

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧОРНОМОРСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ПЕТРА МОГИЛИ

Васильєв Михайло Олександрович

УДК 004.8

**Інтелектуальна система фінансового прогнозування на основі
технологій машинного навчання**

122 – Комп'ютерні науки

Автореферат
магістерської наукової роботи на здобуття освітньої кваліфікації
«Магістр комп'ютерних наук»

Миколаїв – 2020

Магістерська наукова робота є рукопис.

Робота виконана в Чорноморському національному університеті імені Петра Могили Міністерства освіти і науки України на кафедрі інтелектуальних інформаційних систем

Науковий керівник: д.т.н., професор
Гожий О.П.

Рецензент: к.т.н., доцент
Калініна І.О.

Захист відбудеться «24» лютого 2020 р. о 9³⁰ год. на засіданні екзаменаційної комісії (ауд. 2-403) у Чорноморському національному університеті імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

З магістерською науковою роботою можна ознайомитися в бібліотеці Чорноморського національного університету імені Петра Могили за адресою: 54003, м. Миколаїв, вул. 68-ми Десантників, 10.

Автореферат представлений «__» лютого 2020 р.

Секретар
екзаменаційної комісії,
к.пед.н., доцент

Н. М. Болюбаш

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Фінансовий ринок є найважливішим частиною світової економіки. Структура цього ринку здійснює перерозподіл капіталу та має складну структуру. Подібна розгалужена структура впливає з необхідності здійснення координації та організації взаємодії осіб, які бажають інвестувати кошти з власниками інвестиційних проектів. Таким чином, фінансовий ринок здійснює найважливіше завдання в рамках ділового циклу, так як забезпечує залучення вільного капіталу у вигляді інвестицій.

Прийняття рішень на фінансовому ринку, а також розробка інвестиційних стратегій - найважливіше завдання, що стоїть як перед приватними інвесторами, так і перед інвестиційними компаніями. Від якості вирішення цього завдання безпосередньо залежить успішність інвестиційної діяльності, здійснюваної учасником фінансових ринків.

На сьогоднішній день в діловій і науковій літературі опубліковано безліч статей і досліджень, присвячених питанням інвестування на фінансовому ринку. Математичний інструментарій, застосовуваний для вирішення завдань в інвестиційній сфері, великий і багатогранний. Він включає, як відносно прості методи аналізу, такі як згладжування, використання індикаторів технічного аналізу, кореляційний і регресійний аналізи, так і складні методи, які стоять на стику декількох наукових областей. До таких методів можна віднести побудову нейронних мереж, генетичних алгоритмів, моделей опорних векторів і т.д.

Метою магістерської роботи є дослідження та підвищення ефективності методів прогнозування та моделювання поведінки цінних котирувань на валютному ринку на основі моделей та методів машинного навчання.

Об'єктом дослідження є процеси фінансового прогнозування в задачах прийняття рішень.

Предметом дослідження є методи та моделі прогнозування фінансових показників на основі методів машинного навчання.

Практична значимість роботи полягає у можливості використання розробленої програми для прогнозування курсу валюти на валютних біржах

Результати даної магістерської наукової роботи було опубліковано:

1. Васильєв М.О., Гожий О.П. Короткострокове прогнозування валютних коливань із використанням машинного навчання. Могілянські читання – 2019: тези доповідей: Комп'ютерні науки. Технічні науки, Миколаїв: ЧНУ, 11-16 листопада, 2019. – С. 13-15.

2. Васильєв М.О., Гожий О.П. Використання KNN для пошуку цінових шаблонів на валютних ринках – 2020: IV International Scientific and Practical Conference «Priority directions of science development», Львів, 3-4 лютого.

Магістерська наукова робота складається із вступу, 7 розділів, висновків, додатків. Загальний обсяг роботи складає 112 сторінки, 61 рисуноків, 8 таблиць та 33 посилань на літературні джерела.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі магістерської наукової роботи обґрунтовано актуальність обраної теми, сформульовано мету і задачі дослідження, визначено предмет та об'єкт дослідження.

