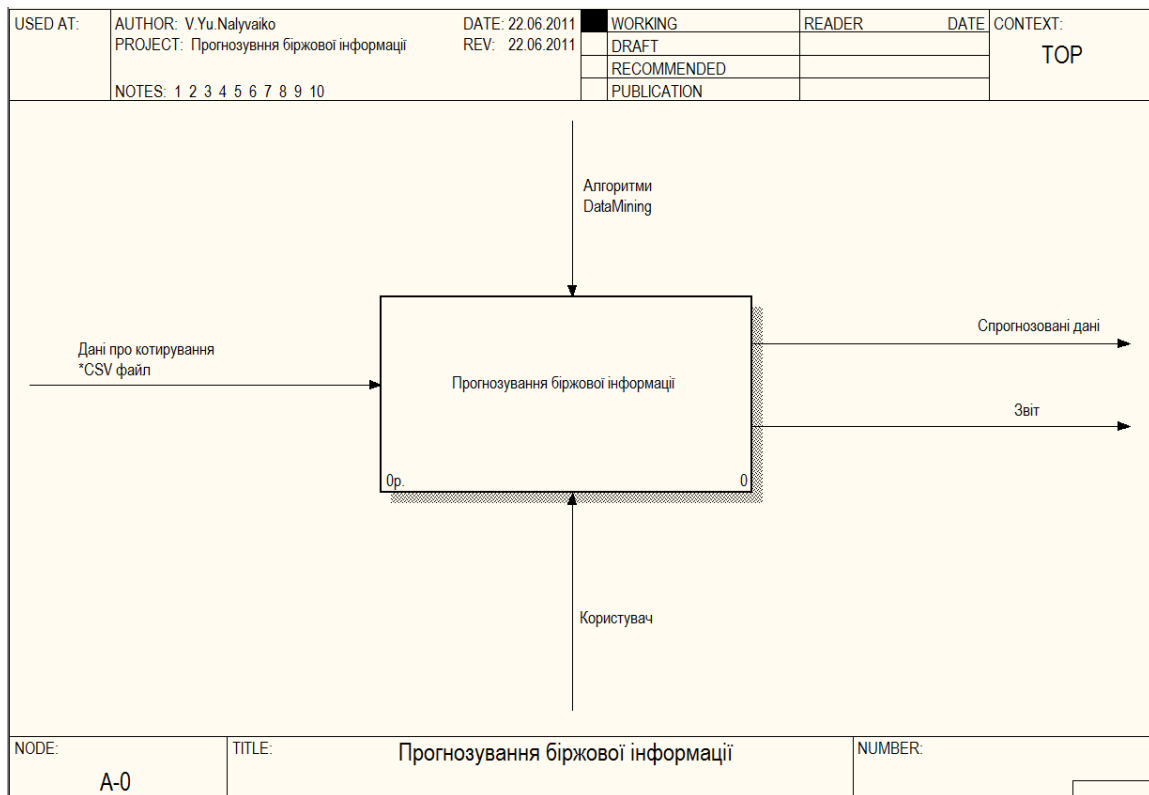


Рисунок Б.1 – Контекстна діаграма процесу Прогнозування біржової інформації

Таблиця Б.1 – Опис основних елементів контекстної діаграми процесу Прогнозування біржової інформації

Назва стрілки	Опис	Тип
Дані про котирування *CSV файл	Дані, за якими будується кластерний аналіз і нейронна мережа, а також прогноз.	Input
Алгоритми DataMining	Алгоритми Microsoft для кластерного аналізу і нейромережевого моделювання.	Control
Спрогнозовані дані	Результуючий набір спрогнозованих даних.	Output
Звіт	Імпорт даних та графіка в MS Excel	Output
Користувач	Користувач, встановлює основні параметри кластерного аналізу, нейромережевого моделювання та прогнозування.	Mechanism

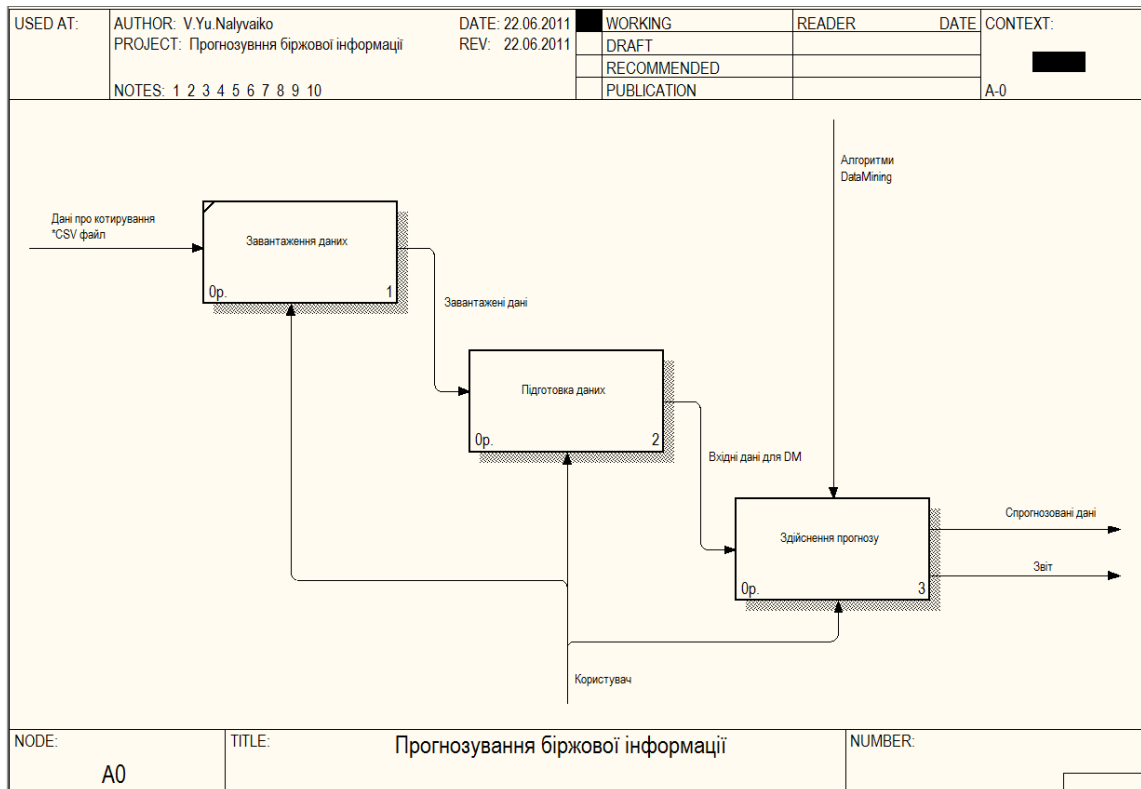


Рисунок Б.2 – Декомпозиція процесу Прогнозування біржової інформації

Таблиця Б.2 – Опис робіт діаграми декомпозиції процесу Прогнозування біржової інформації

Функціональний блок	Опис	Статус
Завантаження даних	Завантаження і зберігання біржової інформації, із засобів звернення до DB Engine через SQL-запити.	Робота
Підготовка даних	Перетворення формату даних, очищення, доповнення, обчислення кореляції, нормалізація даних, створення таблиць вихідних даних для DM.	Робота
Здійснення прогнозу	Створення моделі інтелектуального аналізу даних на базі алгоритмів кластеризації і нейронних мереж та отримання прогнозу	Робота

Таблиця Б.3 – Опис зв'язків між роботами діаграми декомпозиції процесу Прогнозування біржової інформації

Стрілка	Джерело	Тип	Призначення	Тип
Дані про котирування *CSV файл	Контекстна діаграма		Завантаження даних	Input
Завантажені дані	Завантаження даних	Output	Підготовка даних	Input

Продовження таблиці Б.3:

Стрілка	Джерело	Тип	Призначення	Тип
Вхідні дані для DM	Підготовка даних	Output	Здійснення прогнозу	Input
Спрогнозовані дані	Здійснення прогнозу	Output	{Border}	Output
Звіт	Здійснення прогнозу	Output	{Border}	Output
Алгоритми DataMining	Контекстна діаграма		Здійснення прогнозу	Control
Користувач	Контекстна діаграма		Завантаження даних Підготовка даних Здійснення прогнозу	Mechanism

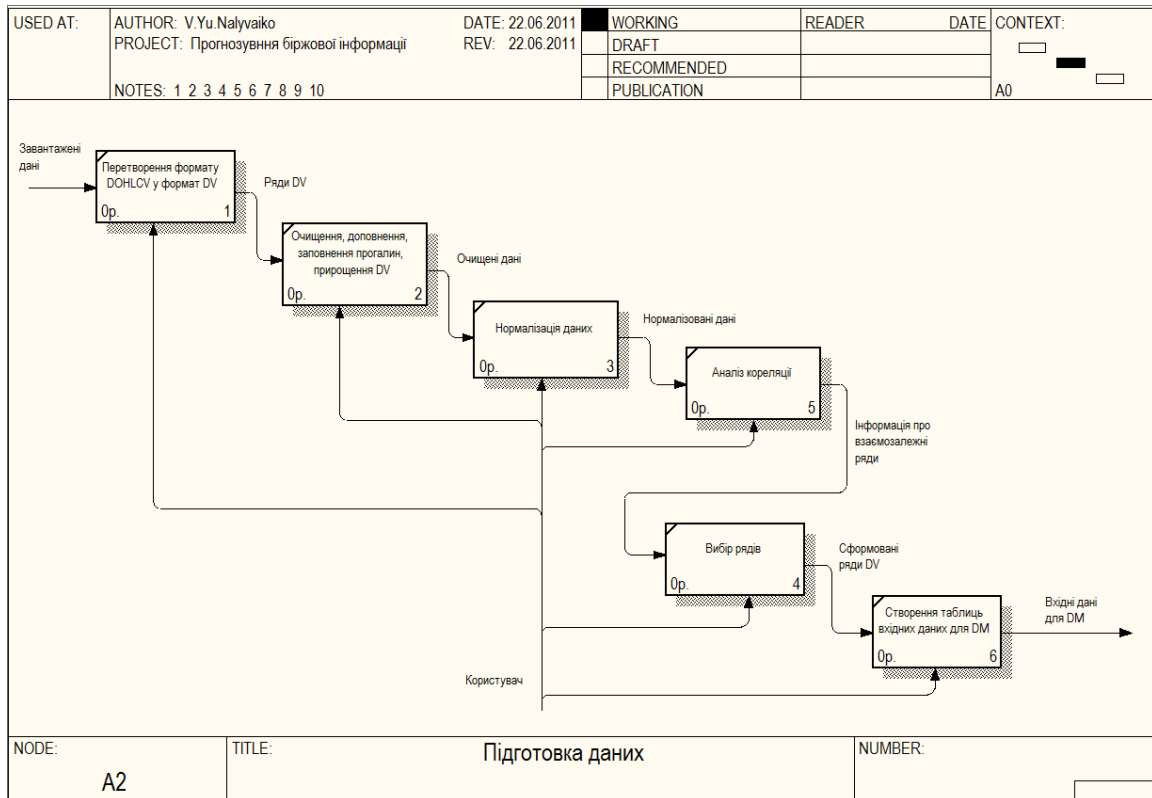


Рисунок Б.3 – Декомпозиція процесу Підготовка даних

Таблиця Б.4 – Опис робіт діаграми декомпозиції процесу Підготовка даних

Функціональний блок	Опис	Статус
Перетворення формату DOHLCV у формат DV	Перетворення формату DOHLCV (Date, Open, High, Low, Close, Volume) у формат DV (Date, Value).	Робота

