ЖУРНАЛ
ЄВРОПЕЙСЬКОЇ ЕКОНОМІКИ
Том 25. № 1 (96). Січень–березень 2026
Видання Західноукраїнського національного університету

Економічна теорія

Леонідас ТЕОДОРАКОПУЛОС,
Александра ТЕОДОРОПУЛУ,
Евангелос СИСКОС
Євген САВЕЛЬЄВ

ЕКОНОМІЧНА ВАРТІСТЬ ФІНАНСОВИХ ФЕЙКОВИХ НОВИН НА ЄВРОПЕЙСЬКИХ РИНКАХ КАПІТАЛУ

Резюме

Дезінформація у фінансових новинах – це системний ризик, який спотворює механізм виявлення цін та алокацію капіталу, збільшуючи інформаційну асиметрію та послаблюючи довіру інвесторів. У цьому дослідженні розроблено придатний до масштабування рамковий підхід, що поєднує обробку природної мови (NLP) та машинне навчання для виявлення оманливих

Теодоракопулос Леонідас, доктор філософії (Великі дані у менеджменті й економіці), позаштатний професор, кафедра управлінських наук і технологій, Патрський університет, Патри, Греція. ORCID: 0000-0002-0891-6780 Email: theodleo@upatras.gr
Теодоропулу Александра, магістр (Цифрові інновації і менеджмент), аспірант, кафедра управлінських наук і технологій, Патрський університет, Патри, Греція. ORCID: 0009-0004-6314-7795 Email: theodoropouloua@upatras.gr
Сискос Евангелос, доктор економічних наук, професор з міжнародних, європейських та чорноморських економічних відносин, кафедра міжнародних та європейських економічних студій, Університет Західної Македонії, Козані, Греція. ORCID: 0000-0002-5221-4444 Email: esiskos@uowm.gr
Савельєв Євген, доктор економічних наук, професор, кафедра міжнародної економіки, Західноукраїнський національний університет, Тернопіль, Україна. ORCID: 0000-0003-0137-2263 Email: savelyev@wunu.edu.ua

фінансових наративів у великих цифрових корпусах текстів з використанням великих даних. Після нормалізації тексту, лематизації та вилучення стоп-слів, у рамках такого підходу виконано порівняння методів TF-IDF та Word2Vec і проведено навчання класифікаторів Logistic Regression, Random Forest та Gradient Boosting. Ефективність підходу оцінено метриками Accuracy (точність), Precision (влучність), Recall (повнота), F1-score (F1-міра) та ROC-AUC (площа під ROC-кривою). У всіх моделях TF-IDF забезпечує кращу дискримінаційну здатність, ніж Word2Vec. TF-IDF у поєднанні з Random Forest показує майже ідеальні результати (ROC-AUC 0,9999; влучність 0,9977). Сфокусованість на прозорих, заснованих на ознаках моделях забезпечує кращу аудитабельність (наприклад, через важливість ознак) і допомагає обмежити кількість шкідливих хибно позитивних результатів, які можуть подавляти справжні сигнали. Результати показують, що високоточні, придатні до інтерпретації, NLP-конвеєри можуть скоротити верифікаційний розрив у швидкозмінних інформаційних середовищах, зменшити макроекономічні втрати від неправдивих наративів та забезпечувати інформацією процеси ринкового нагляду відповідно до вимог Закону про цифрові послуги (Digital Services Act) та Європейського управління з цінних паперів і ринків (European Securities and Markets Authority, ESMA). Такий підхід призначений для розгортання на потокових новинних стрічках і великих архівах платформ.

Ключові слова:

великі дані, європейські ринки капіталу, економічна вартість, маніпулювання ринком, обробка природної мови, прийняття фінансових рішень, фінансова дезінформація.

Класифікація за JEL: G14, D82, C55.

2 рисунки, 4 таблиці, 14 формул, 20 джерел літератури.