Фінансовий ринок є найважливішим частиною світової економіки. Структура цього ринку здійснює перерозподіл капіталу та має складну структуру. Подібна розгалужена структура впливає з необхідності здійснення координації та організації взаємодії осіб, які бажають інвестувати кошти з власниками інвестиційних проектів. Таким чином, фінансовий ринок здійснює найважливіше завдання в рамках ділового циклу, так як забезпечує залучення вільного капіталу у вигляді інвестицій.

Прийняття рішень на фінансовому ринку, а також розробка інвестиційних стратегій - найважливіше завдання, що стоїть як перед приватними інвесторами, так і перед інвестиційними компаніями. Від якості вирішення цього завдання безпосередньо залежить успішність інвестиційної діяльності, здійснюваної учасником фінансових ринків.

На сьогоднішній день в діловій і науковій літературі опубліковано безліч статей і досліджень, присвячених питанням інвестування на фінансовому ринку. Математичний інструментарій, застосовуваний для вирішення завдань в інвестиційній сфері, великий і багатогранний. Він включає, як відносно прості методи аналізу, такі як згладжування, використання індикаторів технічного аналізу, кореляційний і регресійний аналізи, так і складні методи, які стоять на стику декількох наукових областей. До таких методів можна віднести побудову нейронних мереж, генетичних алгоритмів, моделей опорних векторів і т.д.

У першому розділі розглядається поняття валютного ринку, його структура та класифікація операцій.

Валютний ринок - це система економічних та організаційних відносин, які пов'язані з операціями купівлі-продажу іноземних валют та платіжних документів в іноземних валютах.

Валютний ринок надає широке коло операцій щодо зовнішньоторговельних розрахунків, міграції капіталів, робочої сили, туризму, які передбачають обіг іноземної валюти покупцями, продавцями, посередниками, банківськими установами та фірмами. Валютний ринок можна розділити 4 групи:

- юридичні та фізичні особи, що працюють у різноманітних сферах зовнішньоекономічної діяльності;
- державні установи, які зосереджені в центральних банках та казначействах окремих країн;
- комерційні банківські установи, які надають валютне обслуговування зовнішніх клієнтів, особливо брокерські контори;
- валютні відділи та валютні біржі товарних та фондових бірж.

Валютний ринок виконує наступні функції, в яких полягає його призначення і економічна роль. Функції валютного ринку:

- створення суб'єктами валютних відносин передумов для здійснення міжнародних платежів за капітальними і поточними розрахунками та сприяння розвитку зовнішньої торгівлі;
- забезпечення умов та механізмів реалізації валютної політики держави;

- формування та балансування попиту і пропозиції валюти і регулювання валютного курсу;
- страхування валютних ризиків;
- диверсифікація валютних резервів, банків, підприємств, держав, регулювання валютних ресурсів.

У другому розділі проводиться аналіз та дослідження сучасного стану задачі прогнозування валютного ринку. Можна виділити наступні підходи до прогнозування валютного ринку: технічний та фундаментальний аналіз, використання методів машинного навчання.

Технічний аналіз - це метод прогнозування цін за допомогою аналізу графіків рухів ціни за попередні періоди часу.

Методи технічного аналізу на ринку включають в себе графічний аналіз, і як вже впливає з назви, працює виключно з поданням ціни активів на графіку. Тут не використовуються ніякі індикатори (хоча їх застосування не виключено в різних ситуаціях).

Фундаментальний аналіз включає оцінку наступних факторів, що впливають на валютні курси:

- показники економічного зростання (валовий національний продукт, обсяги промислового виробництва та ін.);
- стан торгового балансу, ступінь залежності від зовнішніх джерел сировини;
- зростання грошової маси на внутрішньому ринку;
- рівень інфляції і інфляційні очікування;
- рівень відсоткової ставки;
- платоспроможність країни і довіра до національної валюти на світовому ринку;
- спекулятивні операції на валютному ринку;
- ступінь розвитку інших секторів світового фінансового ринку, наприклад ринку цінних паперів, що конкурує з валютним ринком.

В роботі будуть розглянуті методи машинного навчання для прогнозування валютних ринків.