Продовження таблиці Б.4:

Функціональний блок	Опис	Статус
Очищення, доповнення, заповнення прогалів, прирощення DV	Позбавлення від даних з використанням SMA (Simple Moving Average), які найбільше відрізняються від загальної множини.	Робота
Нормалізація даних	Невільювання можливої проблеми домінування одних змінних над іншими, якщо їх значення розходяться на кілька порядків.	Робота
Аналіз кореляції	Наявність яскраво вираженої кореляції між вхідним і прогнозованим рядами свідчить про те що використання даних рядів явно виправдане.	Робота
Вибір рядів	Вибір рядів для створення таблиць вхідних даних для DM.	Робота
Створення таблиць вхідних даних для DM	Об'єднання числових рядів DV в кінцеву таблицю, яка буде подаватися на вхід алгоритмам кластеризації і нейронних мереж.	Робота

Таблиця Б.5 – Опис зв'язків між роботами діаграми декомпозиції процесу Підготовка даних

Стрілка	Джерело	Тип	Призначення	Тип
Завантажені дані	Прогнозування біржової інформації		Перетворення формату DOHLCV у формат DV	Input
Ряди DV	Перетворення формату DOHLCV у формат DV	Output	Очищення, доповнення, заповнення прогалів, прирощення DV	Input
Очищені дані	Очищення, доповнення, заповнення прогалів, прирощення DV	Output	Нормалізація даних	Input
Нормалізовані дані	Нормалізація даних	Output	Аналіз кореляції	Input
Інформація про взаємозалежні ряди	Аналіз кореляції	Output	Вибір рядів	Input
Сформовані ряди DV	Вибір рядів	Output	Створення таблиць вхідних даних для DM	Input

Продовження таблиці Б.5:

Стрілка	Джерело	Тип	Призначення	Тип
Користувач	Прогнозування біржової інформації		Перетворення формату DOHLCV у формат DV; Очищення, доповнення, заповнення прогалів, прирощення DV; Нормалізація даних; Аналіз кореляції; Вибір рядів; Створення таблиць вхідних даних для DM.	Mechanism

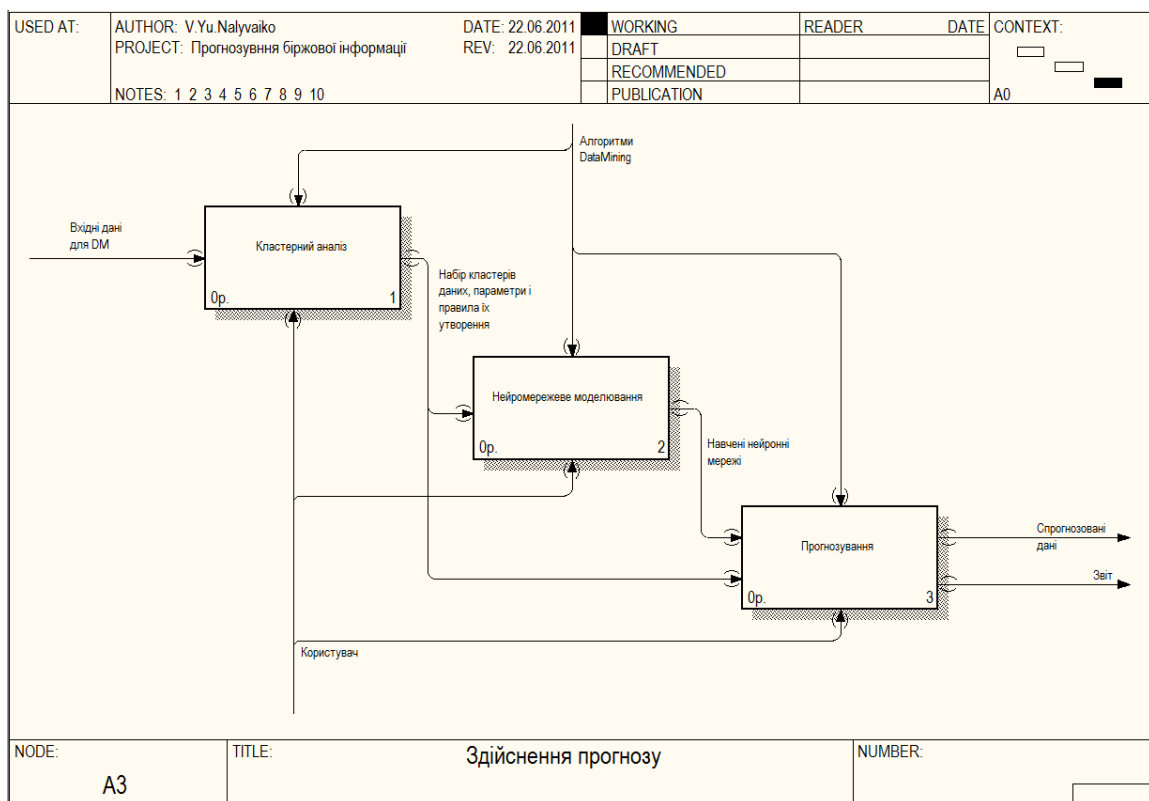


Рисунок Б.4 – Декомпозиція процесу Здійснення прогнозу

Таблиця Б.6 – Опис робіт діаграми декомпозиції процесу Здійснення прогнозу

Функціональний блок	Опис	Статус
Кластерний аналіз	Встановлення користувачем параметрів кластерного аналізу. Виділення зв'язків в наборі даних, формування ряду кластерів на основі вхідних даних.	Робота

Продовження таблиці Б.6:

Функціональний блок	Опис	Статус
Нейромережеве моделювання	Встановлення користувачем параметрів навчання нейронної мережі. Пошук найкращої множини ваг.	Робота
Прогнозування	Визначення кластера до якого відноситься набір входів, відношення точки до однієї з навчених нейронних мереж, здійснення прогнозу для даної точки та через нейронну мережу, яка навчена і натренована по всій вибірці.	Робота

Таблиця Б.7 – Опис зв'язків між роботами декомпозиції діаграми процесу Здійснення прогнозу

Стрілка	Джерело	Тип	Призначення	Тип
Вхідні дані для DM	Прогнозування біржової інформації		Кластерний аналіз	Input
Набір кластерів даних, параметри і правила їх утворення	Кластерний аналіз	Output	Нейромережеве моделювання	Input
Навчені нейронні мережі	Нейромережеве моделювання	Output	Прогнозування	Input
Спрогнозовані дані	Прогнозування	Output	{Border}	Output
Звіт	Прогнозування	Output	{Border}	Output
Алгоритми DataMining	Прогнозування біржової інформації		Кластерний аналіз Нейромережеве моделювання Прогнозування	Control
Користувач	Прогнозування біржової інформації		Кластерний аналіз Нейромережеве моделювання Прогнозування	Mechanism

Додаток В

Декомпозиція процесів Кластерного та Нейромережевого аналізу

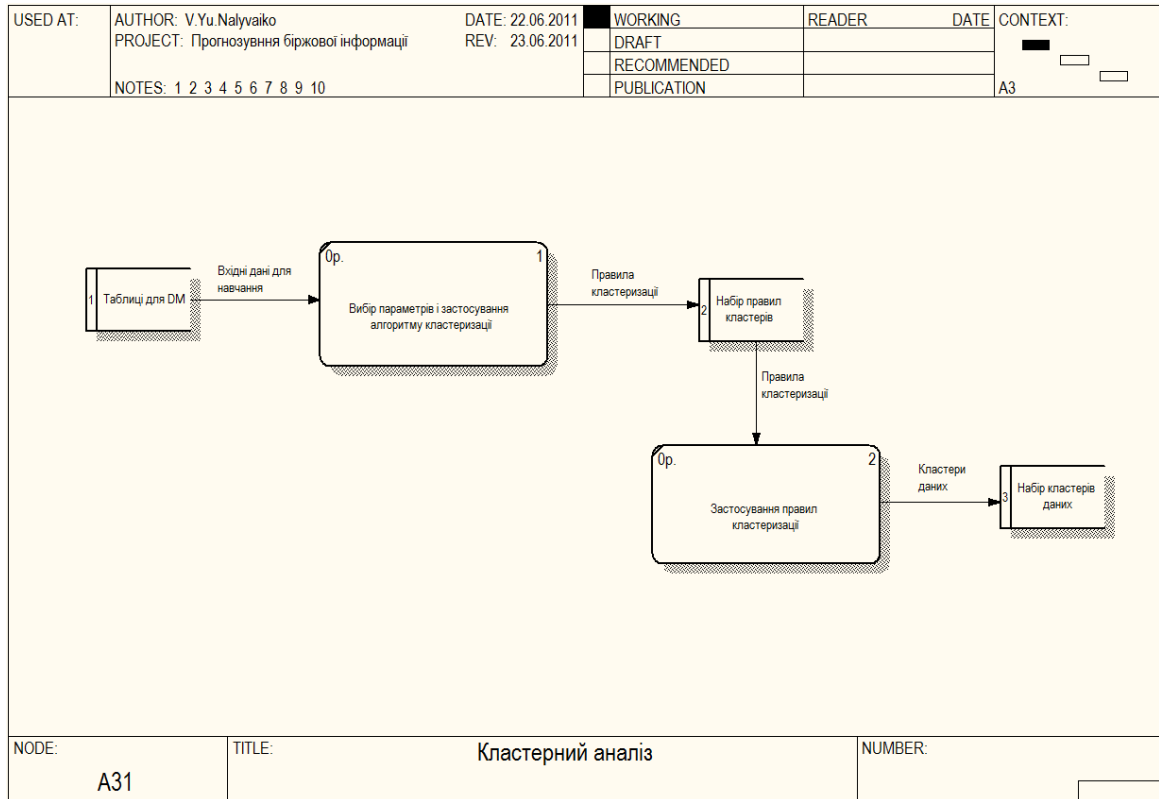


Рисунок В.1 – Декомпозиція процесу Кластерний аналіз

Таблиця В.1 – Опис робіт діаграми декомпозиції процесу Кластерний аналіз

Функціональний блок	Опис	Статус
Вибір параметрів і застосування алгоритму кластеризації	Встановлення необхідних параметрів алгоритму кластеризації (Microsoft) і запуск на виконання	Робота
Застосування правил кластеризації	Знаходження кластерів даних за допомогою отриманих правил кластеризації	Робота
Таблиці для DM	Таблиця з вихідними даними, сформована на етапі підготовки даних	Робота
Набір правил кластерів	Правила, що визначають сегментацію даних	Робота
Набір кластерів даних	Отримані в результаті роботи алгоритму кластери	Робота

Таблиця В.2 – Опис зв'язків між роботами діаграми декомпозиції процесу Кластерний аналіз

Стрілка	Джерело	Призначення
Вхідні дані для навчання	Таблиці для DM	Вибір параметрів і застосування алгоритму кластеризації
Правила кластеризації	Вибір параметрів і застосування алгоритму кластеризації	Набір правил кластерів
Правила кластеризації	Набір правил кластерів	Застосування правил кластеризації
Кластери даних	Застосування правил кластеризації	Набір кластерів даних

