4. Ціни на товарну землю зросли майже до 47 тис. грн/га. AgroPortal.ua. URL: https://agroportal.ua/news/ukraina/cini-na-tovarnu-zemlyu-pidnyalisya-mayzhe-do-47-tis-grn-ga (дата звернення: 15.10.2026).

5. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne : OTexts, 2021. URL: https://otexts.com/fpp3/ (дата звернення: 15.10.2026).

6. Hyndman R. J., Koehler A. B. Another look at measures of forecast accuracy. International Journal of Forecasting. 2006. Vol. 22, No. 4. P. 679-688. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001.

УДК 004.032.26

ПОРІВНЯННЯ ТОЧНОСТІ ТА СТАБІЛЬНОСТІ СТАТИСТИЧНИХ І НЕЙРОМЕРЕЖЕВИХ МЕТОДІВ ПРОГНОЗУВАННЯ МІСЯЧНОГО СЕЗОННОГО РЯДУ

Віолетта Кім

Миколаївський національний аграрний університет, м. Миколаїв, Україна

Науковий керівник: О. Є. Богатєнкова,
асистент Миколаївський національний аграрний університет

Прогнозування часових рядів лежить в основі планування в економіці, енергетиці та агровиробництві, а нейромережеві моделі нині часто протиставляють класичним статистичним. Проте порівняння на 1045 місячних рядах змагання M3 показало, що розглянуті методи машинного навчання поступалися статистичним за точністю на всіх горизонтах і потребували більше обчислювальних ресурсів [1]; водночас гібридні схеми намагаються поєднати переваги обох підходів [2]. Для користувача важлива не лише середня точність, а й відтворюваність результату: нейромережа, навчена двічі з різними початковими вагами, може давати різні прогнози. Ще одна методична проблема – оцінювання на єдиному розбитті «навчання / тест», яке випадково може сприяти одному з методів; тому рекомендують повторювати оцінювання на кількох початках прогнозу [5].

Мета роботи – порівняти точність і стабільність статистичних методів (Holt-Winters, SARIMA) та багатошарового перцептрона (MLP) під час прогнозування сезонного ряду як на одному розбитті з горизонтом 24 місяці, так і за ковзним початком прогнозу з горизонтом 12 місяців.

Використано відкритий ряд міжнародних авіаперевезень пасажирів (тис. осіб) за 144 місяці 1949-1960 рр. [3], отриманий з набору AirPassengers (R datasets) засобами бібліотеки statsmodels. Основне розбиття: навчальна вибірка – 1949-1958 рр. (n = 120), тестова – 1959-1960 рр. (h = 24). Порівнюються: сезонний наївний прогноз (базовий рівень); модель Holt–Winters з адитивним трендом і мультиплікативною сезонністю [4, 5]; SARIMA(0,1,1)(0,1,1)12 для логарифмованого ряду, тобто «модель авіалінії» [3]:

$$(1-B)(1-B^{12})\,ln\,y_t \;=\; (1+\theta B)(1+\Theta B^{12})\,\varepsilon_t \qquad (1)$$

де B – оператор зсуву назад, θ і Θ – параметри ковзного середнього, ε_t – білий шум. Порядок моделі задано апріорі й не підбирався за тестовими даними; на основній навчальній вибірці отримано $\theta = -0{,}342$ і $\Theta = -0{,}541$. Параметри згладжування Holt-Winters оцінено автоматично ($\alpha = 0{,}364$; $\beta \approx 0$; $\gamma = 0{,}636$), тобто нахил тренду фактично не змінюється з часом.

MLP має один прихований шар (tanh), вхід із 12 останніх стандартизованих логарифмічних приростів і один вихід – наступний приріст; навчання методом L-BFGS із L2-регуляризацією 0,01. Кількість нейронів обрано з {4, 8, 16} за помилкою на останніх 12 місяцях навчальної вибірки (обрано 16); прогноз на кілька кроків вперед – рекурсивний, результат перетворюється назад до рівнів ряду. Для оцінки стабільності MLP навчали 30 разів із різними початковими вагами, а «ансамблем» називаємо середнє прогнозів цих запусків. Точність оцінювали за показниками:

$$\text{MAPE} \;=\; \frac{100}{h}\sum_{t=1}^{h}\frac{|y_t-\hat{y}_t|}{y_t} \qquad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{h}\sum_{t=1}^{h}(y_t - \hat{y}_t)^2} \quad (3)$$

$$\text{MASE} = \frac{\frac{1}{h}\sum_{t=1}^{h}|y_t - \hat{y}_t|}{\frac{1}{n-m}\sum_{t=m+1}^{n}|y_t - y_{t-m}|} \quad (4)$$

де y_t і $\hat{y}_t$ – фактичне та прогнозне значення, h – горизонт, m = 12 – сезонний лаг, n = 120 – довжина навчальної вибірки. MASE [6] нормує похибку на середню абсолютну похибку сезонного наївного прогнозу у навчальній вибірці: значення менше за 1 означає кращу точність, ніж у цього наївного прогнозу.

