UDK 004.8:336.76

MOLIYAVIY VAQT QATORLARI ASOSIDA FOND BOZORI DINAMIKASINI PROGNOZLASHNING ADAPTIVE CATBOOST MODELI

Erkinov Farhodjon G‘ulomjon o‘g‘li

Sharof Rashidov nomidagi Samarqand davlat universiteti

Annotatsiya. Maqolada moliyaviy vaqt qatorlarining nostatsionarligi, shovqinliligi, volatilitet klasterlanishi va chiziqli bo‘lmagan bog‘lanishlarini hisobga olgan holda fond bozori dinamikasini prognozlash uchun Adaptive CatBoost modeli taklif etiladi. Yondashuv prognozlash masalasini ikki o‘zaro bog‘liq qismga ajratadi: kelajakdagi logarifmik daromad miqdorini regressiya orqali baholash va bozor harakati yo‘nalishini klassifikatsiya qilish. OHLCV ma’lumotlari asosida laglar, rolling-statistikalar, trend, volatilitet, momentum, texnik indikatorlar, rejim va kalendar belgilari shakllantiriladi. Ma’lumotlar 70/15/15 nisbatda xronologik tarzda train, validation va test qismlariga ajratilib, standartlashtirish parametrlari faqat train qismida baholanadi. GOOGL ma’lumotlari bo‘yicha tajribada Adaptive CatBoost modeli MAE=6,2940, RMSE=8,4677, MAPE=3,1877% va R^2=0,9792 natijalarini qayd etdi. Yo‘nalishni aniqlash aniqligi 56,54% ni tashkil etdi. Natijalar feature-rich, leakage-free gradient boosting yondashuvi narx darajasini prognozlashda samarali ekanini ko‘rsatadi.

Kalit so‘zlar: fond bozori, moliyaviy vaqt qatori, sun’iy intellekt, CatBoost, Adaptive CatBoost, feature engineering, logarifmik daromad, data leakage, prognozlash.

Аннотация. В статье предложена модель Adaptive CatBoost для прогнозирования динамики фондового рынка с учетом нестационарности, шума, кластеризации волатильности и нелинейных зависимостей финансовых временных рядов. Задача прогнозирования разделяется на две взаимосвязанные компоненты: регрессионную оценку величины будущей логарифмической доходности и классификацию направления движения рынка. На основе OHLCV-данных формируются лаговые признаки, скользящие статистики, показатели тренда, волатильности, импульса, технические индикаторы, режимные и календарные признаки. Для предотвращения утечки данных используется хронологическое разделение выборки 70/15/15, причем параметры масштабирования определяются только на обучающей части. На данных GOOGL модель достигла MAE=6,2940, RMSE=8,4677, MAPE=3,1877% и R^2=0,9792; точность направления составила 56,54%. Полученные результаты подтверждают эффективность признаково-насыщенного и leakage-free подхода на основе градиентного бустинга для прогнозирования уровня цен.

Ключевые слова: фондовый рынок, финансовые временные ряды, искусственный интеллект, CatBoost, машинное обучение, логарифмическая доходность, прогнозирование.

Abstract. This paper proposes an Adaptive CatBoost model for forecasting stock market dynamics under non-stationarity, noise, volatility clustering, and nonlinear dependencies in financial time series. The forecasting problem is decomposed into two related components: regression-based estimation of future log-return magnitude and classification-based identification of market movement direction. A feature-rich representation is constructed from OHLCV data using lags, rolling statistics, trend, volatility, momentum, technical indicators, regime variables, and calendar features. To prevent data leakage, the dataset is split chronologically into training,

validation, and test subsets using a 70/15/15 ratio, while scaling parameters are estimated only from the training subset. In the GOOGL experiment, the proposed model achieved MAE=6.2940, RMSE=8.4677, MAPE=3.1877%, and R^2=0.9792; Directional Accuracy was 56.54%. The results demonstrate that a feature-rich, leakage-free gradient boosting framework is effective for price-level forecasting and offers a practical alternative to more complex neural architectures.