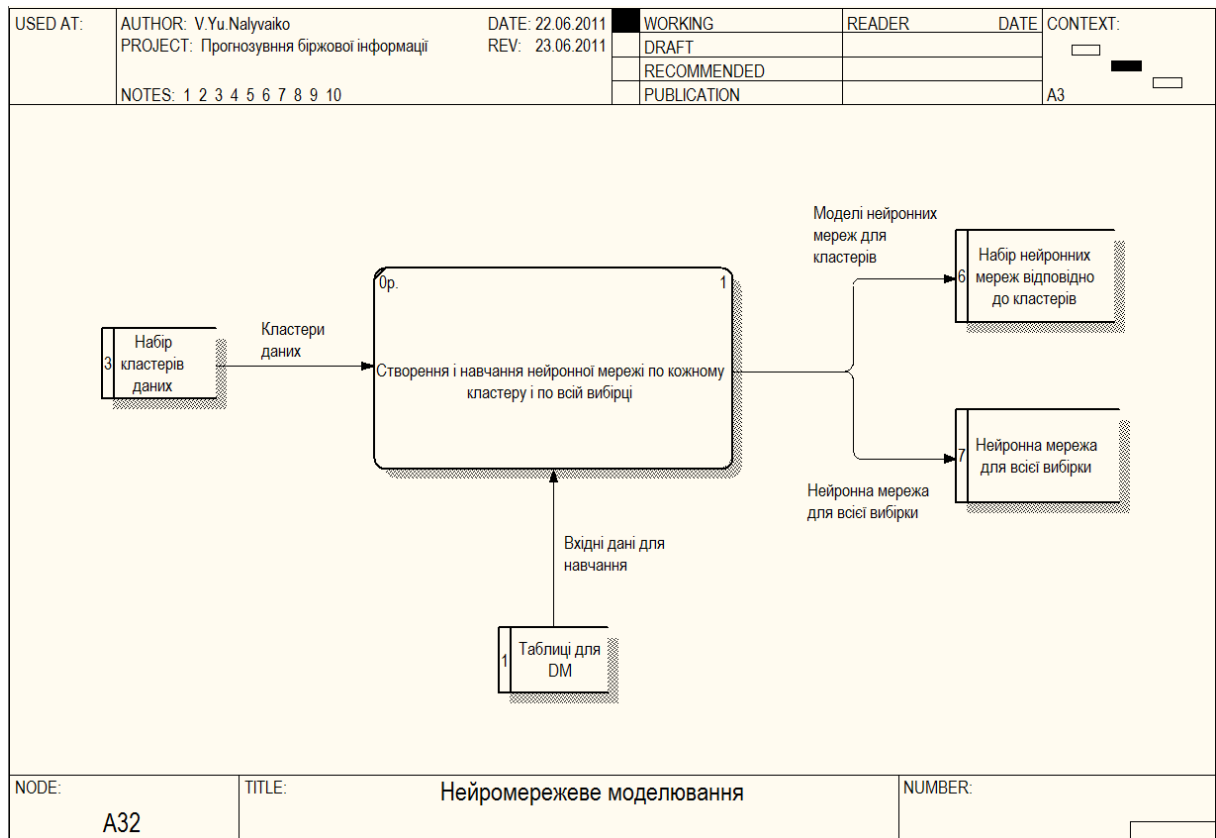


Рисунок В.2 – Декомпозиція процесу Нейромережеве моделювання

Таблиця В.3 – Опис робіт діаграми декомпозиції процесу
Нейромережеве моделювання

Функціональний блок	Опис	Статус
Створення і навчання нейронної мережі по кожному кластеру і по всій вибірці	Створення і навчання нейронних мереж відповідно до кожного кластеру	Робота
Набір кластерів даних	Отримані в результаті роботи алгоритму кластери	Робота
Таблиці для DM	Таблиця з вихідними даними, сформована на етапі підготовки даних	Робота
Набір нейронних мереж відповідно до кластерів	Навчені нейронні мережі відповідно до кожного кластеру	Робота
Нейронна мережа для всієї вибірки	Навчена нейронна мережа по всій вибірці	Робота

Таблиця В.4 – Опис зв'язків між роботами діаграми декомпозиції
процесу Нейромережеве моделювання

Стрілка	Джерело	Призначення
Кластери даних	Набір кластерів даних	Створення і навчання нейронних мереж відповідно до кожного кластеру
Вхідні дані для навчання	Таблиці для DM	Створення і навчання нейронних мереж відповідно до кожного кластеру
Моделі нейронних мереж для кластерів	Створення і навчання нейронних мереж відповідно до кожного кластеру	Застосування правил кластеризації
Нейронна мережа для всієї вибірки	Створення і навчання нейронних мереж відповідно до кожного кластеру	Набір кластерів даних

Додаток Г

Декомпозиція процесу Прогнозування

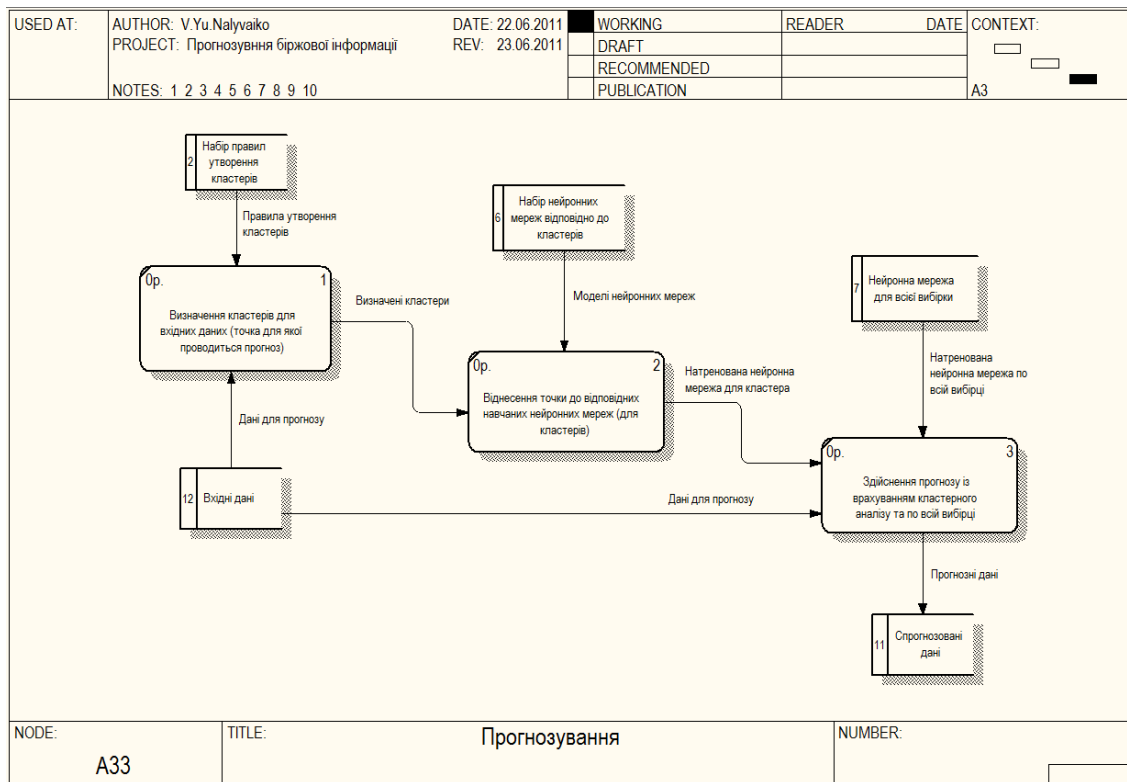


Рисунок Г.1 – Декомпозиція процесу Прогнозування

Таблиця Г.1 – Опис робіт діаграми декомпозиції процесу Прогнозування

Функціональний блок	Опис	Статус
Визначення кластерів для вхідних даних (точка для якої проводиться прогноз)	Знаходження, до якого кластеру найбільше підходить точка	Робота
Віднесення точки до відповідних навчаних нейронних мереж (для кластерів)	Відповідно до того, до якого кластеру належить точка, вона відноситься до навченої нейронної мережі	Робота
Здійснення прогнозу із врахуванням кластерного аналізу та по всій вибірці	За допомогою нейронної мережі, до якої була віднесена точка, здійснюється прогноз. Також прогноз здійснюється окремо по всій вибірці	Робота
Набір правил утворення кластерів	Правила, що визначають сегментацію даних	Робота
Набір нейронних мереж відповідно до кластерів	Навчені нейронні мережі відповідно до кожного кластеру	Робота

Продовження таблиці Г.1:

Функціональний блок	Опис	Статус
Нейронна мережа для всієї вибірки	Навчена нейронна мережа по всій вибірці	Робота
Спрогнозовані дані	Прогнозні дані на основі нейромережевого моделювання	Робота
Вхідні дані	Дані, по яким потрібно побудувати прогноз	Робота

Таблиця Г.2 – Опис зв'язків між роботами діаграми декомпозиції процесу Прогнозування

Стрілка	Джерело	Призначення
Визначені кластери	Визначення кластерів для вхідних даних (точка для якої проводиться прогноз)	Віднесення точки до відповідних навчених нейронних мереж (для кластерів)
Натренована нейронна мережа для кластера	Віднесення точки до відповідних навчених нейронних мереж (для кластерів)	Здійснення прогнозу із врахуванням кластерного аналізу та по всій вибірці
Правила утворення кластерів	Набір правил утворення кластерів	Визначення кластерів для вхідних даних (точка для якої проводиться прогноз)
Дані для прогнозу	Вхідні дані	Визначення кластерів для вхідних даних (точка для якої проводиться прогноз)
Моделі нейронних мереж	Набір нейронних мереж відповідно до кластерів	Віднесення точки до відповідних навчених нейронних мереж (для кластерів)
Натренована нейронна мережа по всій вибірці	Нейронна мережа для всієї вибірки	Здійснення прогнозу із врахуванням кластерного аналізу та по всій вибірці
Дані для прогнозу	Вхідні дані	Здійснення прогнозу із врахуванням кластерного аналізу та по всій вибірці
Прогнозні дані	Здійснення прогнозу із врахуванням кластерного аналізу та по всій вибірці	Спрогнозовані дані

