***Лучишин Богдан Андрійович, магістрант,***

*Західноукраїнський національний університет, м. Тернопіль****Науковий керівник:******Биковий Павло Євгенович,***

***кандидат технічних наук, доцент,*** *Західноукраїнський національний університет, м. Тернопіль*

**ПОРІВНЯЛЬНИЙ АНАЛІЗ МЕТОДІВ МАШИННОГО НАВЧАННЯ ТА ЧАСОВИХ РЯДІВ ДЛЯ ПРОГНОЗУВАННЯ ФОНДОВОГО РИНКУ: ПІДХІД БОТА DISCORD**

У статті розглянуто порівняльний аналіз моделей машинного навчання та методів часових рядів для прогнозування руху фондового ринку. Оцінено такі моделі, як RandomForestRegressor, GradientBoosting, Support Vector Regressor (SVR) та Long Short-Term Memory (LSTM). Моделі було застосовано до даних цін акцій компанії Apple Inc., і їх ефективність оцінювалась за метриками точності та прогнозування. Аналіз реалізовано в рамках роботи бота Discord, що надає користувачам прогнози в реальному часі.

Прогнозування фондового ринку є складним завданням, яке викликає великий інтерес у фінансовій аналітиці та машинному навчанні. Завдяки високій волатильності та випадковості поведінки цін на акції, створення точних моделей прогнозування потребує використання сучасних методів аналізу часових рядів і машинного навчання. Серед перспективних підходів у цій галузі є нейронні мережі, зокрема LSTM, що добре підходять для обробки послідовних даних завдяки своїм властивостям збереження тривалих залежностей. Ансамблеві методи, такі як випадкові ліси та бустинг, також широко використовуються для підвищення стабільності та точності прогнозування.

У прогнозуванні фінансових часових рядів часто застосовуються моделі на основі Long Short-Term Memory (LSTM), що дозволяють ефективно обробляти послідовні залежності в даних. LSTM є потужним інструментом для збереження інформації протягом тривалих періодів, що є важливим для точного прогнозування фінансових змін. Дослідження показують, що LSTM може досягати високої точності при прогнозуванні фінансових ринків, таких як фондові індекси, хоча ці моделі потребують значних обчислювальних ресурсів та налаштування параметрів для досягнення оптимальної ефективності [1, 2].

За умов обмежених обчислювальних ресурсів, ефективнішими можуть бути ансамблеві методи, зокрема RandomForestRegressor, GradientBoostingRegressor та SVR. RandomForestRegressor використовує ансамбль рішень дерев, що мінімізує ризик перенавчання і забезпечує стабільність прогнозів навіть при наявності шуму в даних [3]. GradientBoostingRegressor покроково коригує помилки, що дозволяє значно підвищити точність моделі [3]. SVR є потужним інструментом для роботи з нелінійними та високорозмірними даними, що робить його ефективним для обробки складних фінансових даних [1, 3].

Для дослідження використовувалися історичні дані про ціни акцій AAPL, що охоплюють період з 13 грудня 2010 року по 20 травня 2024 року (3381 запис). Дані були поділені на 80% для навчання моделей і 20% для тестування, і структуровані в датафрейм для подальшої обробки, що включала кілька етапів: вибір ознак, зміна типів даних та масштабування.

Метод LSTM базується на рекурентних нейронних мережах (RNN), здатних моделювати довгострокові залежності в часових рядах. Для оцінки ефективності LSTM модель була навчена з різною кількістю епох: 200 та 2000. Під час цього процесу використовувалося масштабування даних за допомогою MinMaxScaler для стабілізації навчання.

Ансамблеві методи використовують комбінацію кількох моделей для підвищення точності. У дослідженні використовувалися RandomForestRegressor, GradientBoostingRegressor та SVR. Моделі були навчені окремо, а результати прогнозів об'єднано за допомогою VotingRegressor.

Метод LSTM не продемонстрував високої точності у прогнозуванні цін акцій AAPL. Середньоквадратична помилка (MSE) для моделі з 200 епохами склала 20412.54, а середня абсолютна помилка (MAE) — 142.79. При збільшенні кількості епох до 2000, ці значення зросли, що вказує на перенавчання моделі. Негативні значення R² (-2371406.73 для 200 епох) свідчать про те, що модель не змогла адекватно відобразити залежності в даних.

Ансамбль моделей, що об'єднує RandomForestRegressor, GradientBoostingRegressor та SVR, показав набагато кращі результати. Середньоквадратична помилка склала лише 15.02, а середня абсолютна помилка — 2.58. Коефіцієнт детермінації R² дорівнював 0.995, що свідчить про здатність моделі точно пояснювати варіативність у даних. Високий відсоток зміни ціни (697.23%) та висхідний тренд вказують на позитивні прогнози, хоча значна волатильність (56.15%) підтверджує наявність ризиків.

Порівняльний аналіз показав, що метод ансамблю моделей значно перевищує метод LSTM за точністю та стабільністю прогнозів. Моделі RandomForestRegressor, GradientBoostingRegressor та SVR продемонстрували кращу продуктивність у прогнозуванні фондового ринку, тоді як LSTM потребує значних обчислювальних ресурсів і не забезпечив задовільних результатів у цій задачі.

Майбутні дослідження можуть бути спрямовані на інтеграцію більш складних нейронних мереж та гібридних моделей для підвищення точності прогнозів.

**Література**

1. Zhao Z., et al. Unlocking the Power of LSTM for Long Term Time Series Forecasting. 2024. [Електронний ресурс]. – Режим доступу: <https://ar5iv.labs.arxiv.org/html/2408.10006>.

2. Mroua M., Lamine A. Financial time series prediction under Covid-19 pandemic crisis with LSTM network. 2024. [Електронний ресурс]. – Режим доступу: <https://ar5iv.labs.arxiv.org/html/2201.08218>.

3. Chen W., Hussain W., Cauteruccio F., Zhang X. Deep Learning for Financial Time Series Prediction: A State-of-the-Art Review of Standalone and Hybrid Models. Computer Modeling in Engineering & Sciences, 2024, 139(1), 187-224. [Електронний ресурс]. – Режим доступу: <https://doi.org/10.32604/cmes.2023.031388>.