Постановка проблеми

У сучасному цифровому економічному просторі ефективність фінансових ринків все більше залежить від якості та цілісності інформаційних потоків. Хоча швидка диджиталізація засобів масової інформації знизила трансакційні витрати, вона також сприяла появі фінансової дезінформації як значного системного ризику. З економічного погляду фінансові фейкові новини є серйозною формою інформаційної асиметрії, яка спотворює процес виявлення цін на фінансові інструменти (price discovery) і призводить до неоптимальної алокації капіталу.

Макроекономічні наслідки оманливих наративів існують не лише в теорії. Показовий випадок стався у 2013 р., коли одне неправдиве повідомлення про вибух у Білому домі спричинило миттєве падіння ринкової вартості на 130 млрд дол. (Selyukh, 2013). Такі події підкреслюють, як «шум» в інформаційній екосистемі може спровокувати ірраціональну ринкову поведінку і широкомасштабну економічну нестабільність.

Викликом для сучасних європейських економік є те, що обсяг даних, які продукують онлайн-медіа, соціальні мережі та алгоритмічні торгові платформи, зараз випереджає можливості традиційних процедур перевірки людиною. Це створює «прогалину у верифікації», яка може бути використана для маніпулювання ринком. Як наслідок, розробка автоматизованих, придатних до масштабування систем для виявлення недостовірного контенту – це вже не просто технічна задача, а вимога для збереження цілісності ринку та захисту довіри інвесторів.

Метою дослідження є визначення економічних підходів до кількісної оцінки економічної вартості фінансових фейкових новин на європейських ринках капіталу. У дослідженні запропоновано рамковий підхід, у якому поєднуються обробка природної мови (NLP) та машинне навчання (ML) для виявлення характеристик достовірних та неправдивих повідомлень у великих фінансових наборах даних (Du et al., 2024). На відміну від систем виявлення загального призначення, у нашому підході враховано специфічну технічну термінологію та індикатори ринкових емоційних настроїв, притаманні фінансовому дискурсу.

У дослідженні оцінено три моделі керованого навчання – Logistic Regression (логістичну регресію), Random Forest (випадковий ліс) та Gradient Boosting (градієнтний бустинг) за двома методами представлення ознак: TF-IDF та Word2Vec. Ціллю наукового пошуку є створити надійний інструмент моніторингу, здатний виявляти ледь помітні лінгвістичні маркери, що вказують на фінансове шахрайство, що дасть змогу захистити фундаментальні основи стабільної цифрової економіки, роблячи акцент на статистичній стій-

кості моделей та їх інтерпретованості. У Європейському Союзі регуляторний ландшафт набув швидкого розвитку з впровадженням Закону про цифрові послуги (Digital Services Act – DSA), який зобов'язує системно важливі платформи вживати проактивних заходів проти поширення шкідливої дезінформації. З точки зору політики, виклик полягає у розробленні систем, придатних до «аудиту» та «інтерпретації», які здатні відрізнити здорові ринкові спекуляції від скоординованих оманливих наративів, що загрожують цілісності ринків. Зосереджуючись на високоточних і водночас прозорих для аудиторів моделях машинного навчання, у цьому дослідженні запропоновано методологію для операціоналізації цих цілей європейського регулювання.

Стаття охоплює чотири розділи й висновок. У розділі «Огляд літератури» окреслено теоретичні рамки та здійснено огляд релевантних досліджень з інформаційної асиметрії та обчислювальних методів виявлення фейкових новин. У розділі «Методологія» детально описано методологію дослідження, зокрема характеристики набору даних, багатоетапний конвеєр попередньої обробки тексту, а також математичні основи вилучення ознак та використані архітектури машинного навчання. У розділі «Результати дослідження» представлено всебічний аналіз експериментальних результатів із використанням різноманітних метрик та візуалізацій для оцінювання ефективності моделі. У розділі «Обговорення» розглянуто практичну реалізацію рамкового підходу в контексті європейського регулювання та окреслено стратегічні напрямки майбутніх досліджень у сфері фінансової інформатики. У розділі «Висновки» узагальнено ключові результати та відображено широке коло економічних наслідків для цілісності ринку та оцінювання достовірності фінансових новин.