Додаток Д

Структура бази даних, а також структура та призначення таблиць БД

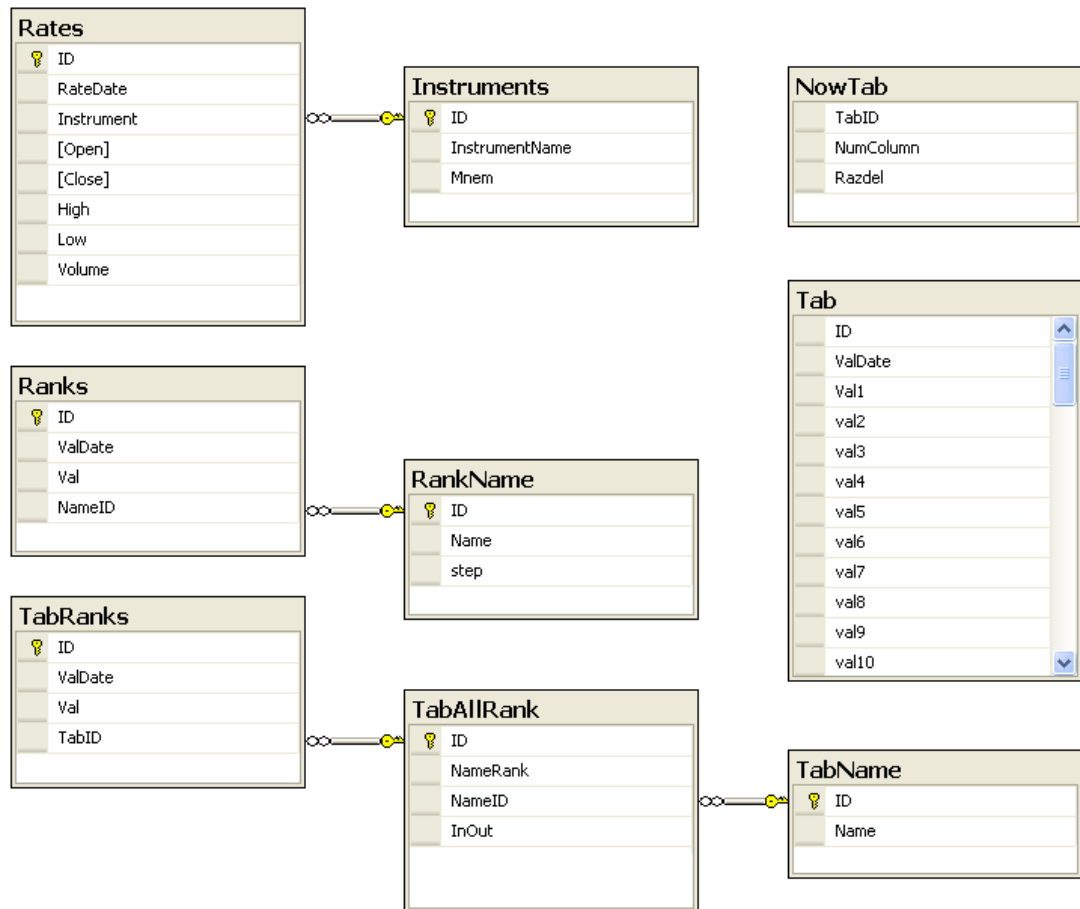


Рисунок Д.1 – Структура БД

Таблиця «Instruments» призначена для збереження назви і мнемоніки інструмента. Структурне представлення таблиці «Instruments» наведено в таблиці Д.1.

Таблиця Д.1 – Структура таблиці Instruments

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	ID	Int	NOT NULL	Первинний ключ
2	InstrumentName	Varchar(40)	NULL	Назва інструмента
3	Mnem	Varchar(3)	NULL	Мнемоніка інструмента

Таблиця «Rates» призначена для зберігання інформації, взятої з файлу *.CSV у форматі DOHLCV. Структурне представлення таблиці «Rates» наведено в таблиці Д.2.

Таблиця Д.2 – Структура таблиці Rates

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	ID	Int	NOT NULL	Первинний ключ
2	RateDate	Datetime	NULL	Дата
3	Instrument	Int	NULL	Зовнішній ключ (Instruments)
4	Open	Decimal(14,5)	NULL	Ціна відкриття діапазону
5	High	Decimal(14,5)	NULL	Максимум діапазону
6	Low	Decimal(14,5)	NULL	Мінімум діапазону
7	Close	Decimal(14,5)	NULL	Ціна закриття діапазону
8	Volume	Decimal(14,5)	NULL	Торговий об'єм діапазону

В таблиці «RankName» зберігаються імена рядів (формату DV). Структурне представлення таблиці «RankName» наведено в таблиці Д.3.

Таблиця Д.3 – Структура таблиці RankName

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	ID	Int	NOT NULL	Первинний ключ
2	Name	Varchar(50)	NOT NULL	Ім'я ряду DV
3	Step	Int	NOT NULL	Крок, на якому був створений ряд

Таблиця «Ranks» створена для збереження рядів у форматі DV, які потім використовуються модулем підготовки даних. Структурне представлення таблиці «Ranks» наведено в таблиці Д.4.

Таблиця Д.4 – Структура таблиці Ranks

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	ID	Int	NOT NULL	Первинний ключ
2	ValDate	Datetime	NOT NULL	Дата (ряд DV)
3	Val	Decimal(14,5)	NOT NULL	Значення ряду DV
4	NameID	Int	NOT NULL	Зовнішній ключ (RankName)

В таблиці «TabName» зберігаються назви таблиць. Структурне представлення таблиці «TabName» наведене в таблиці Д.5.

Таблиця Д.5 – Структура таблиці TabName

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	ID	Int	NOT NULL	Первинний ключ
2	Name	Varchar(50)	NOT NULL	Ім'я таблиці

Таблиця «TabAllRank» використовується для збереження імен рядів. Структурне представлення таблиці «TabAllRank» наведене в таблиці Д.6.

Таблиця Д.6 – Структура таблиці TabAllRank

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	ID	Int	NOT NULL	Первинний ключ
2	NameRank	Varchar(50)	NOT NULL	Ім'я ряду в таблиці
3	NameID	Int	NOT NULL	Зовнішній ключ (TabName)
4	InOut	Int	NOT NULL	Тип ряду (вхід – 1, вихід – 2 або коментар – 3)

В таблиці «TabRanks» зберігаються ряди у форматі DV, які потім використовуються в модулі прогнозування. Структурне представлення таблиці «TabRanks» наведене в таблиці Д.7.

Таблиця Д.7 – Структура таблиці TabRanks

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	ID	Int	NOT NULL	Первинний ключ
2	ValDate	Datetime	NOT NULL	Дата (ряд DV)
3	Val	Decimal(14,5)	NOT NULL	Значення (ряд DV)
4	TabID	Int	NOT NULL	Зовнішній ключ (TabAllRank)

Таблиця «NowTab» використовується для зберігання відомостей про поточну таблицю з рядів котирувань, яка знаходиться на одному з етапів прогнозування. Структурне представлення таблиці «NowTab» наведено в таблиці Д.8.

Таблиця Д.8 – Структура таблиці NowTab

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	TabID	Int	NOT NULL	Первинний ключ поточної таблиці
2	NumColumn	Int	NOT NULL	Кількість стовпців в поточній таблиці
3	Razdel	Int	NOT NULL	Стадія обробки таблиці (1 – кластерний аналіз, 2 – нейромережеве моделювання)

В таблиці «Tab» зберігаються ряди у форматі DV, зібрані в таблицю для прогнозування. Структурне представлення таблиці «Tab» наведено в таблиці Д.9.

Таблиця Д.9 – Структура таблиці Tab

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
1	ID	Int	NOT NULL	Первинний ключ
2	ValDate	Datetime	NOT NULL	Дата (рядів котирувань)

Продовження таблиці Д.9:

Номер атрибуту	Назва атрибуту	Тип даних	Обмеження	Призначення атрибуту
3	Val1...Val50	Decimal(14,5)	NULL	Значення рядів котирувань в поточній таблиці, використовуваний для кластерного аналізу і нейромережевого моделювання

Додаток Е

Блок-схема алгоритму кластеризації

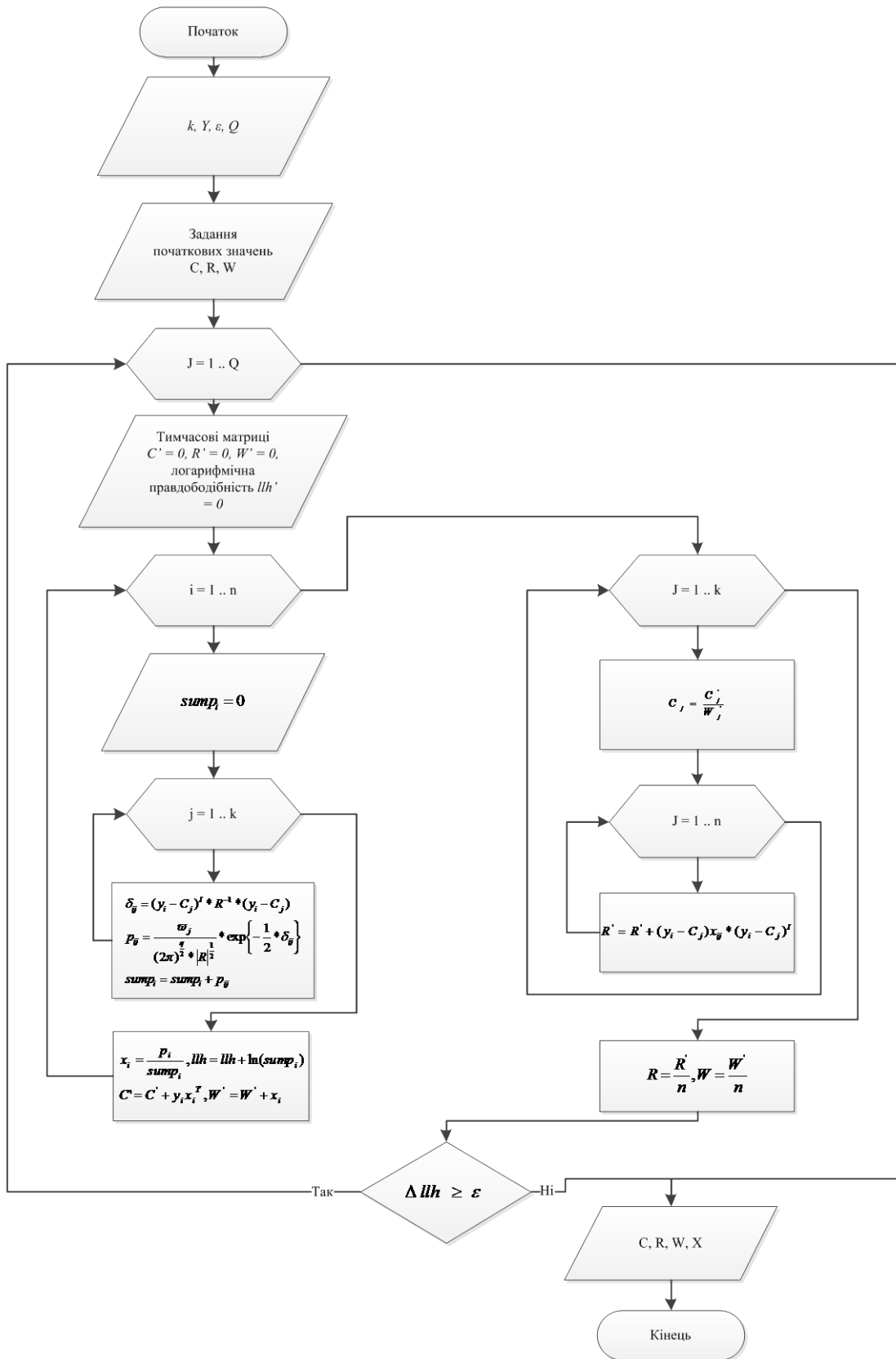


Рисунок Е.1 – Блок-схема алгоритму кластеризації

Додаток Є

Блок-схема алгоритму навчання нейронної мережі



Рисунок Є.1 – Блок-схема алгоритму навчання нейронної мережі

Додаток Ж

Блок-схема алгоритму очищення числового ряду

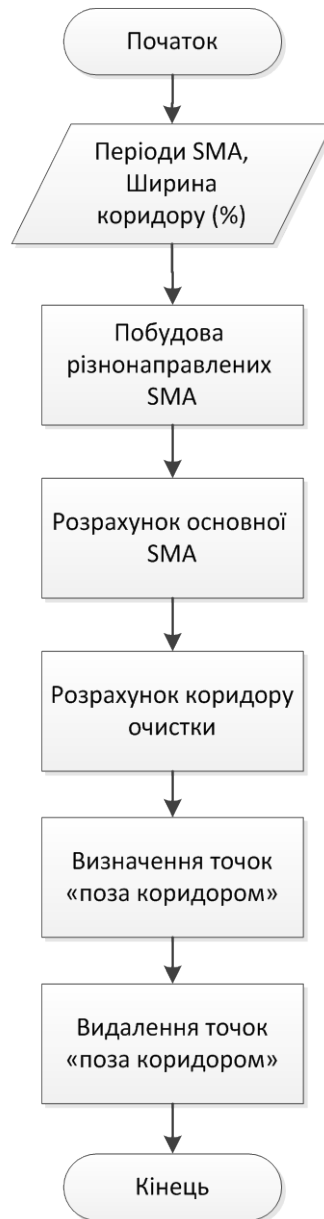


Рисунок Ж.1 – Блок-схема алгоритму очищення числового ряду

Додаток 3

Інтерфейс користувача АІС

Підсистема прогнозування на базі не йроме реже вого моделювання і класте ризації Наливайка Віталія МЕК-12

