

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.9

DOI <https://doi.org/10.35546/kntu2078-4481.2025.3.2.1>

В. В. АРГУНОВ

асистент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-2505-1969

Д. Б. ЗАДНІПРОВСЬКИЙ

аспірант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0004-6637-6315

ФОРМУВАННЯ ІНВЕСТИЦІЙНОГО ПОРТФЕЛЮ НА ОСНОВІ ПРОГНОЗУВАННЯ ФОНДОВОГО РИНКУ З ВИКОРИСТАННЯМ БАГАТОДЖЕРЕЛЬНИХ ДАНИХ

У дослідженні використано дані з різних джерел: історичні ціни акцій, технічні індикатори, новинні повідомлення та публікації з соціальних мереж. Такий підхід дозволяє поєднувати як кількісні, так і якісні фактори, що безпосередньо впливають на динаміку фондового ринку, та формувати більш повну й комплексну інформаційну базу для прогнозування. На відміну від традиційних моделей, які переважно враховують лише статистичні характеристики часових рядів, застосування багатоджерельних даних дає змогу краще відобразити складні взаємозв'язки між економічними, інформаційними та поведінковими чинниками, що визначають коливання вартості цінних паперів. Експериментальні результати підтвердили, що збільшення кількості шарів і нейронів у нейронних мережах, а також застосування регуляризації у вигляді dropout, значно підвищують точність прогнозів і дозволяють уникати перенавчання. Найкраща модель – чотиришарова BiLSTM із 150 нейронами та параметром dropout 0.2 – досягла показника MAPE 4.36, перевершивши аналогічну модель LSTM. Це свідчить про здатність BiLSTM краще враховувати довгострокові залежності та приховані патерни у часових рядах. Водночас важливим обмеженням виявився суттєво більший час навчання BiLSTM, що може ускладнювати або навіть унеможливлювати її застосування в системах реального часу, де швидкість прийняття рішень є критичною. Метою дослідження є розробка ефективної моделі прогнозування динаміки фондового ринку на основі багатоджерельних даних із застосуванням сучасних методів глибинного навчання, здатної забезпечувати високу точність і надійність результатів навіть в умовах нестабільності. Об'єктом дослідження виступає процес прогнозування цін на акції в умовах високої волатильності та непередбачуваності фінансових ринків, що зумовлено як внутрішніми економічними факторами, так і зовнішніми інформаційними впливами. Предметом дослідження є методи та моделі глибинного навчання (зокрема LSTM та BiLSTM), орієнтовані на аналіз часових рядів і текстових даних із різних інформаційних джерел, інтеграція яких дозволяє підвищити якість прогнозування. Отримані результати свідчать про практичну доцільність використання глибинного навчання в управлінні інвестиціями, оскільки такі методи забезпечують більш точні й надійні прогнози в умовах значної невизначеності та нестабільності. Вони можуть стати ефективним інструментом для оптимізації формування інвестиційних портфелів, зниження ризиків і підвищення прибутковості рішень у сфері фінансів. Подальші дослідження можуть бути спрямовані на вдосконалення архітектур моделей для прискорення процесу навчання та розробку гібридних систем, що поєднують глибинне навчання з методами обробки природної мови для ще глибшого аналізу впливу інформаційних факторів на фондовий ринок.

Ключові слова: машинне навчання, прогнозування, інвестиційний портфель, модель, штучний інтелект, аналіз, обробка тексту.