Keywords: stock market, financial time series, artificial intelligence, CatBoost, Adaptive CatBoost, feature engineering, log-return, data leakage, forecasting.

Fond bozori dinamikasini prognozlash moliyaviy tahlil va sun’iy intellektning eng murakkab yo‘nalishlaridan biridir. Aksiya narxlari korxona moliyaviy ko‘rsatkichlari, makroiqtisodiy sharoit, foiz stavkalari, likvidlik, savdo hajmi, investorlar kutilmalari va axborot oqimlarining birgalikdagi ta’siri ostida shakllanadi. Natijada moliyaviy vaqt qatorlari nostatsionarlik, shovqin, volatilitet klasterlanishi, strukturaviy uzilishlar va chiziqli bo‘lmagan bog‘lanishlarga ega bo‘ladi [1–4].

An’anaviy ARIMA oilasidagi modellar nisbatan barqaror va chiziqli strukturalarni ifodalashda foydali bo‘lsa-da, murakkab bozor rejimlari va nochiziqli munosabatlarni to‘liq qamrab olish imkoniyati cheklangan. Shu sababli LSTM, CNN va ularning gibrid shakllari, shuningdek feature engineering bilan boyitilgan mashinaviy o‘qitish yondashuvlari faol qo‘llanmoqda [5–8]. Adabiyotlar tahlili model murakkabligining o‘zi yuqori prognoz aniqligini kafolatlamasligini; targetni to‘g‘ri qurish, informativ belgilarni shakllantirish va real xronologik validatsiya ham hal qiluvchi ahamiyatga ega ekanini ko‘rsatadi [1,3,9].

Mazkur ishning maqsadi — OHLCV ma’lumotlari asosida shakllantirilgan ko‘p o‘lchovli belgilar fazosida kelajakdagi narx

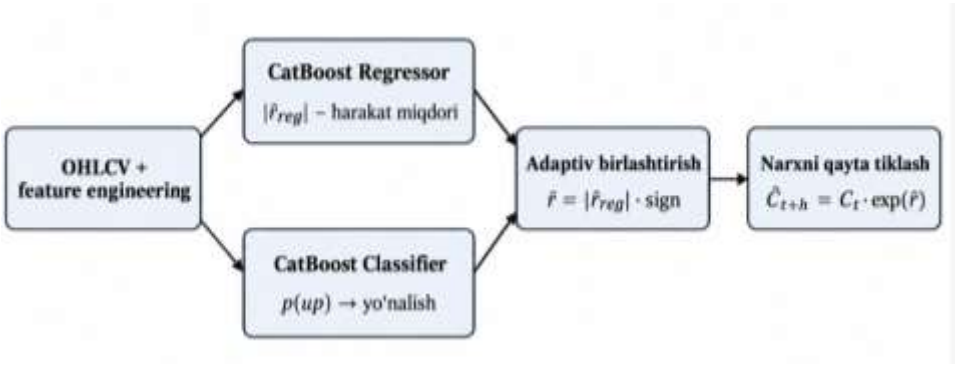
o‘zgarishining miqdori va yo‘nalishini alohida baholab, ularni yagona prognozga birlashtiruvchi Adaptive CatBoost modelini ishlab chiqish va uning samaradorligini ARIMA, LSTM, Multivariate CNN hamda CNN-LSTM Hybrid modellariga nisbatan baholashdir. Taklif etilgan yondashuvda logarifmik daromad maqsad o‘zgaruvchisi sifatida olinadi va barcha preprocessing amallari leakage-free tamoyili bo‘yicha bajariladi.

