

## МЕТОДИ ТА ЗАСОБИ ІНФОРМАЦІЙНОЇ ІНТЕЛЕКТУАЛЬНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ КРИПТОВАЛЮТ

Вінницький національний технічний університет, Україна

### Анотація

Метою роботи є підвищення точності прогнозування курсу цифрових валют за рахунок створення Python-засобів автоматизації методів моделювання цього курсу та вибору оптимальної моделі серед багатьох видів моделей машинного навчання, у т.ч. нейронних мереж та часових рядів. Ефективність технології продемонстрована на прогнозуванні біткоїну за середньодобовими даними у 2021 році. Технологія є універсальною і легко адаптується до будь-якої криптовалюти.

**Ключові слова:** інформаційна інтелектуальна система, цифрова валюта, криптовалюта, нейронна мережа, машинне навчання, біткоїн.

### Abstract

The objective of this work is to provide solution to increase the accuracy of forecasting of exchange rate of digital currencies by creating tools in Python programming language to automate the methods of modeling rate of cryptocurrencies and choosing the optimal model among many types of machine learning models, including neural networks and time series. The effectiveness of the technology is demonstrated by forecasting Bitcoin price changes on average daily data in 2021. The technology is universal and easily adapts to any cryptocurrency.

**Key words:** information intellectual system, digital currency, cryptocurrency, neural network, machine learning, bitcoin.

### Вступ

Обсяг ринку криптовалют вже складає кілька трильйонів доларів США і продовжує зростати. Відомо більше 10 тисяч криптовалют і ця кількість зростає. Все більша кількість людей хоче знати як же можна прогнозувати курс криптовалют. У роботі [1] проведено глибокий аналіз цієї задачі. Зазначені її складнощі та зроблено огляд підходів, моделей, методів, технологій та засобів її розв'язання. Але відкритим є питання створення достатньо зручних засобів для застосування найбільш популярних методів аналізу та прогнозування курсу криптовалют.

Метою даного дослідження є підвищення точності прогнозування курсу цифрових валют за рахунок створення Python-засобів автоматизації методів моделювання цього курсу та вибору оптимальної моделі серед багатьох видів моделей машинного навчання, у т.ч. нейронних мереж та часових рядів.

### Розв'язання задачі

Курс криптовалюти можна прогнозувати як часовий ряд, для чого використати моделі авторегресії та проінтегрованого ковзного середнього (англ. скорочено: «ARIMA») та їх аналоги або Facebook Prophet. Інший підхід щодо прогнозування – це класична задача машинного навчання з вчителем, коли по багатьох ознаках передбачається вихідна. Для цільової ознаки роблять зсув у час назад і тоді по минулих значеннях з'являється можливість прогнозувати майбутнє. За цих умов, можна використовувати класичні моделі машинного навчання: лінійну регресію, дерева рішень, нейронні мережі та ін. Важливо правильно формувати тренувальний, валідаційний (для відбору оптимальної моделі) та тестовий (для фінального оцінювання точності прогнозування) датасети.

Під час розв'язання задачі методами машинного навчання цільову ознаку «*Target*» слід обчислювати по курсу в кінці кожної доби «*Close*» за такою формулою:

$$Target_t = Close_{t-d}, \quad (1)$$

де  $t$  – час, днів;  $d$  – кількість днів, на які слід навчити модель робити прогноз (зсув робиться саме назад, тоді з'являється можливість по значеннях ознак у  $t$  добу прогнозувати цільову ознаку у  $t+d$  добу).

Ще однією особливістю моделей часових рядів є те, що і тестовий, і валідаційний датасети вибираються в кінці даних спостережень. Основна задача валідаційного датасету – допомогти відібрати оптимальну модель, яка найкраще буде прогнозувати майбутні значення, а не значення в минулому.

Як правило, тестовими даними є останні  $d$  («forecasting\_days») значень ряду спостережень. Валідаційними (для перевірки) є наступні за ними, якщо рухатись в минуле.

Було здійснено систематизацію відомих методів машинного навчання [2, 3]: Facebook Prophet; ARIMA; класичні моделі машинного навчання – лінійна регресія, дерева рішень, нейронні мережі та ін.

Розроблено такий алгоритм інтелектуальної інформаційної технології:

1. Проаналізувати та підготувати дані, у т.ч. підготувати вибірку даних (часто варто відкинути дані за доковідний період чи період до певного форку, оскільки закономірності до того часу суттєво відрізняються від сучасних), інтерполювати чи відкинути пропущені дані та ін.

2. Синтез ознак з використанням знань предметної області та з використанням типових для часових рядів бібліотек, наприклад TSFRESH.