Огляд літератури

Попри зростання кількості досліджень щодо виявлення фейкових новин за допомогою обчислювальних методів, питання фінансової дезінформації вивчено недостатньо. Перший етап дослідження зосереджений на основних методах виявлення фейкових новин з використанням текстового аналізу та стилістично-психолінгвістичних методик виявлення сигналів. Так, Ahmad et al. (2020) продемонстрували ефективність ансамблевих методів шляхом проведення беггінг- та бустинг-тестів на різних наборах даних, довівши, що комбінування моделей підвищує надійність класифікації.

Виявлення ринкових чуток і маніпулятивних повідомлень за допомогою обробки природної мови і машинного навчання продемонстровано у фінансовій сфері (Alshuwaier & Alsulaiman, 2025). Наприклад, Majumdar та Bose (2018) представили рамковий підхід на основі великих даних для виявлення фінансових чуток навколо Бомбейської фондової біржі, а Cheng et al. (2023) використали сигнали чуток у соціальних мережах та класифікатори машинного на-

вчання для прогнозування аномальної торгівлі. За допомогою текстових сигналів з платформ мессенджерів також були виявлені маніпулятивні кампанії, зокрема «накачування і скидання» криптовалют (Nghiem et al., 2021). Лінгвістичні закономірності у фінансовому дискурсі, наприклад, специфічні формулювання в процесі розкриття подій, повторення назв суб'єктів та екстремальність формулювань, відповідають маркерам обману, які були задокументовані у фінансових комунікаціях (Larcker & Zakolyukina, 2012), а також галузевим лексиконам, необхідним для моделювання фінансової мови (Loughran & McDonald, 2011). Ця галузь досі характеризується браком загальноприйнятих фінансових бенчмарків для виявлення дезінформації. Саме для подолання цієї прогалини запропоновано низку нових ініціатив, зокрема «Fin-Fact» (Rangapur et al., 2025), а в ширших оглядових дослідженнях наголошується на загальній фрагментованості наявних наборів даних (D'Ulizia et al., 2021).

Alghamdi et al. (2024) і Mishra et al. (2022) у своєму дослідженні виокремили три категорії методів виявлення фейкових новин: контент-орієнтовані, контекст-орієнтовані та гібридні підходи. Автори продемонстрували, що класичні моделі машинного навчання забезпечують високу інтерпретованість і таку саму ефективність завдяки належно проведеній попередній обробці та методикам інженерії ознак. Ці результати покладено в основу дослідження, якому поєднано контентний аналіз із структурованими текстовими ознаками, оскільки ці методи відповідають вимогам фінансових застосунків щодо прозорості та придатності до масштабування.

Ця робота розширює попередні дослідження завдяки застосуванню аналізу фінансових текстів у поєднанні із класичними системами обробки природної мови (NLP) та машинного навчання (ML), які вивчають механізм впливу різних типів ознак та архітектури класифікаторів на специфічні для галузі текстові паттерни. Результати дослідження допомагають визначити як переваги, так і недоліки використання традиційних моделей для виявлення фінансової дезінформації.