У третьому розділі розглянута обробка даних, кластеризація часових рядів та метрика DTW. Попередня обробка і очищення даних - це важливі завдання, які необхідно виконати, перш ніж набір даних можна буде використовувати для навчання моделі. Необроблені дані часто спотворені і ненадійні, і в них можуть бути пропущені значення. Використання таких даних при моделюванні може призводити до невірних результатів. Ці завдання є частиною процесу обробки і аналізу даних групи і зазвичай мають на увазі початкове вивчення набору даних, що використовується для визначення і планування необхідної попередньої обробки.

Кластеризація - це завдання поділу сукупності чи точок даних на ряд груп, так що точки даних у одних і тих же групах більше схожі на інші точки даних цієї ж групи, ніж на інші групи. Простими словами, мета - відокремити групи з подібними ознаками та розподілити їх у кластери.

Динамічне викривлення часу - це алгоритм, який використовується для вимірювання подібності між двома послідовностями, які можуть відрізнятися за часом або швидкістю.

DTW обчислює найкращий глобальний розподіл між двома послідовностями довжиною n та m відповідно. На практиці спочатку обчислюють першу $n \times m$ матрицю попарних відстаней, а потім вирівнюють координати всередині обох рядів. Вартість оптимального вирівнювання остаточно обчислюється за допомогою динамічного формулювання програмування. Це може бути дуже трудомісткою операцією, і зазвичай, обмеження встановлюються на допустимі деформації

Агрегація часових рядів може розглядатися як процес "скорочення даних" в тому сенсі, що він узагальнює набір часових рядів. На практиці такий процес дуже застосовується, особливо в кластеризації чи класифікації.

Агрегація часових рядів являє собою пошук барицентрів. Для пошуку барицентрів використовувалися різні алгоритми, а за метрику точності середні DTW відстань. Результати дослідження представлені в таблиці 1.

Таблиця 1 – Порівняння алгоритмів пошуку барицентрів.

Алгоритм	Середня DTW відстань
Евклідовий барицентр	2.201
DTW	1.603
Soft DTW ($\gamma = 1$)	1.861
Soft DTW ($\gamma = 0.1$)	1.605
Soft DTW ($\gamma = 0.01$)	1.600

У четвертому розділі досліджуються методи машинного навчання для класифікації часових рядів.

Класифікація часових рядів є однією з найбільших наукових областей за останні кілька років, головним чином завдяки великій кількості практичних застосувань у різних областях.

Набір даних часових рядів - це набір даних, який представляє деякі вимірювання величини за певний проміжок часу.

Були досліджені наступні методи машинного навчання:

- дерево рішень;
- логістична регресія;
- машина опорних векторів;
- багат шаровий перцептрон;
- наївний класифікатор Баєса;
- градієнтний бустинг.

Результати дослідження методів машинного навчання для вирішення задачі класифікації зображені в таблиці 2.

Таблиця 2 – Результати алгоритмів для задач класифікації

Алгоритм	Точність класифікації шаблонів, %	Точність класифікації цінового напрямку %	Площа ROCAUC
Дерево рішень	22.22	68.52	0,69
Логістична регресія	25.93	74.07	0,75
Багатошаровий перцептор	35.19	72.2	0,73
Машина опорних векторів	38.89	75.93	0,76
Наївний класифікатор Баєса	37.03	70.37	0,71
Гرادієнтний бустинг	27.78	64.81	0,65

В п'ятому розділі розглядається реалізація програми для прогнозування напрямку валютного курсу. Програма складається з 4 модулів:

- блоку отримання даних;
- блоку обробки даних;
- блоку прогнозування;
- блоку логічного виведення.

На рис. 1. Зображений результат прогнозування валютної пари GBPUSD на 5 днів.