V. V. ARGUNOV

Assistant at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-2505-1969

D. B. ZADNIPROVSKIY

Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0004-6637-6315

FORMATION OF AN INVESTMENT PORTFOLIO BASED ON STOCK MARKET FORECASTING USING MULTI-SOURCE DATA

The study uses data from various sources: historical stock prices, technical indicators, news reports, and social media publications. This approach makes it possible to combine both quantitative and qualitative factors that directly influence stock market dynamics and to build a more complete and comprehensive information base for forecasting. Unlike traditional models, which mainly take into account only the statistical characteristics of time series, the use of multi-source data allows for a better representation of the complex interrelationships between economic, informational, and behavioral factors that determine fluctuations in the value of securities. The experimental results confirmed that increasing the number of layers and neurons in neural networks, as well as applying regularization in the form of dropout, significantly improves forecast accuracy and helps to avoid overfitting. The best-performing model – a four-layer BiLSTM with 150 neurons and a dropout parameter of 0.2 – achieved a MAPE of 4.36, outperforming the comparable LSTM model. This demonstrates BiLSTM's ability to better capture long-term dependencies and hidden patterns in time series. At the same time, a significant limitation was the considerably longer training time required by BiLSTM, which may complicate or even preclude its use in real-time systems where decision-making speed is critical. The aim of the study is to develop an effective model for forecasting stock market dynamics based on multi-source data using modern deep learning methods, capable of ensuring high accuracy and reliability of results even under unstable conditions. The object of the study is the process of stock price forecasting under conditions of high volatility and unpredictability of financial markets, driven by both internal economic factors and external informational influences. The subject of the study is deep learning methods and models (in particular LSTM and BiLSTM) focused on the analysis of time series and textual data from various information sources, the integration of which makes it possible to improve forecasting quality. The obtained results indicate the practical feasibility of applying deep learning in investment management, since such methods provide more accurate and reliable forecasts under conditions of significant uncertainty and instability. They can become an effective tool for optimizing investment portfolio formation, reducing risks, and increasing the profitability of financial decision-making. Further research may be directed toward improving model architectures to accelerate the training process and developing hybrid systems that combine deep learning with natural language processing methods for a deeper analysis of the influence of informational factors on the stock market.

Key words: machine learning, forecasting, investment portfolio, model, artificial intelligence, analysis, text processing.

Постановка проблеми

Прогнозування цін на акції може дати можливості для збільшення прибутку як для продавця, так і для брокера. В минулому робити точні прогнози на фондовому ринку було складно через брак технологій і обчислювальних можливостей. Проте зі стрімким розвитком технологій точність прогнозів значно підвищилася, і сьогодні завдання прогнозування стає значно простішим порівняно з минулими роками. Прогнози на фондовому ринку часто мають хаотичний характер, а не випадковий, що означає можливість їх передбачення шляхом аналізу історії динаміки цін.

Машинне навчання є дієвим методом для моделювання складних процесів, зокрема для прогнозування ринкових значень, близьких до фактичних. Завдяки зростанню точності машинне навчання стає ефективним інструментом для практичного застосування. Багато дослідників дійшли висновку, що машинне навчання є результативним і точним способом оцінювання фондових цін, завдяки чому ця технологія набуває популярності у фінансовій сфері [1–4].

Прогнозування фондового ринку останнім часом привертає значну увагу як дослідників у фінансовій сфері, так і інвесторів. Оскільки ринок характеризується високою нелінійністю та коливаннями часових рядів, для його моделювання потрібні ефективні підходи, здатні працювати з такими різнорідними даними. Традиційні лінійні моделі, зокрема ARMA та ARIMA, раніше застосовувалися для прогнозування, проте вони придатні лише для окремих акцій і не можуть бути поширені на інші. Через стохастичний і неоднозначний характер фондового ринку здійснення прогнозів із використанням класичних методів є надзвичайно складним та ризикованим. У цьому контексті дедалі важливішу роль починають відігравати методи глибинного навчання, які демонструють високу результативність у виявленні закономірностей і трендів фондового ринку.

Технології глибинного навчання мають здатність працювати з високонелінійними даними, що робить їх перспективними для прогнозування руху цін на акції. Доведено, що моделі на основі глибинного навчання перевершують ARIMA у прогнозуванні часових рядів [5], особливо для завдань довгострокового передбачення. Зокрема, LSTM-моделі продемонстрували вищу точність порівняно з ARIMA, однак актуальним завданням залишається перевірка того, чи можна ще більше покращити їхні результати за рахунок використання багатосарових архітектур. Водночас ринково значуща інформація надходить із численних джерел – соціальних мереж, новинних ресурсів, корпоративних оглядів – і саме на ній інвестори будують прогнози. Обробка таких великих і різномірних масивів даних є складним викликом навіть для DL-моделей. Крім того, багатомасштабність часових рядів у фінансовій сфері ускладнює роботу традиційних методів, роблячи їх малоефективними для точних прогнозів. Важливо враховувати, що будь-яка похибка у прогнозі може призвести до значних фінансових втрат, тому інвестори потребують максимально точних оцінок ринкової вартості, що робить завдання прогнозування ще більш складним.

