

Застосування методів машинного навчання до прогнозування часових фінансових рядів

Вінницький національний технічний університет

Анотація

У даному дослідженні було розглянуто алгоритми машинного навчання для прогнозування часових рядів. У ході роботи було розроблене програмне забезпечення для прогнозування кількості проданих за день акцій компанії "Microsoft". Також було оцінено роботу навчених моделей по графічному представленню результатів передбачення разом із реальними значеннями кількості проданих акцій.

Ключові слова: часові ряди, машинне навчання, дерево рішень, ансамблеві алгоритми, додаткове дерево.

Abstract

In this study, machine learning algorithms for time series forecasting were considered. In the course of the work, software was developed for forecasting the number of Microsoft shares sold per day. The performance of the trained models was also evaluated by graphically presenting the prediction results together with the actual values of the number of shares sold.

Keywords: time series, machine learning, decision tree, ensemble algorithms, extra tree.

ВСТУП

Часові ряди[1][2] загалом є набором даних, точок із деякими характеристиками. Серед цих характеристик обов'язковими є деяка цільова характеристика і час до якого прив'язана точка, таким чином можна зіставити графік, на якому видно зміну цільової характеристики із часом. Головною метою аналізу часових рядів є прогнозування цільової характеристики.

Цільовою характеристикою фінансових часових рядів зазвичай є ціна, будь то ціна на нерухомість, якийсь товар чи акції, хоча також може бути, наприклад, кількість акцій, проданих за деякий період. У будь-якому разі, влучне прогнозування часових рядів відкриває широкі можливості для різноманітних фінансових операцій. Вирішити таку важливу задачу, для якої потрібно обробляти великі масиви даних краще всього за допомогою машинного навчання.

Машинним навчанням[3] називають дослідницьку галузь, яка зосереджена на розробці, вивченні та розвитку статистичних алгоритмів. Такі алгоритми здатні навчатися на масивах даних, що дає їм змогу виконувати такі типові завдання як "класифікація" та "регресія", при роботі із даними, які подібні їх навчальному набору даних.

Одним із простих та широко розповсюджених методів машинного навчання є "Дерево рішень"[4]. За цим методом створюється модель, яка розглядає входні дані та приймає серію (хоча може і одне) рішень щодо їх характеристик. Кожне рішення впливає на сутність наступного, адже модель складається із вузлів (умов/рішень) та гілок, якими зв'язані вузли, та має деревоподібну будову. На кінцях гілок знаходяться "Листки", які визначають те, що модель повертає на виході. "Дерева рішень" використовуються для класифікації, прогнозування та підтримки прийняття рішень.

У простого метода є свої недоліки, тому крім нього є сенс спробувати використати ансамблевий метод. Ансамблі[5] використовують декілька об'єднаних алгоритмів для виконання завдання, навчаючись незалежно разом вони перекидають недоліки один одного.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

У дослідженні були взяті дані із відкритого джерела, на платформі Kaggle, а саме "Microsoft Stock – Time Series Analysis"[6]. Даний набір даних спеціально був зіставлений для розробки/випробування методів прогнозування часових рядів. Цільовою характеристикою було взято кількість акцій, якими торгували за день.

Першою побудованою моделлю передбачення була модель, тренувана за алгоритмом дерева рішень.

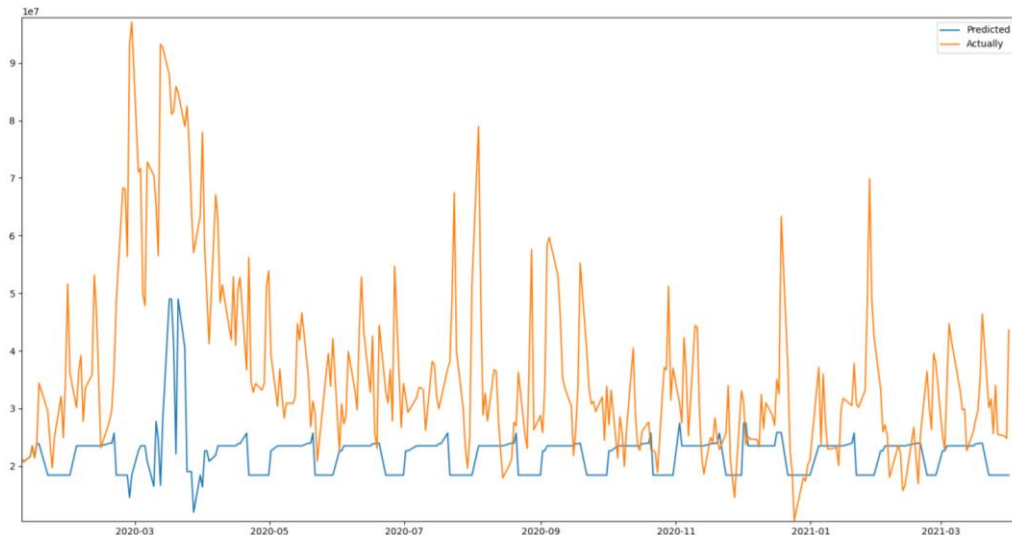


Рисунок 1 – Передбачення акцій компанії “Microsoft” деревом рішень.