1. З'єднання з базою

1. З'єднання з базою

Ім'я сервера

Ім'я користувача

Пароль

Ім'я бази

Ім'я аналітичної бази

З'єднання з БД: з'єднання не встановлено

Рисунок 3.1 – Вкладка з'єднання з базою

Підсистема прогнозування на базі не йроме реже вого моделювання і класте ризації Наливайка Віталія МЕК-12

2. Інструменти

2. Інструменти

Завантаження нових інструментів

Перегляд/видалення інструментів

З'єднання з БД: з'єднання встановлено

Рисунок 3.2 – Вкладка інструменти

Підсистема прогнозування на базі не йроме реже вого моделювання і кластеризації Наливайка Віталія МЕК-12

2.2. Завантаження в БД

Завантаження Графіки

Інструменти

GOLD

Всього строк: 2048
 Max,Min Open:(374.50000,1901.80000)
 Max,Min High:(370.50000,1874.92000)
 Max,Min Low:(374.10000,1901.83000)
 Max,Min Close:(378.10000,1920.53000)
 Max,Min Volume:(29,134170)

Завантажити Видалити Зберегти в БД

Дата
 Ціна відкриття
 Найбільша ціна
 Найменша ціна
 Ціна закриття
 Значення

Назад

З'єднання з БД: з'єднання встановлено

DateTime	Open	High	Low	Close	Volume
12/17/2003 00:00:00	407.50000	413.80000	406.10000	412.20000	2696(GOLD)
12/18/2003 00:00:00	412.10000	412.80000	407.10000	410.80000	2882(GOLD)
12/19/2003 00:00:00	411.10000	412.00000	406.60000	409.10000	2584(GOLD)
12/22/2003 00:00:00	409.00000	412.80000	407.10000	410.10000	2179(GOLD)
12/23/2003 00:00:00	410.10000	412.00000	408.30000	410.80000	1906(GOLD)
12/24/2003 00:00:00	410.80000	413.50000	409.70000	412.10000	1204(GOLD)
12/25/2003 00:00:00	412.50000	414.80000	411.30000	413.20000	53(GOLD)
12/26/2003 00:00:00	413.60000	414.80000	410.10000	412.26000	292(GOLD)
12/29/2003 00:00:00	412.30000	416.50000	410.80000	415.30000	1852(GOLD)
12/30/2003 00:00:00	415.10000	418.10000	413.30000	416.30000	1965(GOLD)
12/31/2003 00:00:00	416.30000	418.60000	413.60000	415.60000	1306(GOLD)
01/02/2004 00:00:00	415.30000	417.00000	414.30000	416.10000	172(GOLD)
01/05/2004 00:00:00	414.80000	425.50000	414.30000	424.20000	2212(GOLD)
01/06/2004 00:00:00	424.18000	431.10000	420.60000	422.60000	2313(GOLD)
01/07/2004 00:00:00	422.70000	425.00000	418.80000	420.00000	4603(GOLD)
01/08/2004 00:00:00	420.00000	425.60000	416.20000	423.30000	5768(GOLD)
01/09/2004 00:00:00	423.40000	427.80000	418.30000	425.70000	4760(GOLD)
01/12/2004 00:00:00	427.20000	430.10000	422.20000	425.60000	4168(GOLD)
01/13/2004 00:00:00	425.60000	427.80000	421.50000	423.10000	4734(GOLD)
01/14/2004 00:00:00	423.00000	425.00000	416.70000	421.20000	5532(GOLD)
01/15/2004 00:00:00	420.20000	421.70000	406.90000	408.00000	6505(GOLD)
01/16/2004 00:00:00	408.00000	411.10000	404.50000	406.30000	4404(GOLD)
01/20/2004 00:00:00	406.80000	413.80000	406.30000	412.50000	4257(GOLD)
01/21/2004 00:00:00	412.30000	414.30000	406.50000	410.70000	4743(GOLD)
01/22/2004 00:00:00	410.60000	414.00000	407.60000	409.40000	4736(GOLD)
01/23/2004 00:00:00	409.50000	413.80000	406.00000	407.50000	5798(GOLD)
01/26/2004 00:00:00	407.50000	410.30000	404.10000	406.00000	4967(GOLD)
01/27/2004 00:00:00	406.10000	413.20000	402.10000	409.60000	11961(GOLD)
01/28/2004 00:00:00	409.60000	416.00000	407.80000	414.20000	3962(GOLD)
01/29/2004 00:00:00	414.10000	414.70000	395.80000	399.10000	6386(GOLD)
01/30/2004 00:00:00	399.10000	403.60000	397.80000	401.60000	5369(GOLD)
02/02/2004 00:00:00	402.60000	404.60000	394.10000	398.50000	4572(GOLD)
02/03/2004 00:00:00	398.00000	404.40000	397.50000	398.70000	5053(GOLD)
02/04/2004 00:00:00	398.70000	402.10000	397.60000	400.50000	4706(GOLD)
02/05/2004 00:00:00	400.60000	401.80000	396.10000	397.40000	4723(GOLD)
02/06/2004 00:00:00	397.60000	410.10000	394.30000	403.00000	5388(GOLD)

Рисунок 3.3 – Вкладка «Завантаження в БД»

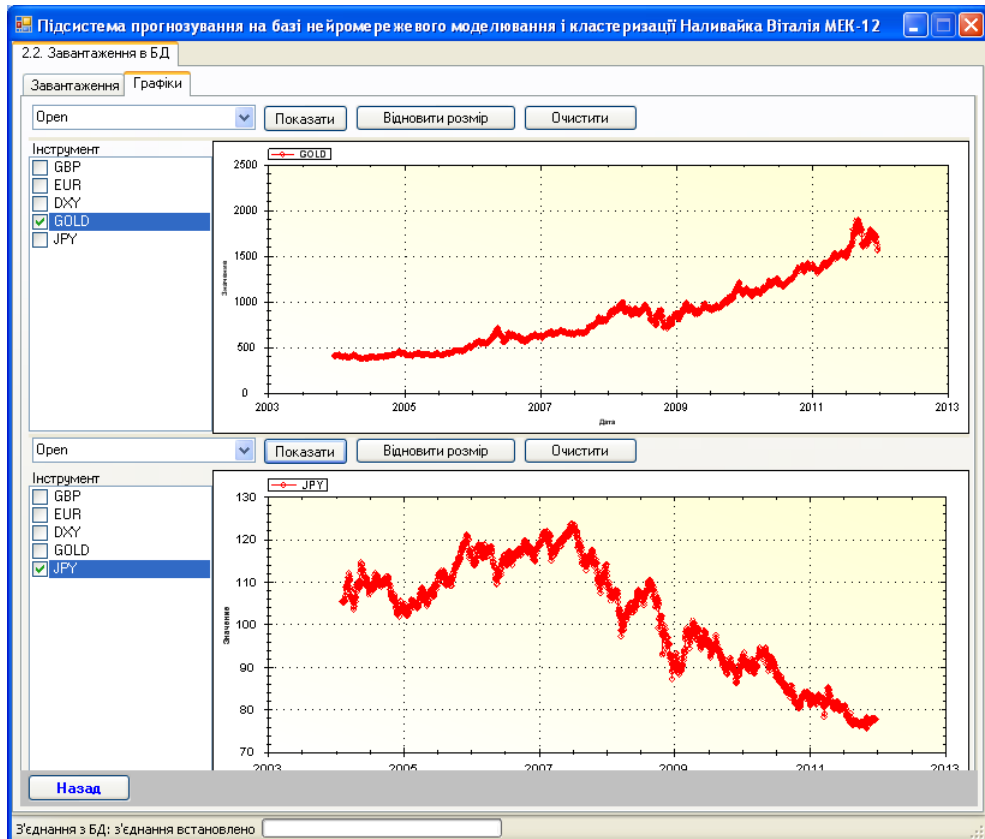


Рисунок 3.4 – Вкладка «Графіки»

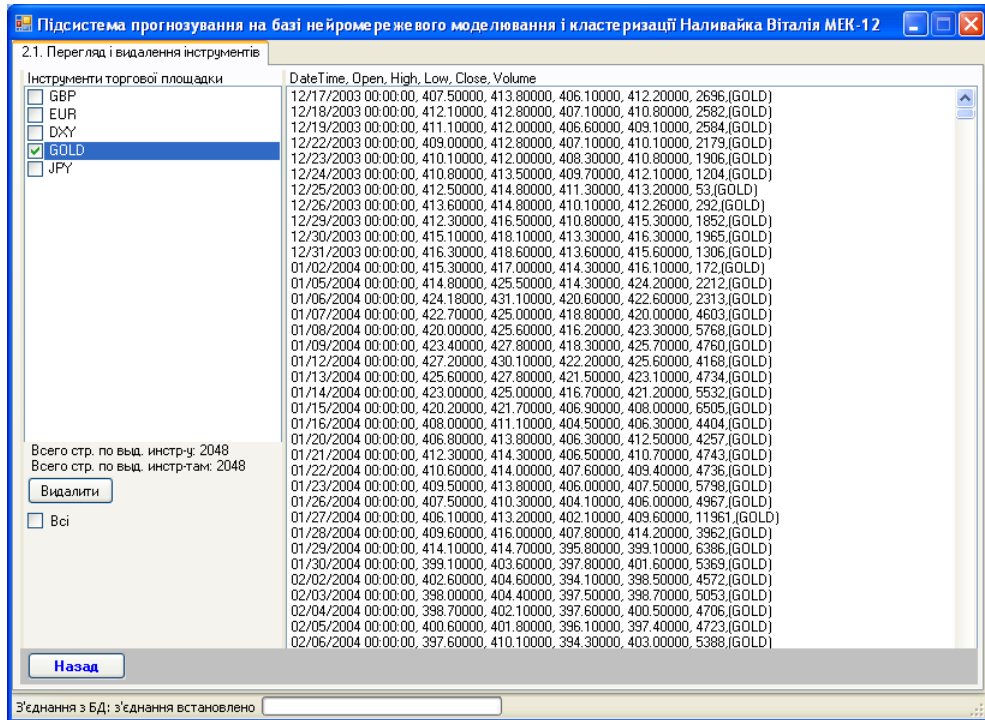


Рисунок 3.5– Вкладка «Перегляд і видалення інструментів»