Економічна вартість фінансової інформаційної асиметрії

З економічної точки зору фейкові фінансові новини становлять серйозну форму інформаційної асиметрії, що спотворює ефективність ринку. Поширення недостовірного контенту створює «шум», який заважає процесу виявлення цін (price discovery), що призводить до неоптимальної алокації капіталу. Злам Associated Press у 2013 р. (Selyukh, 2013) показав, що одна одиниця дезінформації може призвести до миттєвого руйнування багатства в макроекономічному масштабі. Тому автоматизовані системи виявлення слугують критично важливою інфраструктурою для збереження цілісності ринку

та захисту довіри інвесторів – фундаментальних принципів стабільної цифрової економіки. У літературі все частіше визнається, що фінансова дезінформація – це не просто соціальна проблема, а структурна неефективність економіки (Lyzun et al., 2019; Lyzun et al.,2023). Якщо ранні дослідження в цій сфері зосереджувалися на простому фільтруванні за ключовими словами, то перехід до аналітики великих даних дозволяє глибше зрозуміти, як оманливі наративи оминають ринкові фільтри. У європейському контексті, де фрагментація ринку між різними країнами є високою, здатність NLP-моделей підтримувати високу влучність під час використання різноманітних джерел фінансових новин має вирішальне значення для запобігання транскордонному інформаційному арбітражу.

Методологія

Методологія дослідження складається з чотирьох основних етапів: (1) підготовка набору даних, (2) попередня обробка тексту, (3) вилучення ознак і (4) класифікація моделей. У дослідженні також реалізовано два паралельні NLP-конвеєри з використанням представлень TF-IDF і Word2Vec.

Набір даних та анотація

Джерелом набору даних, які використані у дослідженні, є загальнодоступний датасет «Фейкові та справжні новини» (Fake and Real News Dataset) на Kaggle (kaggle.com, n.d.), відфільтрований для статей на фінансову та економічну тематику. Датасет складається з двох окремих CSV-файлів:

- Fake.csv: містить сфабриковані новини (23 481 запис);
- True.csv: містить правдиві новини (21 417 записів).

Кожен файл містить такі колонки:

- «Title»: заголовок новини;
- «Text»: повний текст статті;
- «Subject»: широка категорія теми;
- «Date»: дата публікації статті.

Для цілей дослідження вручну присвоєно бінарні класифікаційні мітки:

- Мітка 0: фейкові новини (з файлу Fake.csv);
- Мітка 1: правдиві новини (з файлу True.csv).

Обидва набори даних об'єднано в єдиний корпус із 44 898 новинних статей для аналізу. Поля «title» (заголовок) і «text» (текст) об'єднано в єдиний вхідний елемент («content») для кожного випадку. Колонки «subject» (тема) і «date» (дата) не використовувалися для моделювання, але дослідники можуть їх використовувати для подальших експлоративних досліджень і часового аналізу. В об'єднаному датасеті записи перемішані у випадковому порядку, щоб запобігти упередженню через порядок розташування під час навчанні моделі.

Попередня обробка тексту

Текстові дані потребували попередньої обробки для отримання корисних ознак, оскільки виконано багато операцій з нормалізації та очищення. Ці операції були важливим інструментом, який забезпечив усунення непотрібних точок даних і неузгодженої інформації, водночас мінімізуючи мовні відмінності, що в результаті покращило якість вхідних даних для моделі машинного навчання.

- **Застосування малих літер:** увесь текст було переведено в нижній регістр для забезпечення однорідності. Завдяки цьому кроку, модель не сприйматиме такі слова, як «Trump» і «trump», як різні сутності, що зменшує надлишковість словника.
- **Видалення цифр і розділових знаків:** за допомогою регулярних виразів та методів обробки рядків видалено всі числа та розділові знаки з тексту. Такі елементи в цьому контексті не містять значущого смислового навантаження, тому їх видалення спрощує текст.
- **Токенізація:** Потім текст було розділено на окремі слова, або токени. Це дало змогу застосувати до кожного слова такі подальші операції, як видалення стоп-слів і лематизацію. У TF-IDF-конвеєрі використовувалося базове розбиття рядків, а в конвеєрі на основі Word2Vec – gensim.simple_preprocess, що виявилося достатнім через його неконтекстну природу.
- **Видалення стоп-слів:** Поширені англійські стоп-слова (такі як «the», «is» та «and») було вилучено за допомогою списку NLTK. Такі високочастотні слова мало сприяють розрізненню фейкових і справжніх новин та часто додають шум у модель.
- **Лематизація:** Кожне слово було лематизовано за допомогою WordNetLemmatizer з NLTK. Цей інструмент приводить слова до їхньої базової або словникової форми (наприклад, «running» перетворюється на «run»), що допомагає згрупувати різні форми слова

Том 25. № 1 (96). Січень–березень 2026
ISSN 1684-906X

під єдиним представленням, зменшуючи розмірність і покращуючи семантичну узгодженість.