Модель притримується дивної сезонності, яка на перший погляд не корелює із реальними змінами значення кількості акцій, проте по ній видно період максимального попиту на акції. Хоча загалом, значення прогнозу близьке до реального значення, або збігається із ним так рідко, що модель не можна вважати придатною для використання. По характеру прогнозного графіку можна зробити висновок, що модель вважає, що кількість проданих акцій тримається приблизно одного значення, і може лише трохи відхилятися від нього. Використовувати модель у такому вигляді не доцільно.

Результатом аналізу графіку та навчального набору даних був наступний висновок: “на модель сильно впливає остання точка навчального набору, модель вважає її значення рівнем, на околицях якого тримається кількість проданих акцій”. Насправді це не є чимось несподіваним, адже уся суть навчання моделі за даними часового ряду у послідовній обробці даних, адже наступні дані у часовому ряді сильно залежать від актуальної/попередньої ситуації. Для покращення роботи моделі можна перенавчати її перед кожним передбаченням, додаючи до навчальної вибірки точку, яку передбачували попереднього разу.

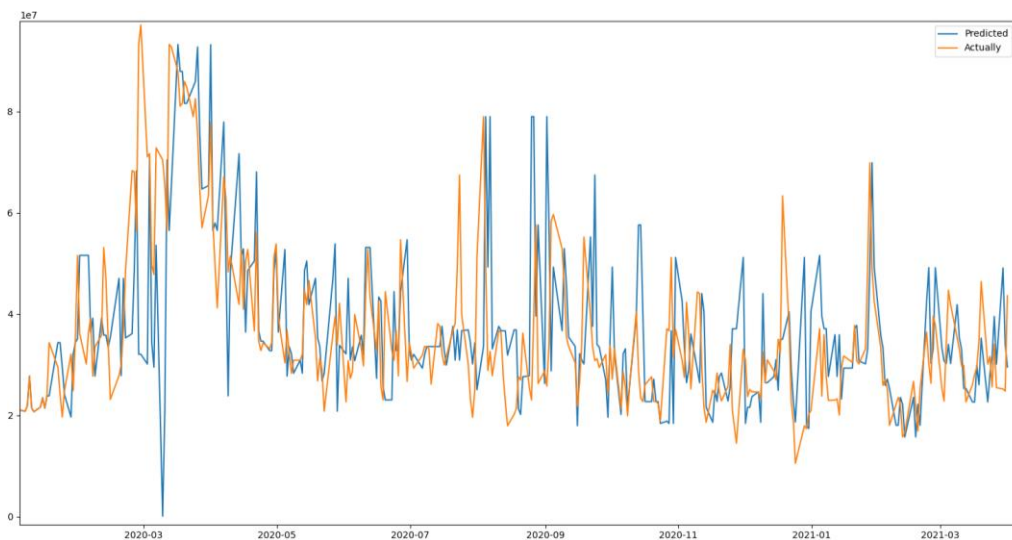


Рисунок 2 – Передбачення акцій компанії “Microsoft” модифікованим деревом рішень.

Ситуація із трендом покращилася, проте точність такої моделі все одно залишає бажати на краще, до того ж деінде з’явилися доволі дивні зростання та спади передбачення акцій. Схоже це усе що можна отримати від такого простого алгоритму.

Наступною моделлю для побудови було взято більш складну, ансамблевую модель. Як саму по собі, так і використовуючи вже досліджену техніку для врахування тренду. Використано було ансамбль, який називається “Додаткове дерево”.

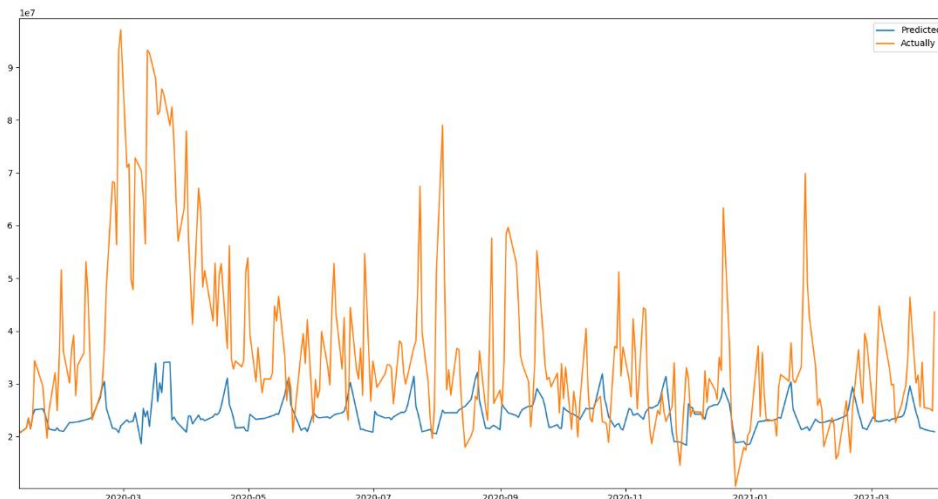


Рисунок 3 – Передбачення акцій компанії “Microsoft” додатковим деревом.