3. Формування цільової ознаки «Target» (1).

4. Передоброблення даних – здійснити стандартизацію ознак.

5. Формування тренувального, валідаційного та тестових датасетів для різних типів моделей.

6. Вибір і тренування моделей (у т.ч. їх ансамблів) на тренувальному датасеті («train»). Прогнозування за кожною ідентифікованою моделлю валідаційних даних («valid»).

7. Вибір оптимальної моделі за певною метрикою.

8. Тренування оптимальної моделі за розширеним датасетом («train»+«valid»). Прогнозування за ідентифікованою моделлю тестових даних («test»).

9. Оцінювання точності прогнозу тестових даних.

10. Аналіз точності, аналіз викидів (суттєвих відхилень), аналіз важливості ознак різними способами. За потреби – видалення неважливих ознак чи інше постоброблення та перехід на п. 1 чи 2.

На Python здійснена автоматизація тюнінгу типових методів машинного навчання з можливістю їх розширення та можливістю автоматичної та ручної оптимізації [2].

Налаштування параметрів ARIMA здійснено з використанням автоматичного методу SARIMAX. Також, передбачено ручне налаштування за результатами аналізу автокореляційних функцій ряду та його першої і другої різниць.

Для Facebook Prophet задано такі параметри моделі:

– Стандартні параметри річної та тижневої сезонності встановлено як відсутні, для того, щоб мати можливість їх задати вручну окремо і здійснити контрольований вибір їх оптимальних параметрів: `yearly_seasonality = False, weekly_seasonality = False;`

– Денна сезонність відсутня (можливо вона і є, але за щодобовими даними це ідентифікувати неможливо): `daily_seasonality = False;`

– Точки зміни напрямку варто шукати на всьому діапазоні даних (за замовчуванням – тільки на перших 80% даних): `changepoint_range = 1;`

– Ряд є дуже волатильним (має високу дисперсію), тому, у першому наближенні варто встановлювати велике значення можливого розкиду значень біля точок зміни (зазвичай: від 0,01 до 0,5), тобто – низьку регуляризацію: `changepoint_prior_scale = 0,5` (у разі погіршення прогнозних якостей це число варто зменшувати);

– Аномальні дати можна ідентифікувати окремо і теж задати вручну `holidays = holidays_df`.

Моделі машинного навчання налаштовуються з використанням повного перебору усіх заданих комбінацій параметрів з використанням GridSearchCV.

### Результати дослідження

Програма [2] була застосована для щодобових даних по біткоїну 2021 рік і зазначених вище моделей. Результат наведено на рис. 1.

name_model	type_data	r2_score	rmse	mape
MLP Regressor	valid	-0.198116	1351.685712	2.446971
Random Forest Regressor	valid	-1.346753	1891.733477	3.600522
XGB Regressor	valid	-1.681957	2022.330814	3.910245
ARIMA_manual	valid	-2.134046	2186.144947	4.155419
ARIMA_auto	valid	-2.134046	2186.144947	4.155419
Prophet_4_days_3_order	valid	-3.500235	2619.653422	4.378228
Prophet_7_days_3_order	valid	-3.484681	2615.122624	4.434402
Prophet_5_days_3_order	valid	-3.451903	2605.548101	4.459947
Prophet_14_days_12_order	valid	-3.980607	2755.924964	4.67282
KNeighbors Regressor	valid	-2.878011	2431.816508	4.742777
Prophet_7_days_12_order	valid	-4.348056	2855.776728	4.823952
Prophet_14_days_3_order	valid	-4.210608	2818.840293	4.854828
Prophet_5_days_12_order	valid	-4.572239	2915.017259	4.903258
Prophet_4_days_12_order	valid	-4.853504	2987.680986	4.904383
Linear SVR	valid	-3.263582	2549.843676	4.947899
Support Vector Machines	valid	-3.347282	2574.750635	5.000937
Linear Regression	valid	-7.766195	3656.215868	7.098678
Bagging Regressor	valid	-11.517699	4369.06509	7.56984

Рисунок 1 – Похибки усіх моделей, побудованих для багатьох ознак для валідаційних даних курсу криптовалюти [2]

На рис. 2 наведено приклад моделювання даних за моделлю FB Prophet.