Tadqiqotda kundalik OHLCV (Open, High, Low, Close, Volume) kuzatuvlari asosiy xom ma’lumot sifatida qabul qilindi. Xom qiymatlar bilan bir qatorda, modelga bozor holatini kengroq ifodalash imkonini beruvchi hosila belgilar — lagged qiymatlar, harakatlanuvchi o‘rtachalar, rolling-statistikalar, volatilitet tavsiflari, momentum ko‘rsatkichlari, RSI, MACD, ATR va Bollinger Bands kabi texnik indikatorlar, rejim belgilari hamda kalendar xususiyatlari kiritildi [9,10]. Shu tariqa t-vaqt momentidagi holat x_t=(x_t1,…,x_tm) belgilar vektori bilan ifodalandi.

Xom narxni bevosita prognozlash o‘rniga, h gorizont uchun kelajakdagi logarifmik daromad asosiy regressiya targeti sifatida tanlandi:

$$r_{t,h} = ln(C_{t+h} / C_t).$$

Bunda C_t joriy yopilish narxi, C_{t+h} esa h savdo kunidan keyingi yopilish narxidir. Yo‘nalish klassifikatsiyasi uchun $d_{t,h}=1$, agar $r_{t,h}>0$ bo‘lsa, aks holda $d_{t,h}=0$ ko‘rinishidagi ikkilamchi target shakllantiriladi. Bu konstruksiya narx masshtabining uzoq muddatli o‘zgarishlariga nisbatan barqarorroq bo‘lib, turli davrlar bo‘yicha statistik solishtirishni soddalashtiradi.



1-rasm. Adaptive CatBoost modelining ikki tarmoqli arxitekturasi

Taklif etilgan model ikki parallel komponentdan tashkil topadi. Regressiya tarmog‘i kelajakdagi logarifmik daromad miqdorini $f_reg(x_t)$ orqali baholaydi, klassifikatsiya tarmog‘i esa yuqoriga harakatlanish ehtimolini $p_t=f_cls(x_t)$ ko‘rinishida beradi. Klassifikator natijasi 0,5 chegaraga nisbatan +1 yoki −1 yo‘nalish belgiga aylantiriladi. Yakuniy return quyidagicha shakllantiriladi:

$$\hat{r}_{t,h} = |\hat{r}_t reg| \cdot \hat{s}_t, \quad bunda \quad \hat{s}_t = +1 \ (p_t \geq 0{,}5), \ -1 \ (p_t < 0{,}5).$$

Shundan so‘ng prognozlangan kelajak narxi log-return fazosidan original narx fazosiga qaytariladi:

$$\hat{C}_{t+h} = C_t \cdot exp(\hat{r}_{t,h}).$$

Mazkur mexanizm bir tomondan harakat kattaligini, ikkinchi tomondan esa uning yo‘nalishini alohida modellashtirish imkonini beradi. Dissertatsiya implementatsiyasida CatBoost Regressor va CatBoost Classifier uchun 700 ta boosting iteratsiyasi, daraxt chuqurligi 6 va learning rate 0,03 parametrlaridan foydalanilgan. Regressor RMSE bo‘yicha, klassifikator esa Logloss funksiyasi bo‘yicha o‘qitilib, klassifikatsiya sifati validatsiya bosqichida AUC orqali nazorat qilinadi.

1-jadval. Eksperimental konfiguratsiyaning asosiy parametrlari

Parametr	Qiymat
Moliyaviy instrument	GOOGL
Ma’lumotlar davri	01.01.2010 – 05.04.2026
Prognoz gorizonti	5 savdo kuni
Sequence length	60
Train / Validation / Test	70% / 15% / 15%
CatBoost iterations	700
Tree depth	6
Learning rate	0,03

Moliyaviy vaqt qatorlarida random train-test split kelajak davrga tegishli axborotning bilvosita training jarayoniga kirib qolishiga olib kelishi mumkin. Shu sababli barcha kuzatuvlar vaqt tartibi bo‘yicha ketma-ket 70% training, 15% validation va 15% test qismlariga ajratildi. StandardScaler parametrlari faqat training to‘plamida aniqlanib, ayni parametrlar validation va test ma’lumotlariga qayta baholanmasdan qo‘llandi. Bunday yondashuv test qismidagi taqsimot haqida oldindan axborot olishning oldini oladi va real deployment holatini yaxshiroq imitatsiya qiladi [11].