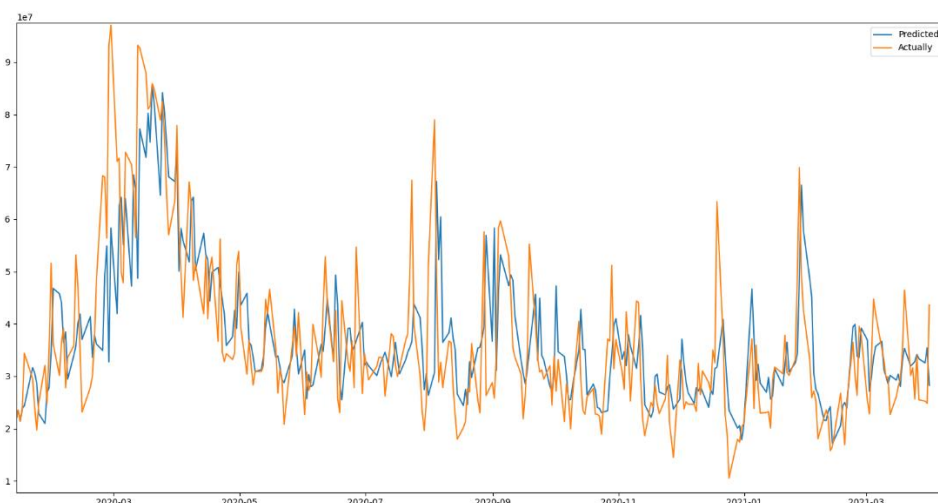


Рисунок 4 – Передбачення акцій компанії “Microsoft” модифікованим додатковим деревом.

Використаний самий по собі метод показав дивні результати. Циклічні відрізки графіку ускладнилися та почали дуже приблизно збігатися по характерам зростань/спадань із реальною ситуацією. Циклічний характер графіку може вказувати на те, що алгоритм відшукав у даних залежність від часу, те що у часових рядах називають сезонністю. Сезонність присутня у товарів чи послуг, попит на які значно підіймається у певний період часу. Проте, на відмінну від дерева рішень, голий алгоритм додаткового дерева не прогнозує значного (максимального) зростання кількості проданих акцій четвертого місяця 20-го року. Модифікований же алгоритм чудово передбачає рівень, до якого близьке значення поданих акції. Дана версія алгоритму показала себе найкраще.

ВИСНОВКИ

У даній роботі було проведено дослідження прогнозування часового ряду методами машинного навчання. У ході дослідження було розроблене програмне забезпечення, яким реалізовувалися чотири моделі передбачення кількості акцій компанії “Microsoft”.

Також було виявлено, що у певних аспектах передбачення часових рядів, більш складний ансамблевий метод може працювати гірше за звичайне дерево рішень, якщо використовувати ансамбль без урахування особливостей даних часового ряду.

Ансамблевий алгоритм “Додаткове дерево”, при використанні із дослідженою технікою для урахування тренду, показав найбільшу та задовільну точність передбачення.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Бакай С. І., Кабачій В. В., Маслій Р. В. Модель прийняття рішень для фінансових часових рядів на основі пари середніх з використанням оцінки різних часових вимірів [Електронний ресурс] – URL: <http://ir.lib.vntu.edu.ua/handle/123456789/21777>
2. Бідюк П. І. Аналіз часових рядів [Електронний ресурс] – URL: <https://ela.kpi.ua/handle/123456789/862>

3. Машинне навчання [Електронний ресурс] - URL: <http://www.mmf.lnu.edu.ua/ar/1739>
4. Ансамблі [Електронний ресурс] – URL: <https://evergreens.com.ua/ua/articles/ensembles.html>
5. Дерево рішень [Електронний ресурс] – URL: <https://ua5.org/algorithm/1976-derevo-rishen.html>
6. Microsoft Stock – Time Series Analysis [Електронний ресурс] – URL: <https://www.kaggle.com/datasets/vijayvkenkitesh/microsoft-stock-time-series-analysis/data>

Червінський Ростислав Олександрович – студент групи ЗАКІТР-23м, факультет комп'ютерних систем і автоматизації, Вінницький національний технічний університет, м. Вінниця, e-mail: rostislav.chervinskiy@gmail.com

Науковий керівник: Кабачій Владислав Володимирович – к. т. н., доцент кафедри Автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, e-mail: kabachij.v.v@vntu.edu.ua

Chervinskyu Rostislav Oleksandrovych – student of group ЗАКІТР-23m, Faculty of Computer Systems and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: rostislav.chervinskiy@gmail.com

Supervisor: Kabachy Vladyslav Volodymyrovych – Ph.D., Associate Professor of Automation and Intelligent Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: kabachij.v.v@vntu.edu.ua