Таким чином, актуальним постає дослідження та розробка сучасних підходів до прогнозування фондового ринку, здатних враховувати складну нелінійну динаміку часових рядів і різномірність джерел інформації. Використання методів глибинного навчання, зокрема моделей LSTM та BiLSTM, відкриває можливості для підвищення точності передбачення та формування більш надійних інструментів підтримки інвестиційних рішень, що має вагоме практичне значення в умовах високої волатильності фінансових ринків.

Аналіз останніх досліджень і публікацій

У сучасних наукових дослідженнях спостерігається зростаючий інтерес до проблеми прогнозування фондових ринків. Значна кількість робіт присвячена використанню статистичних методів, таких як ARMA та ARIMA, які тривалий час вважалися основними інструментами для моделювання часових рядів. Водночас дослідники відзначають, що такі моделі обмежені у застосуванні, оскільки вони враховують лише окремі аспекти динаміки ринку і не здатні адекватно відображати складні нелінійні залежності, характерні для фінансових даних. Це створило передумови для пошуку нових методологічних підходів, здатних забезпечити більш високу точність і стійкість прогнозів.

Останніми роками значна увага приділяється застосуванню методів машинного та глибинного навчання для прогнозування біржових показників. Зокрема, у наукових публікаціях активно досліджуються можливості моделей на основі нейронних мереж, серед яких важливе місце займають LSTM та BiLSTM, що демонструють здатність ефективно працювати з нелінійними часовими рядами. Крім того, у сучасних публікаціях акцент робиться на інтеграції ринкових і текстових даних (новин, соціальних мереж, аналітичних звітів), що дозволяє підвищити точність прогнозів. Такий підхід поступово витісняє традиційні методи, утверджуючи глибинне навчання як провідний інструмент досліджень у сфері фінансового прогнозування.

У статті [6] автори демонструють новий оптимізаційний підхід до прогнозування фондових цін, заснований на моделі Multi-Layer Sequential LSTM (MLS LSTM), яка використовує оптимізатор Adam. Алгоритм MLS LSTM застосовує нормалізовані часові ряди, поділені на часові кроки, щоб визначати взаємозв'язок між минулими та майбутніми значеннями з метою підвищення точності прогнозів. Крім того, він усуває проблему зникання градієнта, притаманну простим рекурентним нейронним мережам. Модель MLS LSTM демонструє надзвичайно високі показники точності, однак її недоліками є можливе перенавчання, використання обмеженого набору лише цінкових даних без урахування зовнішніх факторів та відсутність перевірки стабільності прогнозів у різних ринкових умовах.

У роботі [7] пропонується проста архітектура на основі багатосарових перцептронів (MLP) під назвою StockMixer, яка легко оптимізується та демонструє високу прогнозну ефективність. StockMixer виконує послідовно: змішування індикаторів, потім змішування часових характеристик, а далі змішування акцій. На відміну від стандартного MLP-змішування, у нашому підході часовий блок передбачає обмін багатомасштабною часовою інформацією, тоді як змішування акцій реалізується через явне врахування впливів «акція-ринок» та «ринок-акція». Результати експериментів на реальних фондових даних показали, що StockMixer перевершує сучасні методи прогнозування з помітним відривом, одночасно знижуючи використання пам'яті та витрати часу на обчислення.

У дослідженні представлено у статті [8] запропоновано метод, спрямований на підвищення ефективності системи SPF на основі сучасних алгоритмів машинного й глибинного навчання. Спершу застосовано екстремальний градієнтний бустинг (XGBoost) як техніку відбору ознак для виокремлення найважливіших характеристик із багатовимірних часових рядів та усунення надлишкових параметрів. Далі відібрані ознаки були подані на вхід глибинної мережі довготривалої короткочасної пам'яті (LSTM) для прогнозування цін акцій. Така архітектура дозволила відобразити часову природу вхідних даних та повністю використати контекстну інформацію для майбутніх прогнозів. Складна структура моделі дала змогу захопити більше стохастичних залежностей у динаміці цін. Запропонований метод є універсальним і може застосовуватися як до даних про акції, так і до валютних ринків (Forex). Недоліком цього підходу є його висока обчислювальна складність, що зумовлена використанням XGBoost для відбору ознак та глибинної LSTM-мережі для прогнозування. Крім того, модель значною мірою залежить від якості та обсягу вхідних даних, що може обмежувати її узагальнюваність на інших ринках або в умовах різкої зміни економічних факторів.