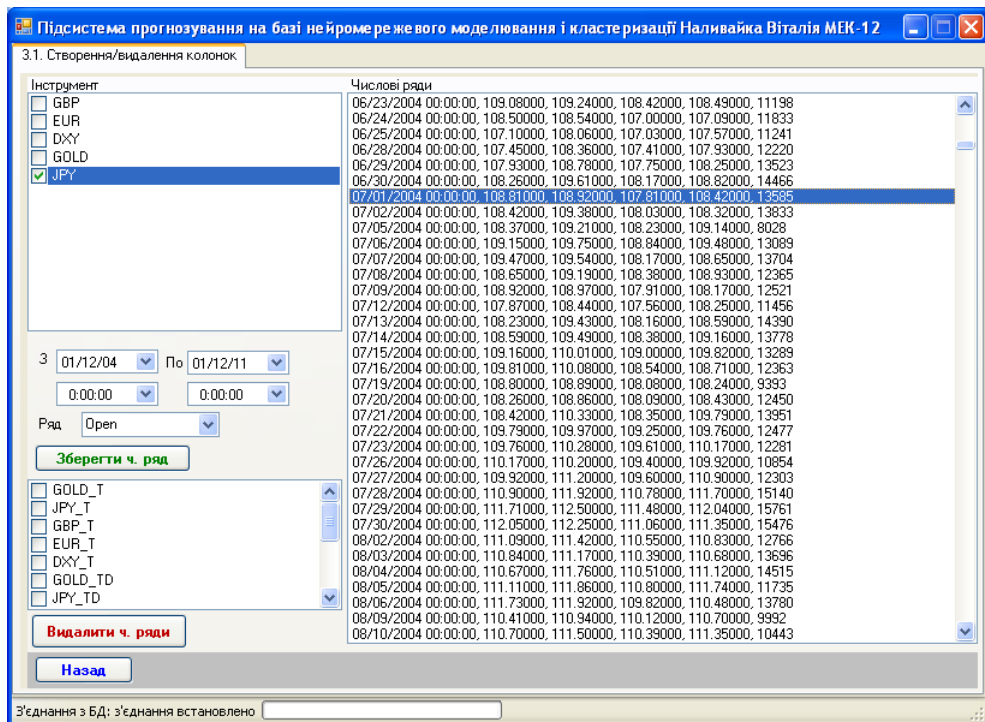


Рисунок 3.6 – Вкладка «Створення / Видалення колонок»

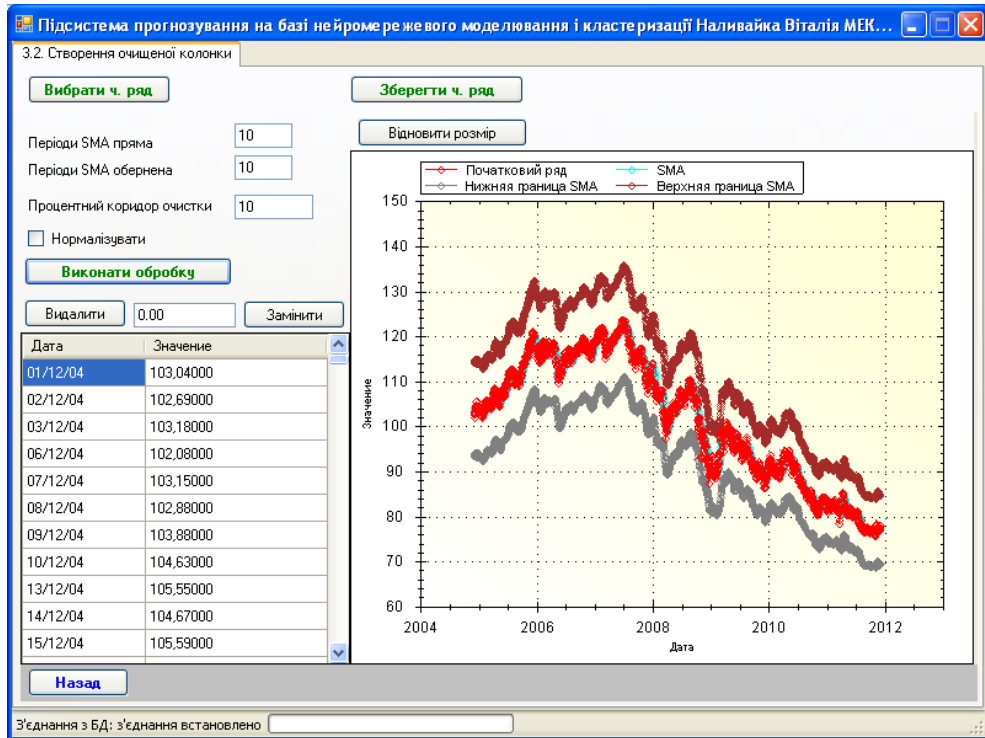


Рисунок 3.7 – Вкладка «Створення очищеної колонки»

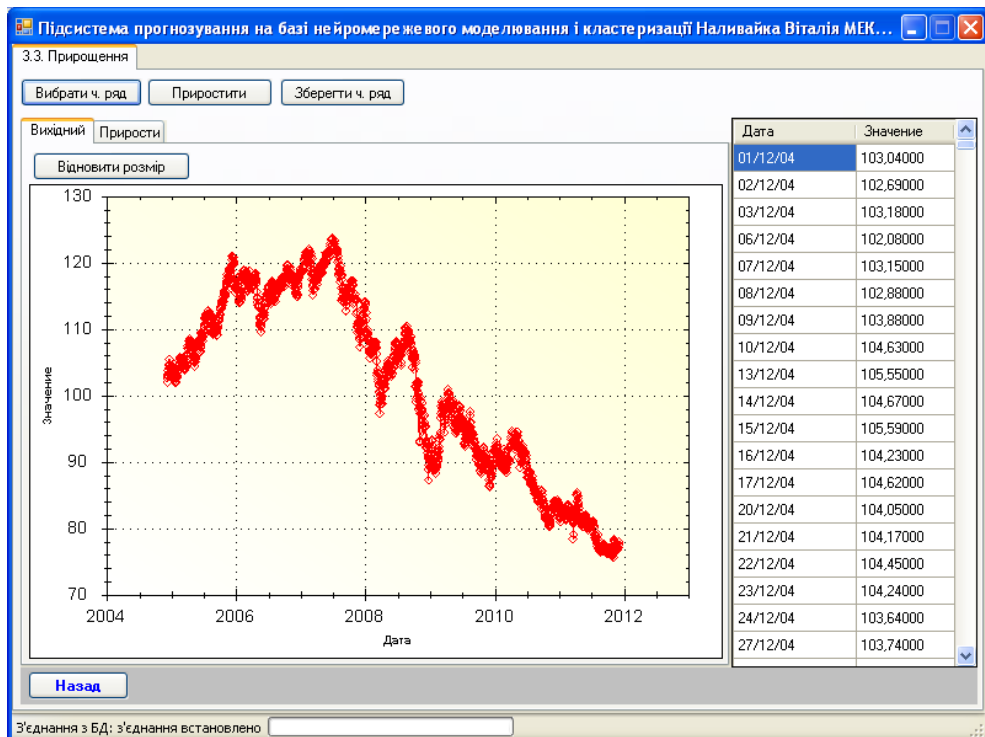


Рисунок 3.8 – Вкладка «Прирощення»

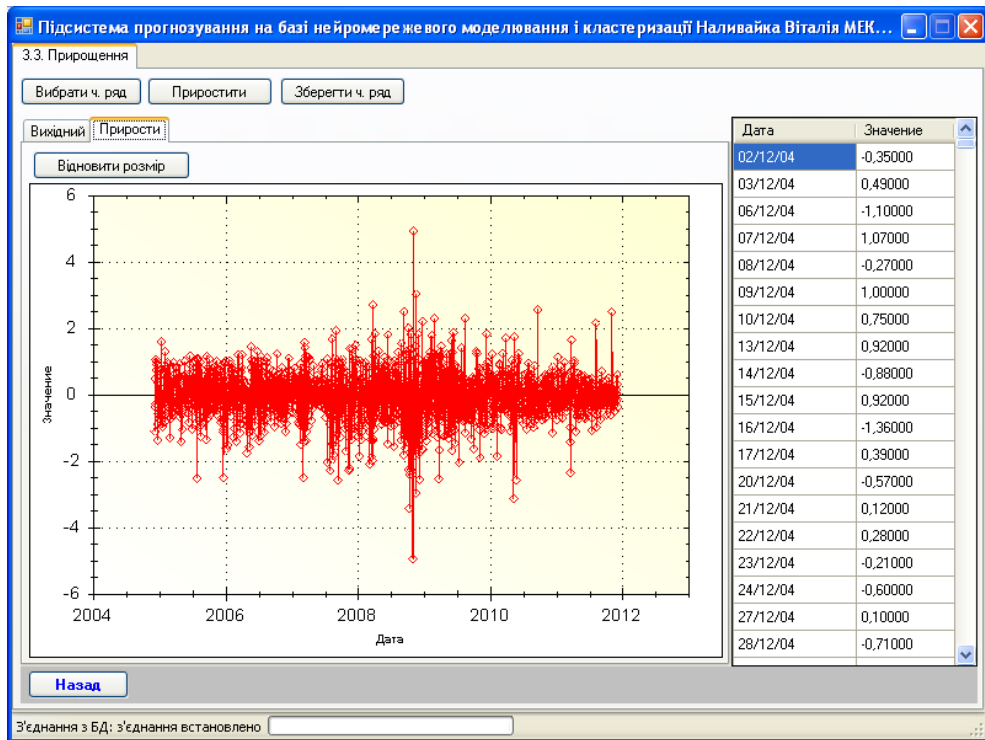


Рисунок 3.9 – Вкладка «Прирости»

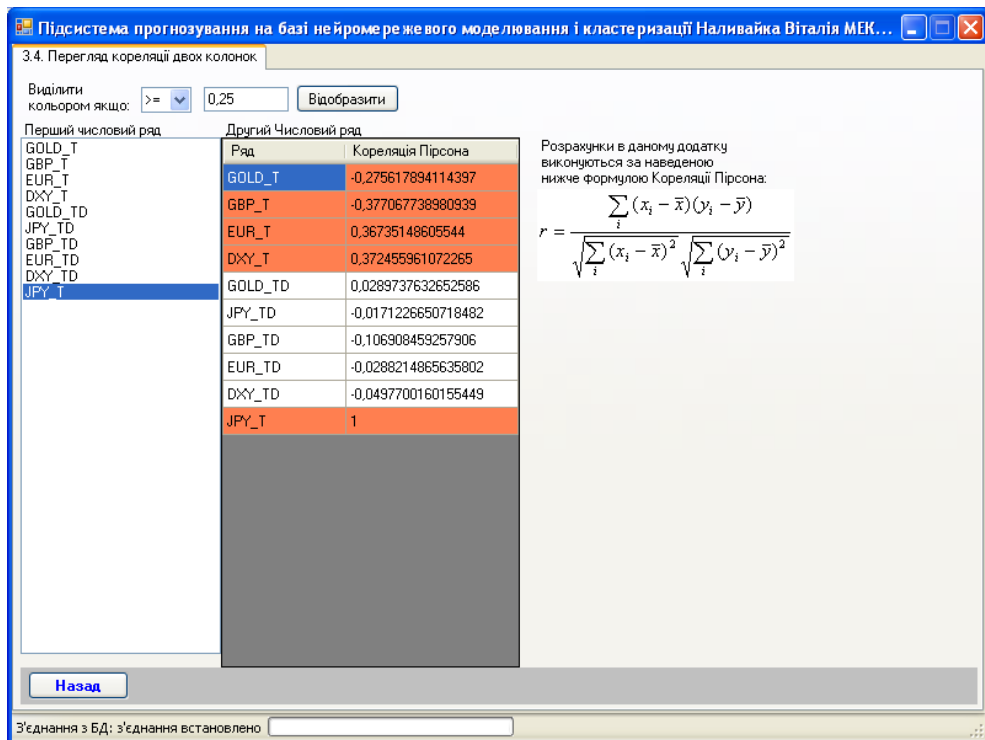


Рисунок 3.10 – Вкладка «Перегляд кореляції двох колонок»