Такі методи попередньої обробки тексту застосовано до обох конвеєрів (Word2Vec і TF-IDF), щоб вилучення ознак виконувалося на чистому і семантично значущому тексті.

Вилучення ознак

Для підготовки тексту до аналізу за допомогою машинного навчання застосовано два окремі методи числового перетворення – вбудовування Word2Vec та векторизацію TF-IDF. Ці методи становлять два протилежні підходи до представлення тексту, оскільки в них для створення представлень використовуються контекстне значення та статистична частота слів.

Вбудовування Word2Vec

Word2Vec – це методика обробки природної мови (NLP) на основі неглибокої двошарової нейромережевої моделі, розроблена компанією Google; вона вивчає розподілені представлення слів у неперервному векторному просторі таким чином, щоб семантично схожі слова були розташовані на сусідніх точках. У цьому експерименті ми тренували модель Skip-gram Word2Vec на корпусі тексту з розміром вектора d =100, розміром вікна w = 5 і мінімальним порогом частоти слова – 2.

Модель Skip-gram максимізує ймовірність оточуючого контексту для заданого слова. Для послідовності навчальних слів w1,w2,...,wT цільова функція має вигляд:

$$\max \frac{1}{T} \sum_{t=1}^{T} \sum_{-w \le j \le w, j \ne 0} \log P(w_{t+j} \mid w_t), \qquad (1)$$

де умовна ймовірність $P(w_{t+j} \mid w_t)$ обчислюється за допомогою softmax-функції:

$$P(w_0 \mid w_1) = \frac{\exp(u_{w_0}^T u_{w_1})}{\sum_{w=1}^{|V|} \exp(u_w^T u_{w_1})}. \qquad (2)$$

Тут: u_{w_1} – вхідний вектор центрального слова; u_{w_0} – вихідний вектор контекстного слова; $|V|$ – розмір словника.

Після того, як вбудовування слів вивчені, кожен документ перетворюється на вектор фіксованої довжини через усереднення векторів слів, що

входять до його складу. Для документа, представленого набором з n допустимих токенів $\{w_1, w_2, \ldots, w_n\}$, вектор документа $\vec{d}$ обчислюється як:

$$\vec{d} = \frac{1}{n}\sum_{i=1}^{n} \vec{w}_i. \quad (3)$$

Векторизація TF-IDF

Другою методикою вилучення ознак була Term Frequency–Inverse Document Frequency або TF-IDF (з англ. частота терміна – обернена частота документа) – широко використовувана статистична міра, яка відображає важливість слова в документі відносно корпусу тексту.

Вага TF-IDF для терміна *t* в документі *d* визначається як:

$$TF - IDF(t,d) = TF(t,d) * IDF(t), \quad (4)$$

де:

- **Частота терміна (Term Frequency – TF)** – це нормалізована частота появи терміна в документі:

$$TF(t,d) = \frac{f_{t,g}}{\sum_{t' \in d} f_{t',d}}, \quad (5)$$

де $f_{t,d}$ – початкове число згадувань терміна *t* в документі *d*.

- **Зворотна частота документа (Inverse Document Frequency – IDF)** показує, наскільки рідко вживається термін у всіх документах:

$$IDF(t) = \log\left(\frac{N}{1+n_t}\right), \quad (6)$$

де: N – загальна кількість документів; n_t – кількість документів, що містять термін *t*.