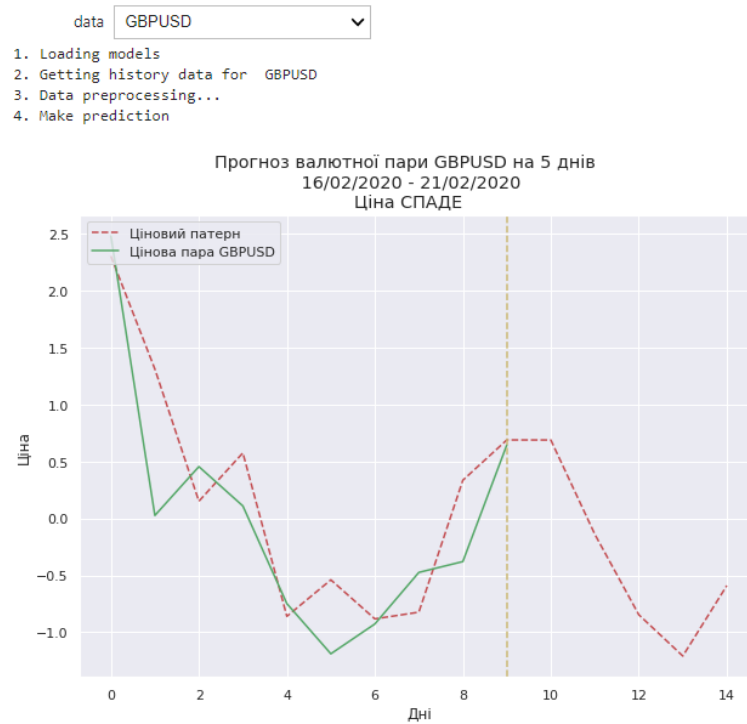


Рис. 1. Результати роботи програми

В результаті дослідження методів класифікації часових рядів був обраний метод SVM, оскільки він показав найкращу точність класифікації порівняно з іншими методами. Для агрегації кластеру KNN був обраний метод Soft-DTW. В результаті тестування системи в основу яких були покладені дані алгоритми на тестових історичних даних – дані, котрі не використовувалися для навчання моделей були отримані наступні результати прогнозування на валютних парах (рис. 2).

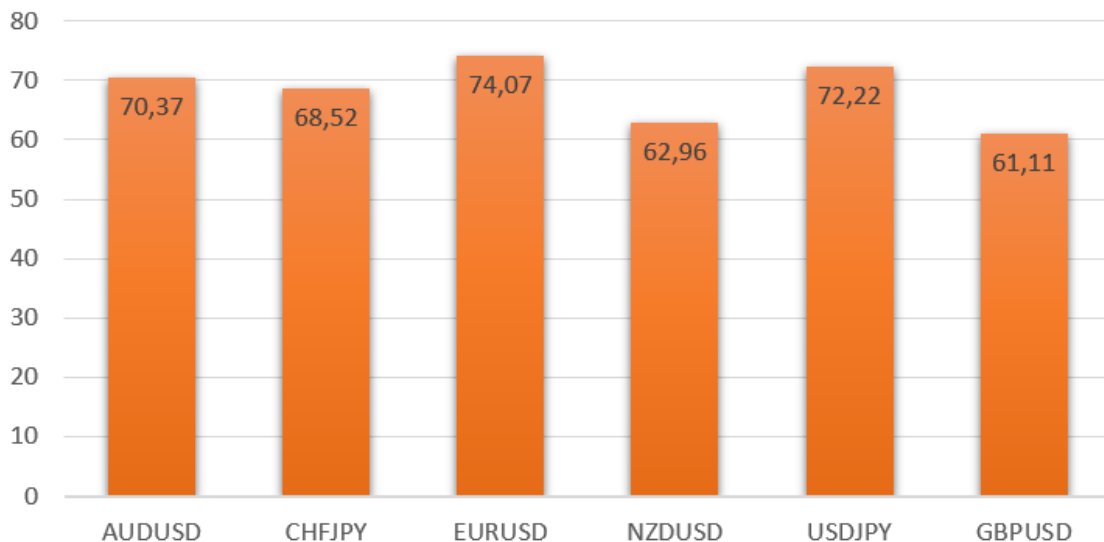


Рис. 2. Результати тестування програми на історичних даних за 2019 рік

В методичному розділі розроблено практичну роботу на тему «Класифікація даних методом KNN».

В спеціальному розділі до магістерської наукової роботи розглянуті основні положення охорони праці та безпеки у надзвичайних ситуаціях на підприємстві.

ЗАГАЛЬНІ ВИСНОВКИ

Прийняття рішень на фінансовому ринку, а також розробка інвестиційних стратегій - найважливіше завдання, що стоїть як перед приватними інвесторами, так і перед інвестиційними компаніями. Від якості вирішення цього завдання безпосередньо залежить успішність інвестиційної діяльності, здійснюваної учасником фінансових ринків.