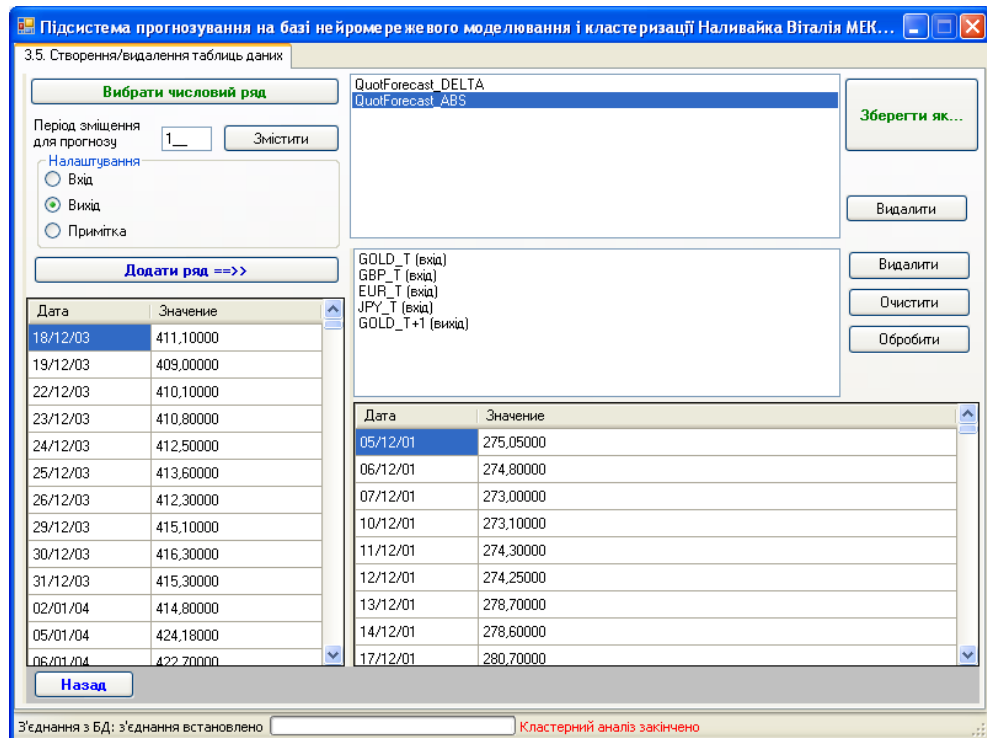


Рисунок 3.11 – Вкладка «Створення / Видалення таблиці даних»

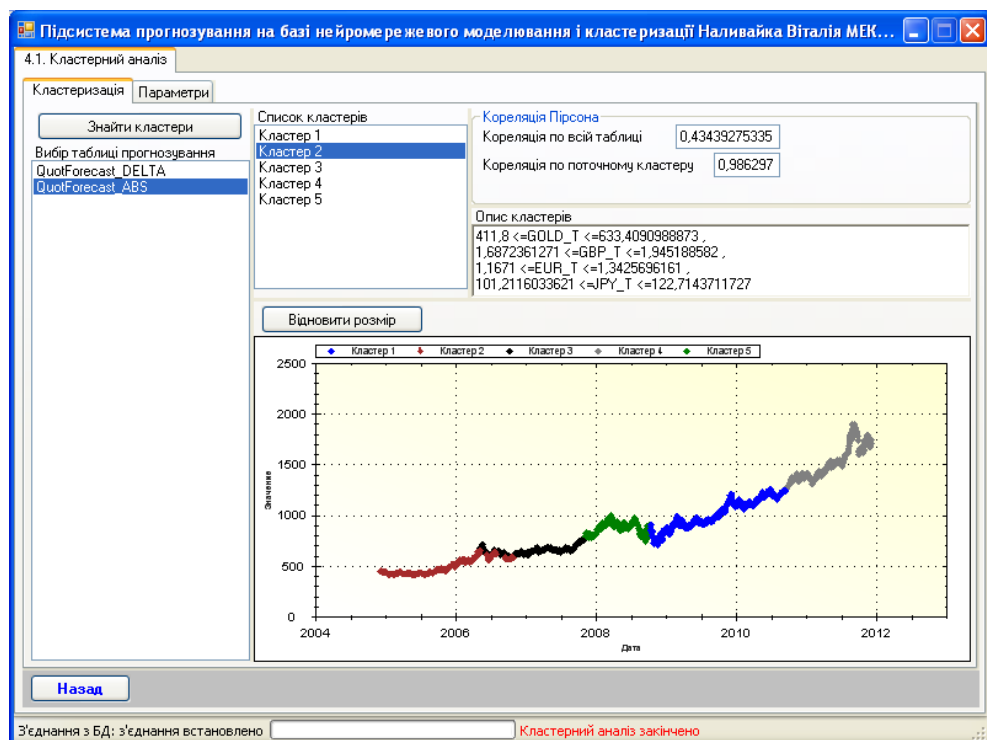


Рисунок 3.12 – Вкладка «Кластерний аналіз»

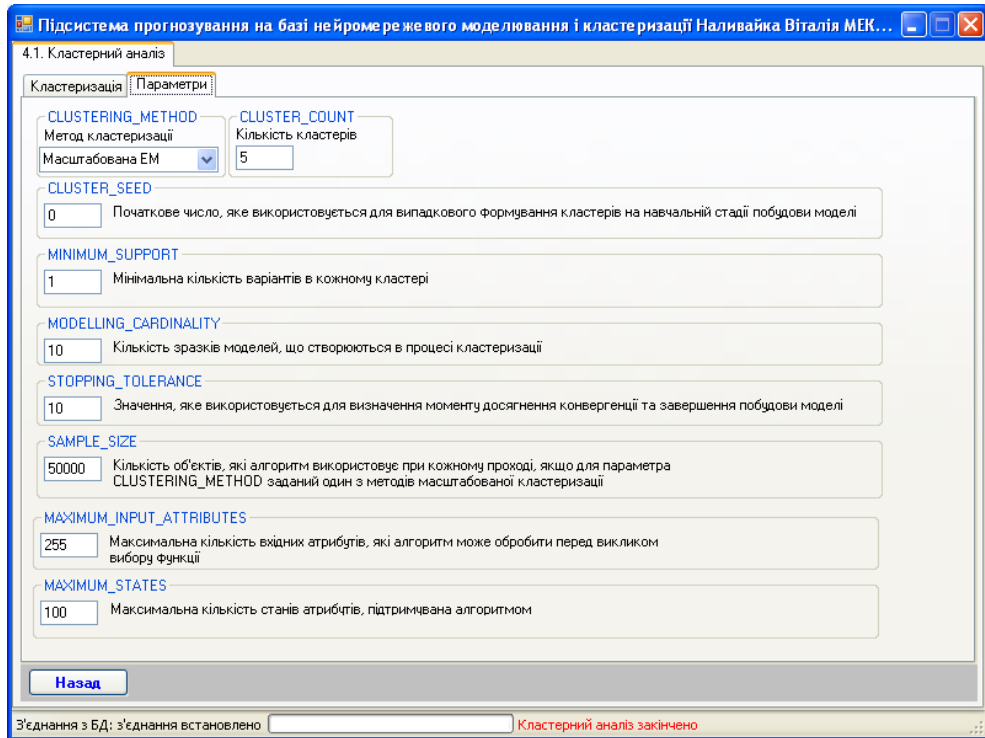


Рисунок 3.13 – Вибір параметрів алгоритму кластерного аналізу

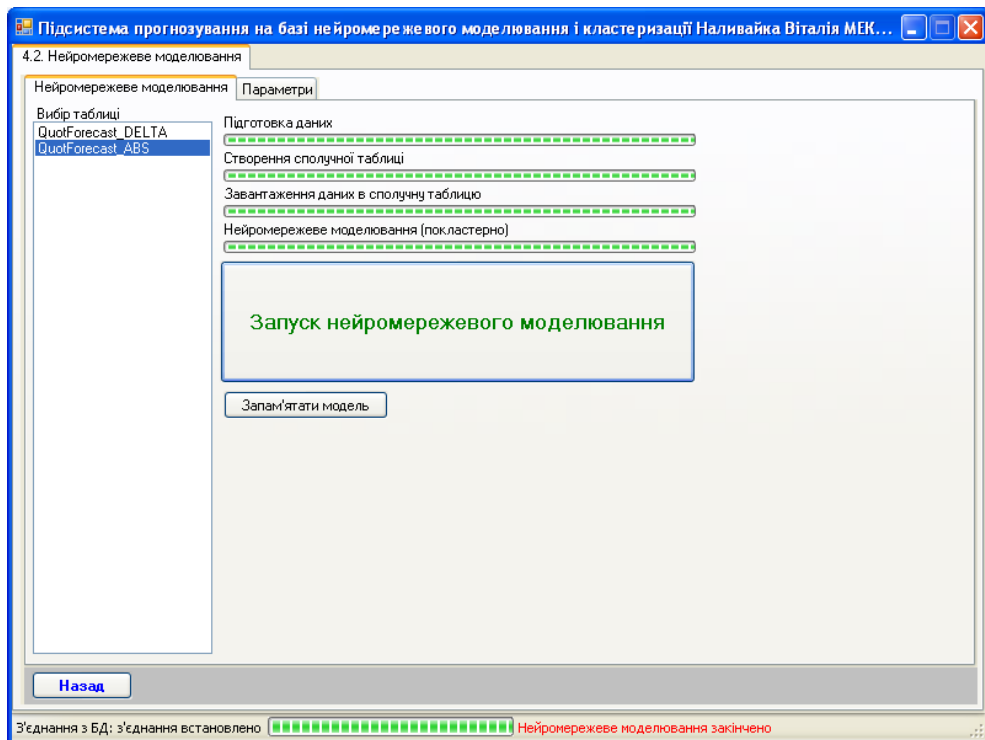


Рисунок 3.14 – Підкладка «Неймережеве моделювання»