Отже, узагальнюючи результати останніх досліджень, можна стверджувати, що формування інвестиційного портфеля на основі прогнозування фондового ринку в реальному часі з використанням багатоджерельних даних є надзвичайно актуальним напрямом, оскільки поєднання глибинного навчання та різномірної інформації дає змогу підвищити точність прогнозів і забезпечити ефективність прийняття фінансових рішень.

Формулювання мети дослідження

Метою роботи є розробка та оцінка підходу до формування інвестиційного портфеля на основі прогнозування фондового ринку в реальному часі з використанням багатоджерельних даних.

Викладення основного матеріалу дослідження

Для ефективного управління інвестиціями необхідно мати надійні інструменти прогнозування, які враховують складність та динамічність фондового ринку. В умовах швидкої зміни ринкової інформації традиційні підходи до формування інвестиційного портфелю втрачають свою ефективність. Саме тому в цьому дослідженні пропонується використання багатоджерельних даних – зокрема новинних повідомлень, публікацій у соціальних мережах, технічних індикаторів та історичних цін акцій – для побудови моделі, здатної прогнозувати рух ринку в реальному часі.

Рисунок 1 зображує структурну модель формування інвестиційного портфелю, яка базується на прогнозах фондового ринку, отриманих за допомогою сучасних методів аналізу та обробки даних. Ця модель поєднує етапи збору, підготовки, аналізу інформації та прийняття рішень, що дозволяє формувати збалансований портфель з урахуванням ризиків і потенційного прибутку в умовах високої ринкової волатильності.

Дана діаграма ілюструє структурну модель формування інвестиційного портфелю на основі прогнозування фондового ринку в реальному часі з використанням багатоджерельних даних. У центрі моделі – поетапна обробка різних типів інформації: твіти з Twitter, новинні статті з вебсайтів і фондові дані з біржі NASDAQ. Після збору ці дані проходять очищення від помилок, дублювань та пропущених значень. Текстові дані додатково обробляються шляхом видалення хештегів, емодзі, посилань, а також застосовується лематизація.

Кількісні характеристики нормалізуються, а ключові ознаки виділяються за допомогою методу головних компонент (PCA). Паралельно застосовується аналіз настроїв за допомогою попередньо навченої трансформерної моделі BERT, що дозволяє класифікувати новини та твіти за емоційним забарвленням і додати відповідні мітки до текстів.

Після підготовки всі ознаки передаються до моделі Bi-LSTM, яка адаптована за допомогою алгоритму оптимізації Adam. Результати прогнозу оцінюються за допомогою метрик точності. Такий підхід дає змогу побудувати ефективну систему прогнозування змін на фондовому ринку, що є основою для формування обґрунтованого інвестиційного портфелю.

У цій статті як метрику оцінки було обрано MAPE. Підсумкові значення MAPE для моделей LSTM отримувалися як середнє всіх значень після крос-валідації.

Одношарова модель LSTM з 32 нейронами та часовим кроком у 1 день (look-back) досягла MAPE на рівні 12.55 для прогнозування скоригованої ціни закриття Google на наступний день. Для порівняння, модель BiLSTM з такою ж конфігурацією показала дещо кращий результат – 12.49, але при цьому час навчання зріс: LSTM тренувався 10.19 секунд, тоді як BiLSTM – 13.50 секунд.

Додавання ще одного шару та збільшення кількості нейронів до 64 покращило точність: MAPE склав 9.28 для LSTM і 9.07 для BiLSTM. Водночас час тренування скоротився до 8.08 секунд для LSTM та 12.95 секунд для BiLSTM. Хоча BiLSTM демонструє кращу точність, його навчання займає майже удвічі більше часу.