2-rasm. Leakage-free xronologik o‘qitish va baholash jarayoni

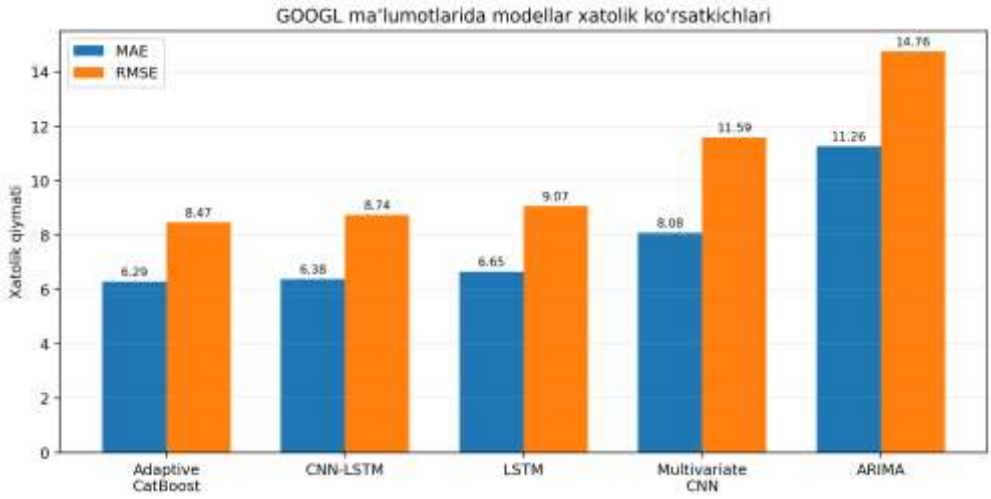
Model sifati MAE, RMSE, MAPE, R^2 va Directional Accuracy (DA) ko‘rsatkichlari orqali baholandi. Birinchi to‘rtta metrika narx darajasini son jihatdan qanchalik aniq tiklanganini, DA esa prognozlangan va haqiqiy return ishoralarining mos kelish ulushini ifodalaydi. Bu ikki turdagi mezonlarni birgalikda qo‘llash zarur, chunki yuqori R^2 yoki past RMSE har doim ham bozor yo‘nalishini to‘g‘ri aniqlashni anglatmaydi.

Asosiy eksperiment GOOGL aksiyasi bo‘yicha 2010–2026-yillar oralig‘idagi ma’lumotlarda amalga oshirildi. Barcha modellar bir xil davr, bir xil 5 kunlik prognoz gorizonti, bir xil target konstruktsiyasi va xronologik bo‘linish sharoitida baholandi. Taqqoslashga an’anaviy ARIMA, LSTM, Multivariate CNN va CNN-LSTM Hybrid modellari kiritildi [5–8].

2-jadval. GOOGL ma’lumotlarida prognozlash modellarining taqqoslama natijalari

Model	MAE	RMSE	MAPE, %	R^2	DA, %	Vaqt, s
Adaptive CatBoost	**6,2940**	**8,4677**	**3,1877**	**0,9792**	**56,54**	**26,18**

Model	MAE	RMSE	MAPE, %	R^2	DA, %	Vaqt, s
CNN-LSTM Hybrid	6,3831	8,7386	3,2493	0,9778	58,05	39,90
LSTM	6,6520	9,0734	3,3961	0,9761	59,73	20,71
Multivariate CNN	8,0840	11,5942	4,0310	0,9610	50,67	26,35
ARIMA	11,2634	14,7638	5,5447	0,9367	46,64	7,26