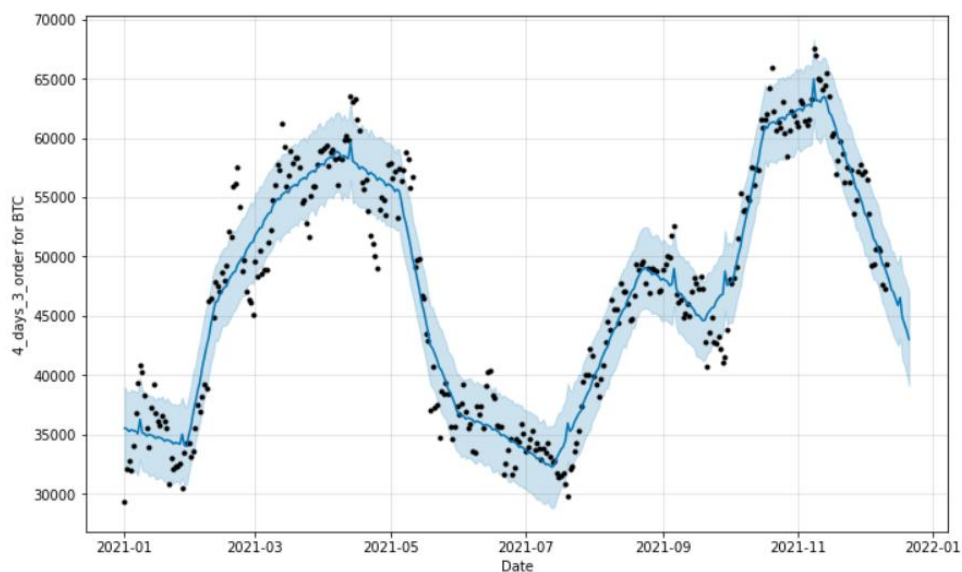


Рисунок 2 – Результат прогнозування курсу біткоїна за найкращою моделлю за валідаційними даними на основі Facebook Prophet [2]

На рис. 1 видно, що оптимальною моделлю є модель MLP Regressor, хоча вона не є достатньо адекватною, оскільки напрямок зміни курсу прогнозує неправильно ( $r_2\_score$  є від'ємним), що означає, що слід продовжувати тюнінг. Але, якщо піти далі за алгоритмом, тоді її варто натренувати на датасеті train+valid і зробити прогнозування тестових даних test (рис. 3).



Рисунок 3 – Реальні тестові дані курсу криптовалюти біткоїн та їх прогноз за моделлю MLP Regressor, оптимальної за усіма трьома метриками [2]

### Висновки

Створено та охарактеризовано Python-засоби автоматизації методів моделювання курсу заданої цифрової криптовалюти та вибору оптимальної моделі серед багатьох видів моделей машинного навчання, у т.ч. нейронних мереж та часових рядів, що дозволяє підвищувати точність прогнозування, особливо у разі застосування ансамблів цих моделей. Ефективність технології продемонстрована на прогнозуванні біткоїну за середньодобовими даними у 2021 році. Технологія є універсальною і легко адаптується до будь-якої криптовалюти.

Були проведені випробування на декількох десятках різних криптовалют, які мають найбільшу капіталізацію. Досягнута точність оптимальної моделі за метрикою MAPE (відносна похибка) склала від 1,97 до 11% (переважно – 3-4%).

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Жуков С. О. Інформаційна технологія прогнозування курсу криптовалют на основі комплексної інженерії ознак // В. Б. Мокін, С. О. Жуков, Л. М. Куперштейн, О. В. Слободянюк // Вісник Вінницького політехнічного інституту, 2022, № 2. Режим доступу: <https://visnyk.vntu.edu.ua/index.php/visnyk/article/view/2757>
2. Mokin V.B. Kaggle Notebook «Crypto - BTC : Advanced Analysis & Forecasting» [Електронний ресурс]. URL: <https://www.kaggle.com/vbmokin/crypto-btc-advanced-analysis-forecasting>
3. Mokin V.B. Kaggle Dataset «Forecasting Top Cryptocurrencies» [Електронний ресурс]. URL: <https://www.kaggle.com/datasets/vbmokin/forecasting-top-cryptocurrencies>

**Мокін Віталій Борисович** – д-р техн. наук, професор, завідувач кафедри системного аналізу та інформаційних технологій, e-mail: [vbmokin@gmail.com](mailto:vbmokin@gmail.com);

**Лосенко Арсен Володимирович** – аспірант кафедри системного аналізу та інформаційних технологій, e-mail: [arsenloosenko@gmail.com](mailto:arsenloosenko@gmail.com).

**Mokin Vitalii B.** – Dr. Sc. (Eng.), Professor, Head of the Chair of Systems Analysis and Information Technology, e-mail: [vbmokin@gmail.com](mailto:vbmokin@gmail.com);

**Losenko Arsen V.** – Post-Graduate Student of the Chair of Systems Analysis and Information Technology, e-mail: [arsenloosenko@gmail.com](mailto:arsenloosenko@gmail.com).