Результати основного розбиття наведено в табл. 1 і на рис. 1. Найкращі показники за всіма метриками має Holt-Winters (MAPE 6,39 %, MASE 1,01); SARIMA дає 8,52 %, а ансамбль MLP – 7,80 %. Усі методи значно точніші за сезонний наївний прогноз (15,52 %). Різниця між Holt-Winters і ансамблем MLP невелика, а на другому році тестового періоду ансамбль навіть дещо точніший (7,66 % проти 7,99 %), проте на одному ряді таку різницю не можна вважати достовірною.

Таблиця 1. Показники точності 24-місячного прогнозу (1959-1960 рр.)

Метод	MAPE, %	RMSE, тис. пас.	MASE	MAPE, 1-й рік, %	MAPE, 2-й рік, %
Сезонний наївний	15,52	76,99	2,49	11,06	19,99
Holt–Winters	6,39	32,49	1,01	4,79	7,99
SARIMA(0,1,1)(0,1,1)12	8,52	43,19	1,38	6,76	10,28
MLP, ансамбль із 30 запусків	7,80	39,29	1,25	7,95	7,66
MLP, окремий запуск (середнє ± σ)	11,86 ± 6,91	62,45 ± 35,30	1,88 ± 1,13	9,20 ± 4,61	14,51 ± 9,89

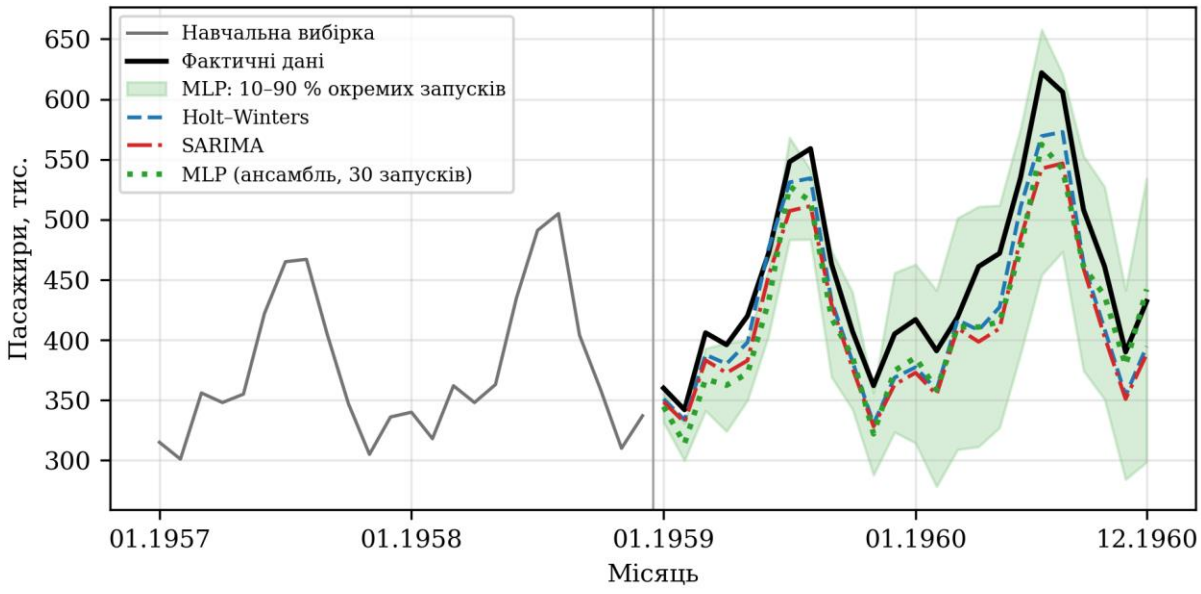


Рис. 1. Прогнози на 1959-1960 рр. і фактичні значення

Принципово інша картина – в окремих запусках MLP: середній MAPE становить 11,86 % зі стандартним відхиленням 6,91 в. п., а розкид – від 2,66 до 34,82 % (медіана 9,61 %; 10-й і 90-й процентилі – 5,63 і 19,46 %). Лише 5 із 30 запусків (17 %) виявилися точнішими за Holt-Winters і 11 із 30 (37 %) – за SARIMA, а 24 із 30 перевершили сезонний наївний прогноз (рис. 2а). Отже, висока точність ансамблю досягається усередненням, яке гасить дисперсію, зумовлену початковими вагами.