Подальші експерименти включали використання невеликого dropout у 0.2 при двошаровій структурі з 64 нейронами. Це дало суттєве покращення: MAPE знизився до 5.24 у LSTM і 5.14 у BiLSTM. Загалом, чим більше шарів та нейронів використовувалося, тим точнішою ставала модель, хоча час тренування збільшувався. Результати дослідження представлені в таблиці 1

Найкращою архітектурою виявилась 4-шарова модель з 150 нейронами та dropout 0.2, яка при часовому кроці 1 день показала результати: 4.85 для LSTM і 4.36 для BiLSTM. Таким чином, BiLSTM стабільно перевершував LSTM за точністю, проте потребував значно більше часу на навчання.

Висновки

Прогнозування цін на акції є надзвичайно актуальним завданням сучасної фінансової сфери, адже від точності передбачень залежить ефективність інвестиційних рішень. Традиційні статистичні моделі, такі як ARMA та ARIMA, виявилися обмеженими у застосуванні через неспроможність адекватно відображати складну нелінійну динаміку фондового ринку. Це зумовило перехід до використання методів глибинного навчання, серед яких LSTM та BiLSTM довели свою високу ефективність у роботі з нелінійними часовими рядами та багатоджерельними даними.

Експериментальні результати показали, що збільшення кількості шарів і нейронів, а також застосування dropout, суттєво знижує похибку прогнозування. Найкраща архітектура – чотиришарова модель із 150 нейронами та dropout 0.2 – забезпечила MAPE 4.85 для LSTM і 4.36 для BiLSTM. Порівняння моделей підтвердило, що