3-rasm. GOOGL ma'lumotlarida MAE va RMSE bo'yicha modellar taqqoslanishi

Natijalarga ko'ra, Adaptive CatBoost modeli MAE=6,2940, RMSE=8,4677, MAPE=3,1877% va R^2=0,9792 qiymatlari bilan narx darajasini prognozlashda eng past xatolikni qayd etdi. CNN-LSTM Hybrid eng yaqin raqib bo'lib, MAE=6,3831 va RMSE=8,7386 natijalarini ko'rsatdi. LSTM modeli regressiya metrikalarida Adaptive CatBoostdan ortda qolgan bo'lsa-da, Directional Accuracy bo'yicha 59,73% bilan yuqori natijaga erishdi; Adaptive CatBoostda bu ko'rsatkich 56,54% ni tashkil etdi.

Adaptive CatBoostning ARIMAga nisbatan MAE kamayishi taxminan 44,1%, RMSE kamayishi 42,6%, MAPE kamayishi esa 42,5% ni tashkil etdi. Multivariate CNN bilan solishtirilganda MAE taxminan 22,1% va RMSE 27,0% ga pasaydi. Bu natijalar structured feature space va gradient boosting kombinatsiyasi ayrim moliyaviy prognozlash vazifalarida chuqur neyron arxitekturalarga raqobatbardosh, hatto narx darajasi bo‘yicha ustun natija berishi mumkinligini ko‘rsatadi.

Shu bilan birga, Directional Accuracy natijalari muhim metodologik xulosani beradi: narxning aniq qiymatini tiklash va harakat yo‘nalishini to‘g‘ri topish bir xil vazifa emas. Yuqori R^2 va past RMSE modelning haqiqiy narx trayektoriyasiga yaqinligini ko‘rsatadi, biroq kichik returnlar atrofida ishora xatosi Directional Accuracy ko‘rsatkichini pasaytirishi mumkin. Shu sababli kelgusida binary direction classification o‘rniga “up–neutral–down” uch sinfli target, threshold-based labeling, shuningdek sentiment va makroiqtisodiy belgilarni qo‘shish maqsadga muvofiq.

Taklif etilgan Adaptive CatBoost yondashuvining ilmiy yangiligi prognozlash masalasini faqat bitta regression target bilan cheklamasdan, kelajakdagi harakat miqdori va yo‘nalishini ikki alohida, ammo o‘zaro bog‘langan CatBoost komponentlarida baholashdan iborat. Regressor qiymatning absolyut miqdorini, klassifikator esa ishorasini belgilaydi; shu sababli yakuniy forecast miqdor va yo‘nalish axborotining adaptiv kombinatsiyasi sifatida quriladi. Ushbu arxitektura narx darajasini bevosita uzoq gorizontga ekstrapolyatsiya qilish o‘rniga log-return fazosida ishlashi bilan ham farqlanadi.

Metodologik jihatdan yana bir muhim natija — feature engineering, target construction va validatsiya strategiyasining bitta yaxlit forecasting pipeline ichida muvofiqlashtirilganidir. Ayniqsa, scaler parametrlarini faqat training ma’lumotlarida baholash va test segmentini yakuniy bosqichgacha

yopiq saqlash olingan metrikalarning amaliy sharoitga yaqinligini oshiradi. Moliyaviy vaqt qatorlarida bu jihat algoritm turidan kam bo‘lmagan ahamiyatga ega, chunki noto‘g‘ri preprocessing yuqori ko‘ringan, biroq real prognozlashda takrorlanmaydigan natijalarni yuzaga keltirishi mumkin.

Amaliy nuqtai nazardan model structured tabular feature space bilan ishlagani sababli katta chuqur neyron tarmoqlarga nisbatan soddaroq deployment va interpretatsiya imkonini beradi. U real fond bozori ma’lumotlarida takroriy eksperimentlar, turli aksiyalar bo‘yicha model taqqoslashlari va qarorlarni analitik qo‘llab-quvvatlash tizimlari uchun asos bo‘lishi mumkin. Biroq mazkur model savdo signali yoki investitsiya tavsiyasi sifatida emas, prognozlash va ilmiy tahlil vositasi sifatida talqin qilinishi lozim.