Дослідник, який повідомив би лише про найкращий запуск (2,66 %), зробив би висновок про суттєву перевагу нейромережі, а той, хто взяв би найгірший (34,82 %), – про її непридатність; обидва висновки були б хибними.

Показово й систематичне зміщення прогнозів. Середня процентна похибка (MPE; від'ємне значення означає заниження) становить −6,32 % для Holt-Winters, −8,52 % для SARIMA, −7,61 % для ансамблю MLP і −15,52 % для сезонного наївного прогнозу. SARIMA занижує прогноз у всіх 24 місяцях, а Holt–Winters і ансамбль MLP – у 23 із 24. Тому MAPE кожного методу майже збігається з модулем його MPE: похибка зумовлена не випадковими коливаннями, а тим, що сезонна амплітуда у 1959–1960 рр. зростала швидше, ніж у навчальній вибірці. Для окремих запусків MLP середнє зміщення становить −7,61 % (σ = 10,17 в. п.), а 23 із 30 запусків (77 %) занижували прогноз.

Імовірною причиною нестабільності MLP є співвідношення розміру моделі й вибірки: мережа з 16 прихованими нейронами має 225 параметрів (12·16 вхідних ваг, 16 зміщень, 16 вихідних ваг і 1 зміщення), тоді як основна навчальна вибірка дає лише 107 навчальних прикладів (119 приростів без 12 лагів). За такого співвідношення цільова функція має багато локальних мінімумів, а результат оптимізації залежить від початкових ваг, що узгоджується зі спостережуваним розкидом; проте цю гіпотезу спеціально не перевіряли.

Щоб перевірити, чи не зумовлена картина конкретним розбиттям, оцінювання повторено для чотирьох початків прогнозу (навчання до кінця 1956, 1957, 1958 і 1959 рр.) з горизонтом 12 місяців (табл. 2, рис. 2б); кількість нейронів MLP залишено 16, тобто підібрану на основному розбитті. Ранжування методів змінюється залежно від вікна: у 1957 р. найточнішою була SARIMA (1,66 %), у 1958 р. – навіть сезонний наївний прогноз (3,14 %), у 1959 і 1960 рр. – Holt-Winters (4,79 і 2,21 %). За середнім MAPE по чотирьох вікнах Holt-Winters залишається найкращим (3,59 %), далі йдуть SARIMA (4,73 %), ансамбль MLP (5,64 %) і сезонний наївний прогноз (8,73 %). Ансамбль точніший за середній окремий запуск MLP (6,52 %) у кожному з чотирьох вікон, а розкид окремих запусків істотно залежить від вікна: σ коливається від 0,87 до 4,61 в. п.

Таблиця 2. MAPE, %, 12-місячного прогнозу за чотирма початками прогнозу

Вікно (12 міс.)	Сез. наївний	Holt–Winters	SARIMA	MLP, ансамбль	MLP, окремий запуск (± σ)
1957 р.	10,76	3,66	1,66	2,40	2,95 ± 0,87
1958 р.	3,14	3,68	7,60	9,04	9,06 ± 1,46
1959 р.	11,06	4,79	6,76	7,95	9,20 ± 4,61
1960 р.	9,99	2,21	2,90	3,18	4,88 ± 2,24
Середнє	8,73	3,59	4,73	5,64	6,52 ± 2,29