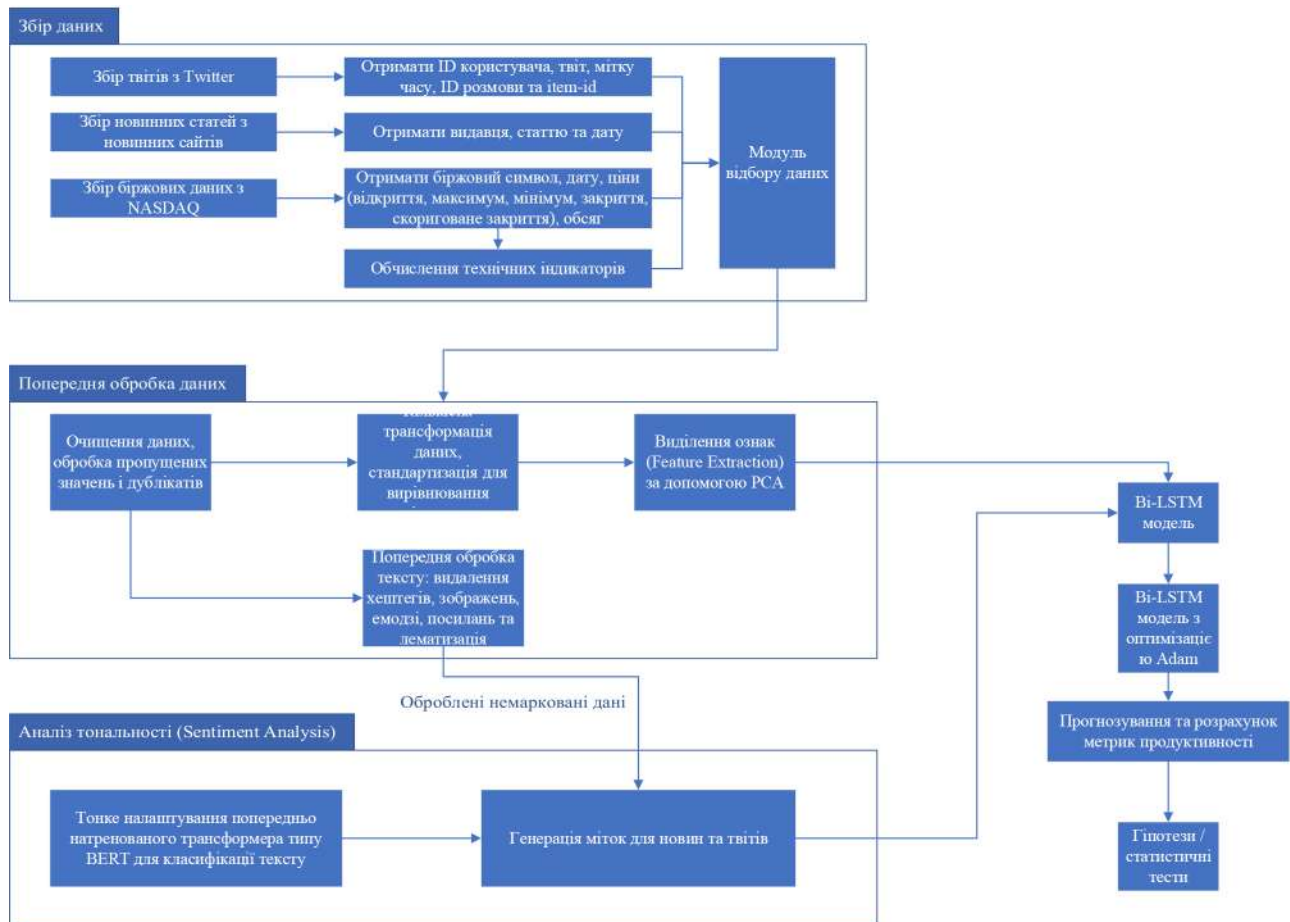


Рис. 1. Структурна модель формування інвестиційного портфелю на основі прогнозування фондового ринку в реальному часі з використанням багатоджерельних даних

Таблиця 1

Порівняння результатів моделей LSTM та BiLSTM

Параметри	Базова	Розширена	З dropout	Оптимальна
Шари	1	2	2	4
Нейрони	32	64	64	150
Dropout	–	–	0.2	0.2
Часовий крок	1 день	1 день	1 день	1 день
MAPE (LSTM)	12.55	9.28	5.24	4.85
Час навчання (LSTM)	10.19 с	8.08 с	–	–
MAPE (BiLSTM)	12.49	9.07	5.14	4.36
Час навчання (BiLSTM)	13.50 с	12.95 с	–	–

BiLSTM стабільно демонструє кращу точність, проте потребує значно більше часу на навчання, що може бути критичним для завдань у реальному часі.

Практичне значення дослідження полягає у використанні багатоджерельних даних, таких як історичні ціни, технічні індикатори, новини та твіти, що в комплексі з методами глибинного навчання дозволяє підвищити точність прогнозів і створити надійні інструменти для управління інвестиційними портфелями в умовах високої волатильності фінансових ринків.

Список використаної літератури

1. Nti, I.K., Adekoya, A.F. & Weyori, B.A. A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. J Big Data 8, 17 (2021), doi: <https://doi.org/10.1186/s40537-020-00400-y>
2. X. Zhang, S. Qu, J. Huang, B. Fang and P. Yu, "Stock Market Prediction via Multi-Source Multiple Instance Learning", in IEEE Access, vol. 6, pp. 50720-50728, 2018, doi: 10.1109/ACCESS.2018.2869735

3. Wu, Shengting, et al. "S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis". *Connection Science* 34.1 (2022): 44–62.
4. Zhang, Qiuyue, et al. "Graph-based stock prediction with multisource information and relational data fusion". *Information Sciences* 690 (2025): 121561.
5. Mashadihasanli, Tamerlan. "Stock market price forecasting using the ARIMA model: an application to Istanbul, Turkiye". *İktisat Politikası Araştırmaları Dergisi* 9.2 (2022): 439–454.
6. Md, Abdul Quadir, et al. "Novel optimization approach for stock price forecasting using multi-layered sequential LSTM". *Applied Soft Computing* 134 (2023): 109830, doi: <https://doi.org/10.1016/j.asoc.2022.109830>
7. Fan, J., & Shen, Y. (2024). StockMixer: A Simple Yet Strong MLP-Based Architecture for Stock Price Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8), 8389-8397, doi: <https://doi.org/10.1609/aaai.v38i8.28681>

Дата першого надходження рукопису до видання: 21.09.2025

Дата прийнятого до друку рукопису після рецензування: 16.10.2025

Дата публікації: 28.11.2025