З самого початку фінансових операцій на біржах люди розробляли методи для прогнозування вартості активів у майбутньому. З розвитком обчислювальних технологій та штучного інтелекту, точність методів зростає кожного дня.

У результаті даного дослідження було:

- проведений аналіз сучасних методів та підходів прогнозування показників фінансового ринку;
- проаналізовані існуючі методи для прогнозування показників фінансових ринків на основі історичних даних;
- досліджено методи класифікації та кластеризації часових рядів в задачах машинного навчання;
- досліджено метрику DTW. Знайдені цінових шаблонів;
- розроблений алгоритм прогнозування валютного курсу на основі цінових шаблонів;
- розроблено програмне забезпечення для прогнозування валютного курсу.

АНОТАЦІЯ

Васильєв М.О. «Інтелектуальна система фінансового прогнозування на основі технологій машинного навчання». – На правах рукопису.

Магістерська наукова робота на здобуття освітньої кваліфікації «Магістр комп'ютерних наук» – Чорноморський національний університет імені Петра Могили, Миколаїв, 2020.

Магістерська робота спрямована на дослідження методів машинного навчання для кластеризації та класифікації часових рядів. Розглянуті алгоритми машинного навчання: KNN, SVM, MLP, Дерево рішень, Наївний класифікатор Басса, Градієнтний бустинг. Практичне значення результатів дослідження та розроблення полягає у можливості їх запровадження в практику для прогнозування динамічних фінансових часових рядів.

Пояснювальна записка магістерської роботи складається зі вступу, п'яти розділів, висновків. У вступі визначається актуальність теми, сформульовані мета, об'єкт, предмет та завдання дослідження магістерської роботи. У першому розділі досліджується ринок курсів акцій та валют, що являться як об'єктом дослідження; проводиться аналіз його економічних показників та основних методів аналізу. У другому розділі проводиться аналіз існуючих методів та підходів прогнозування часових рядів для короткострокового прогнозування курсу валют. У третьому розділі досліджується метод кластеризації часових рядів та агрегування результатів (пошук барицентрів). Порівнюються різні алгоритми пошуку барицентрів. У четвертому розділі розглянуті основні алгоритми машинного навчання для задачі класифікації часових рядів. У п'ятому розділі описується архітектура розробленої програми та узагальнюються результати дослідження. В цілому, магістерська робота без додатків містить 104 сторінки, 61 рисунків, 8 таблиць, 33 джерела посилання.

Ключові слова: *часовий ряд, фінансовий ринок, кластеризація, препроцесинг, машинне навчання, класифікація, ціновий шаблон, градієнтний бустинг, дерева рішень, глибокі нейронні мережі*

ABSTRACT

of the student of the group 601m of the
Black Sea National University named after Petro Mohyla

Mykhailo Vasyliiev

Master Thesis on the topic:

«Intellectual financial forecasting system based on machine learning technologies»

The master's thesis is aimed at studying machine learning methods for clustering and classifying time series. The machine learning algorithms are considered: KNN, SVM, MLP, Decision Tree, Naive Bayes Classifier, Gradient Boosting. The practical significance of the R&D results is the ability to put them into practice to predict dynamic financial time series.

The explanatory note of the master's thesis consists of an introduction, five sections, conclusions. The introduction defines the relevance of the topic, formulated the purpose, object, subject and objectives of the master's study. The first section examines the market for stock and currencies that are the subject of research; its economic indicators and basic methods of analysis are analyzed. The second section analyzes the current time series forecasting methods and approaches for short-term currency exchange rates. The third section explores the method of clustering time series and aggregating results (search for bar centers). Comparison of different algorithms for search of bar centers. The fourth section discusses basic machine learning algorithms for the time series classification problem. The fifth section describes the architecture of the program developed and summarizes the research findings. In total, the master's work without applications contains 104 pages, 61 figures, 8 tables, 33 sources of links.

Keywords: *time series, financial market, clustering, preprocessing, machine learning, classification, price pattern, gradient boosting, decision trees, deep neural networks*