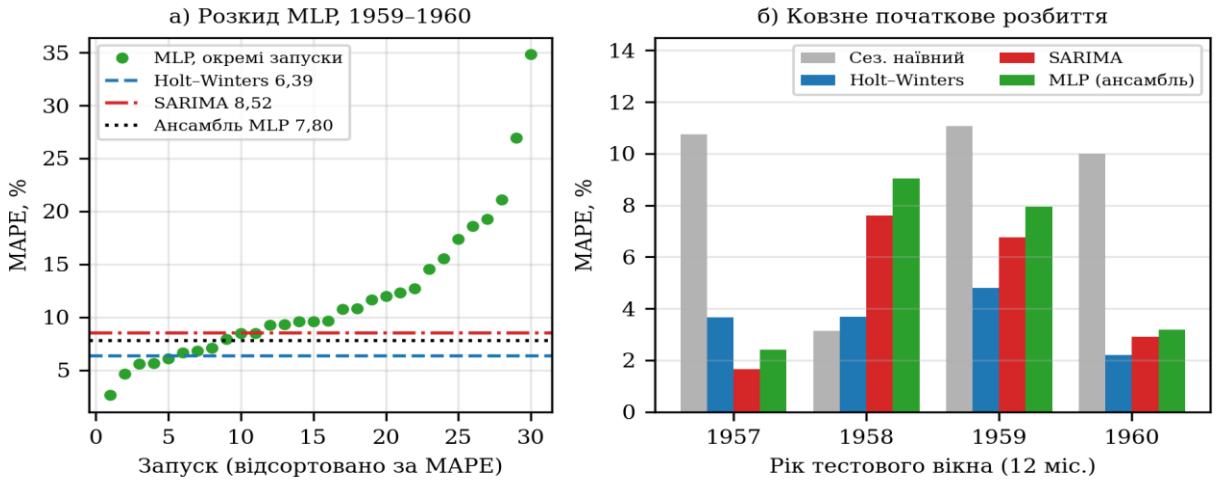


Рис. 2. Розкид MAPE окремих запусків MLP (а) та MAPE за ковзним початком прогнозу (б)

Практичне значення полягає в тому, що для коротких місячних рядів (до кількох сотень спостережень) статистичні моделі доцільно використовувати як обов'язковий еталон, а нейромережу – лише з багаторазовим навчанням (ансамблем) і звітуванням про розкид помилки, а не про один «вдалий» запуск. Додатково слід перевіряти результат на кількох початках прогнозу, бо ранжування методів на одному розбитті може не відтворюватися. Для агровиробництва, де сезонні ряди (споживання ресурсів, обсяги продажу, надходження врожаю) зазвичай короткі, це означає, що простий Holt-Winters або SARIMA є надійною відправною точкою, а перехід до нейромереж виправданий лише за наявності суттєвого приросту точності на незалежних вікнах.

З отриманих результатів випливає простий протокол порівняння статистичних і нейромережевих методів на коротких рядах: (1) обов'язково включати сезонний наївний прогноз і масштабовану за ним похибку MASE; (2) навчати нейромережу не менше кількох десятків разів із різними початковими вагами та наводити середнє, розкид і результат ансамблю; (3) повторювати оцінювання на кількох початках прогнозу; (4) підбирати гіперпараметри лише за валідаційною частиною навчальної вибірки, не користуючись тестовими даними. Дотримання цих кроків не потребує значних ресурсів: усі розрахунки цієї роботи виконуються за лічені хвилини на звичайному комп'ютері.

Перспективи подальших досліджень пов'язані з розширенням набору рядів (зокрема аграрних і енергетичних), перевіркою рекурентних архітектур і гібридних схем «ARIMA + нейромережа» [2], а також із формальним порівнянням точності прогнозів за допомогою статистичних тестів.

Обмеження дослідження: один ряд, фіксований порядок SARIMA, невеликий обсяг навчальних даних (від 96 до 132 спостережень), неповний перебір гіперпараметрів MLP і лише одна архітектура; тому результати не узагальнюються на інші ряди й архітектури, зокрема рекурентні та гібридні [2]. Статистичну значущість відмінностей між методами не перевіряли.

Висновки. На відкритому місячному сезонному ряді Holt-Winters мав найменшу похибку як на основному розбитті (MAPE 6,39 %, MASE 1,01), так і в середньому за чотирма ковзними вікнами (3,59 %). Ансамбль MLP показав 7,80 % і 5,64 % відповідно, SARIMA – 8,52 % і 4,73 %. Окремі запуски MLP мають середній MAPE 11,86 % і значний розкид (σ = 6,91 в. п.), тож точність нейромережі потрібно оцінювати за набором запусків. Результати підтверджують доцільність порівняння з простими статистичними еталонами, але не доводять переваги жодного методу в загальному випадку.

Список використаних джерел

1. Makridakis S., Spiliotis E., Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE. 2018. Vol. 13, No. 3. e0194889. DOI: https://doi.org/10.1371/journal.pone.0194889.

2. Zhang G. P. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing. 2003. Vol. 50. P. 159-175. DOI: https://doi.org/10.1016/S0925-2312(01)00702-0.

3. Box G. E. P., Jenkins G. M., Reinsel G. C., Ljung G. M. Time Series Analysis: Forecasting and Control. 5th ed. Hoboken : Wiley, 2015.

4. Winters P. R. Forecasting sales by exponentially weighted moving averages. Management Science. 1960. Vol. 6, No. 3. P. 324-342.

5. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne : OTexts, 2021. URL: https://otexts.com/fpp3/ (дата звернення: 15.10.2026).

6. Hyndman R. J., Koehler A. B. Another look at measures of forecast accuracy. International Journal of Forecasting. 2006. Vol. 22, No. 4. P. 679-688. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001.