Maqolada moliyaviy vaqt qatorlari uchun Adaptive CatBoost prognozlash modeli taklif etildi. Yondashuvning asosiy jihatlari: OHLCV ma’lumotlarini kengaytirilgan feature engineering bilan ifodalash, xom narx o‘rniga future log-returnni regressiya targeti sifatida tanlash, return miqdori va yo‘nalishini alohida CatBoost komponentlarida baholash hamda leakage-free xronologik o‘qitish strategiyasidan foydalanishdir.

GOOGL ma’lumotlarida model MAE=6,2940, RMSE=8,4677, MAPE=3,1877% va R^2=0,9792 natijalarini ko‘rsatib, narx darajasini prognozlash bo‘yicha taqqoslangan modellar ichida eng past xatolikka erishdi. Biroq Directional Accuracy bo‘yicha LSTMning yuqoriroq natijasi bozor yo‘nalishini aniqlash alohida takomillashtirishni talab qilishini ko‘rsatdi. Umuman, natijalar moliyaviy prognozlashda algoritm murakkabligi bilan bir qatorda target dizayni, feature engineering, xronologik validatsiya va data leakage’ning oldini olish hal qiluvchi ahamiyatga ega ekanini tasdiqlaydi.

Foydalanilgan adabiyotlar

1. Kumbure M.M., Lohrmann C., Luukka P., Porras J. Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications. 2022;197:116659. doi:10.1016/j.eswa.2022.116659.

2. Ajiga D.I. et al. Machine learning for stock market forecasting: A review of models and accuracy. Finance & Accounting Research Journal. 2024;6(2):112–124. doi:10.51594/farj.v6i2.783.

3. Sezer O.B., Gudelek M.U., Ozbayoglu A.M. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing. 2020;90:106181. doi:10.1016/j.asoc.2020.106181.

4. Saberironaghi M., Ren J., Saberironaghi A. Stock market prediction using machine learning and deep learning techniques: A review. AppliedMath. 2025;5(3):76. doi:10.3390/appliedmath5030076.

5. Hiransha M., Gopalakrishnan E.A., Menon V.K., Soman K.P. NSE stock market prediction using deep-learning models. Procedia Computer Science. 2018;132:1351–1362. doi:10.1016/j.procs.2018.05.050.

6. Nabipour M., Nayyeri P., Jabani H., Mosavi A. Deep learning for stock market prediction. Entropy. 2020;22(8):840. doi:10.3390/e22080840.

7. Mehtab S., Sen J. Stock price prediction using convolutional neural network on a multivariate time series. Proc. 3rd National Conference on Machine Learning and Artificial Intelligence (NCMLAI). New Delhi; 2020.

8. Sahib M.R., Elkina H., Zaki T. From technical indicators to trading decisions: A deep learning model combining CNN and LSTM. International Journal of Advanced Computer Science and Applications. 2024;15(6).

9. Htun H.H., Biehl M., Petkov N. Survey of feature selection and extraction techniques for stock market prediction. Financial Innovation. 2023;9:26. doi:10.1186/s40854-022-00441-7.

10. Patsiarikas M., Papageorgiou G., Tjortjis C. Using machine learning on macroeconomic, technical, and sentiment indicators for stock market forecasting. Information. 2025;16(7):584. doi:10.3390/info16070584.

11. Shi L. Machine learning for stock prediction by different models. 2021 IEEE Symposium Series on Computational Intelligence (SSCI). 2021:1–10. doi:10.1109/SSCI50451.2021.9660134.

12. Sonkavde G., Dharrao D.S., Bongale A.M., et al. Forecasting stock market prices using machine learning and deep learning models: A systematic review, performance analysis and discussion of implications. International Journal of Financial Studies. 2023;11(3):94. doi:10.3390/ijfs11030094.