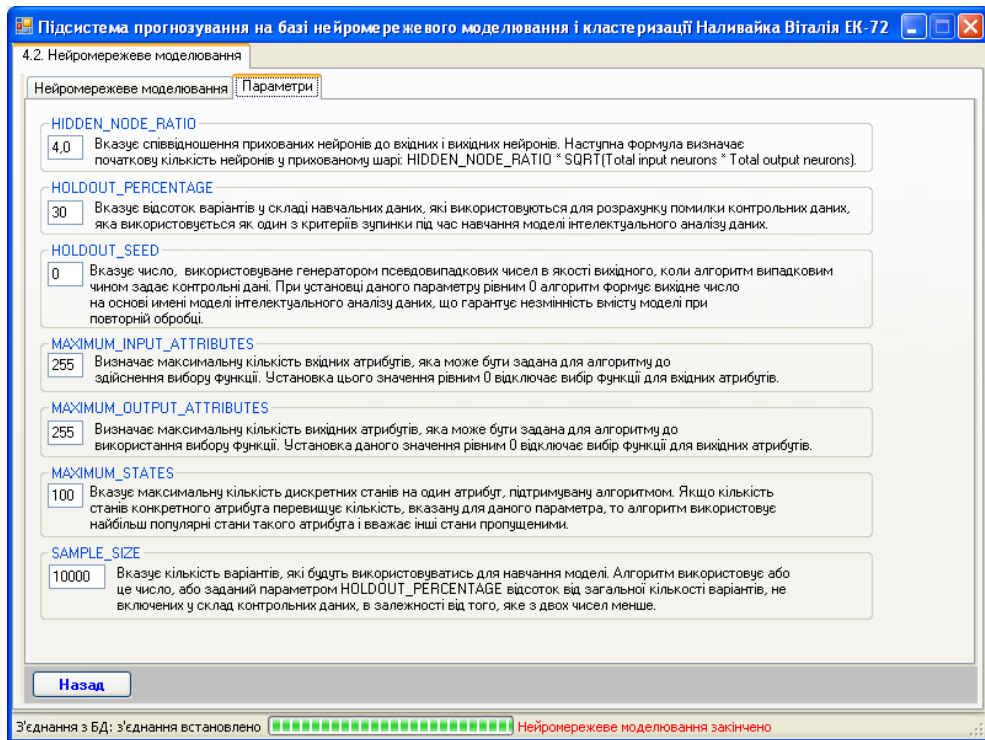


Рисунок 3.15 – Підвкладка «Параметри»

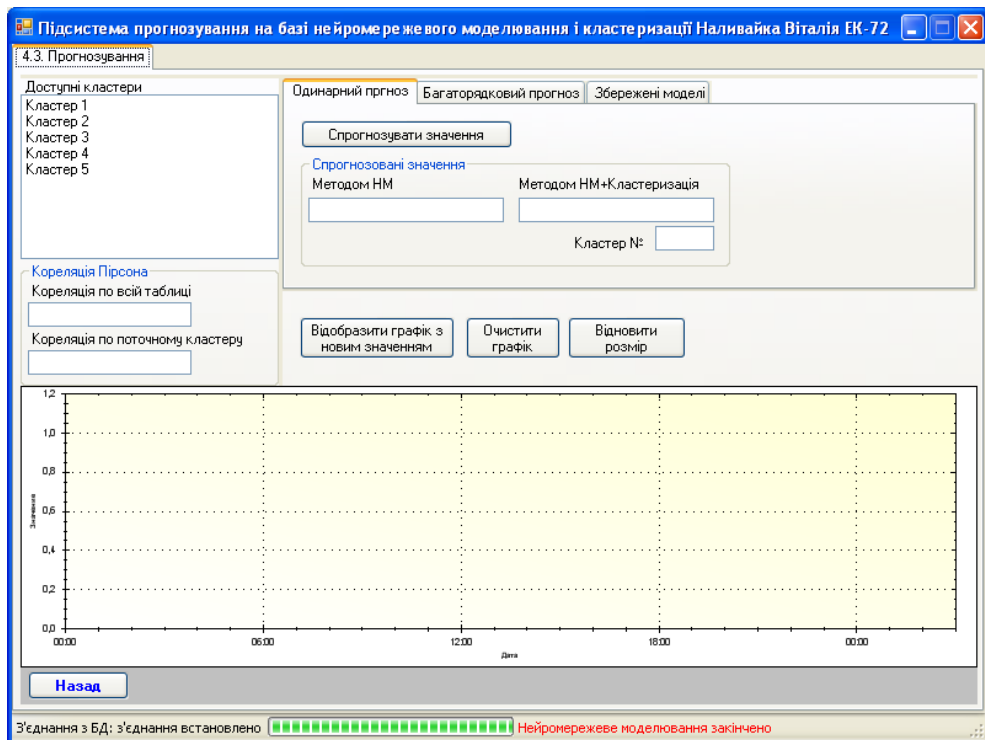


Рисунок 3.16 – Підсистема прогнозування, вкладка «Одинарний прогноз»



Рисунок 3.17 – Задання значень (тіків), «Однорядковий прогноз»

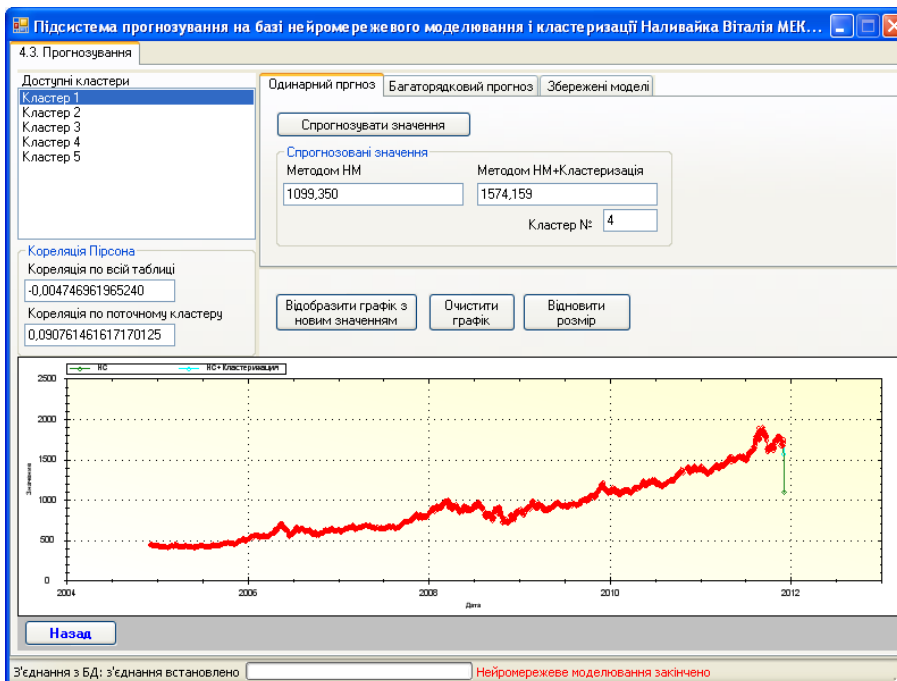


Рисунок 3.18 – Відображення кореляції та графічне представлення результату прогнозування

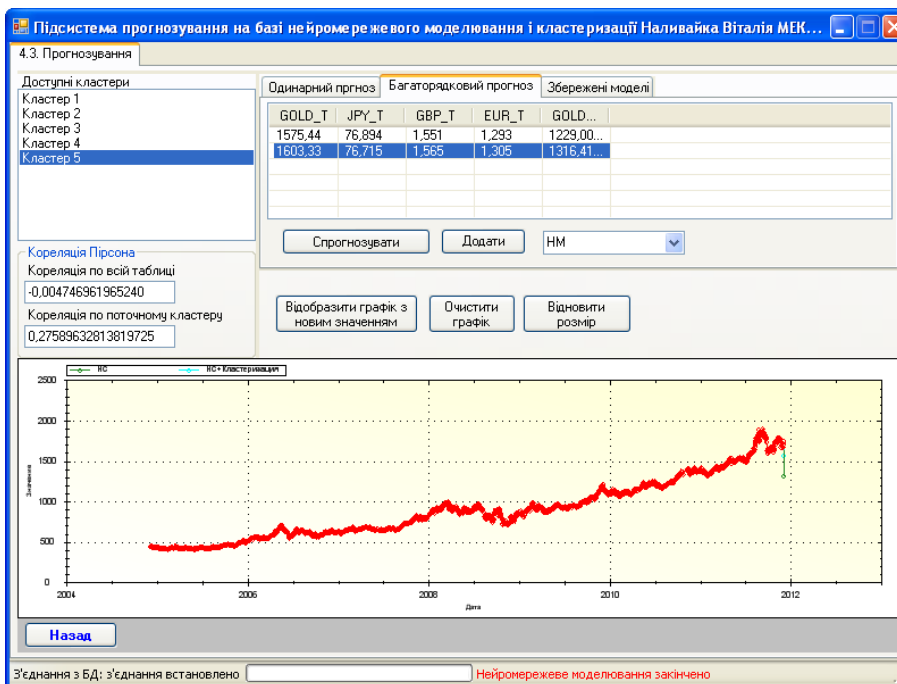


Рисунок 3.19 – Результати прогнозування, «Багаторядковий прогноз»

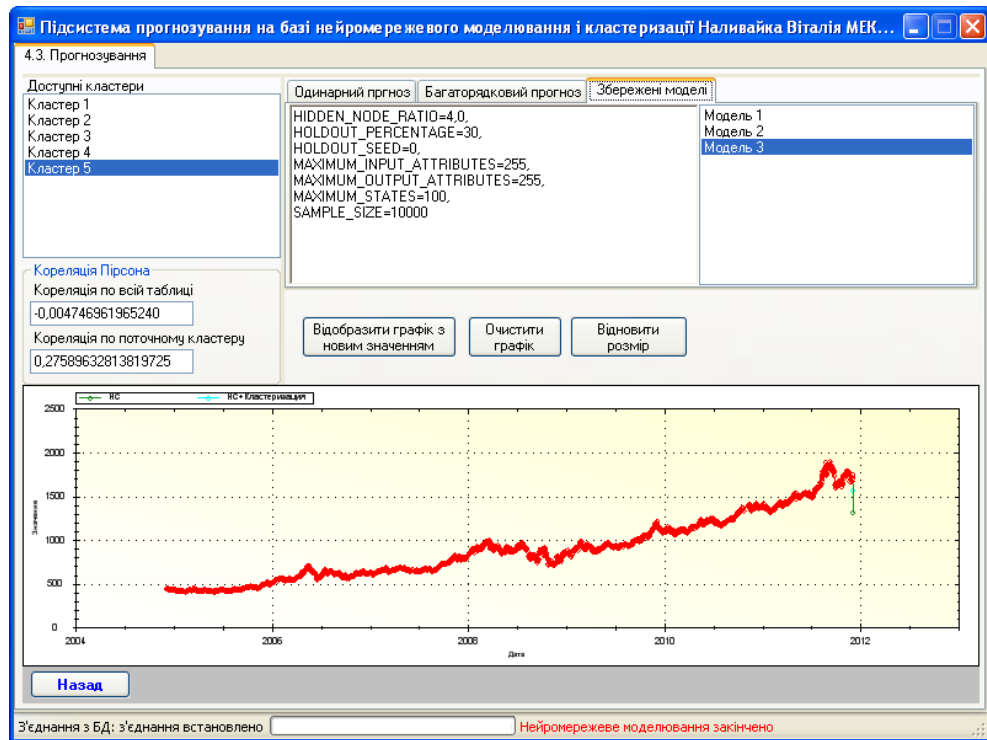


Рисунок 3.20 – Результати прогнозування, «Збережені моделі»