TF-IDF присвоює більшу важливість термінам, які часто зустрічаються в цьому документі, але рідко трапляються в усьому корпусі тексту, що допомагає зменшити вплив загальновживаних, але менш інформативних слів. У нашій реалізації використано TfidfVectorizer з scikit-learn зі словником із 5,000 найпоширеніших термінів, впорядкованих за частотою. У результаті отримано розріджену матрицю розмірністю N×5000, де кожен рядок відповідає вектору документа, а кожен стовпець – терміну в словнику.

Тренування та оцінка моделі

Для класифікації новинних статей на фейкові та правдиві використано три класичні класифікатори машинного навчання: Logistic Regression (логістичну регресію), Random Forest (випадковий ліс) і Gradient Boosting (градієнтний бустинг). Кожна модель застосовувалася з використанням обох представлень ознак (Word2Vec і TF-IDF), а їхня ефективність оцінювалася за допомогою єдиного набору кількісних метрик.

Logistic Regression

Logistic Regression (логістична регресія) – це лінійний імовірнісний класифікатор, який моделює ймовірність бінарної мітки за допомогою логістичної (сигмоїдної) функції. Для вхідного вектора $\vec{x} \in R^d$ ймовірність того, що вихідна мітка $y \in \{0,1\}$ є істинною (тобто, новина є реальною), задається формулою:

$$P(y = 1 \mid \vec{x}) = \sigma(\vec{w}^T \vec{x} + b) = \frac{1}{1 + e^{-(\vec{w}^T \vec{x} + b)}}, \quad (7)$$

де: $\vec{w}$ – ваговий вектор; b – зміщення; $\sigma(.)$ – сигмоїдна функція активації.

Параметри моделі $\vec{w}$ та b визначаються шляхом мінімізації логарифмічної функції втрат (крос-ентропії) на навчальній вибірці:

$$L = -\frac{1}{N} \sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]. \quad (8)$$

Random Forest

Random Forest – це алгоритм ансамблевого навчання, який працює за допомогою побудови великої кількості дерев прийняття рішень під час навчання і виводить моду (більшість голосів) прогнозів окремих дерев прийняття рішень для задач класифікації.

Кожне дерево рішень навчається на випадковій бутстрап-вибірці навчальних даних, а при кожному розгалуженні дерева розглядається випадкова підмножина ознак. Це забезпечує різноманітність дерев і знижує перенавчання.

Прогноз для вхідних даних $\vec{x}$ задається формулою:

$$y = majority\_vote(h_1(\vec{x}), h_2(\vec{x}), ..., h_T(\vec{x})), \quad (9)$$

де: $h_T(\vec{x})$ – прогноз t-го дерева прийняття рішень, T – загальна кількість дерев у лісі.

Gradient Boosting

Gradient Boosting (градієнтний бустинг) – це ще один ансамблевий метод, у якому поетапно будується послідовність слабких моделей або учнів (як правило, неглибоких дерев прийняття рішень). Кожен новий учень зосереджується на виправленні залишків (помилок) попередніх учнів шляхом мінімізації диференційованої функції втрат (як правило, логарифмічної функції втрат для бінарної класифікації).

Нехай $F_m(\vec{x})$ позначає модель після *mmm* ітерації. Вона оновлюється так:

$$F_{m+1}(\vec{x}) = F_m(\vec{x}) + \eta * h_m(\vec{x}), \tag{10}$$

де: $h_m(\vec{x})$ – базовий учень, пристосований до від'ємного градієнту втрат; η – коефіцієнт навчання (параметр стиснення); $F_0(\vec{x})$ ініціалізується константою (наприклад, середніми лог-шансами).

Метрики оцінки

Для комплексної оцінки ефективності кожної моделі застосовано кілька стандартних класифікаційних метрик:

- **Точність**: частка правильно передбачених випадків.

$$Точність = \frac{Істинно\ позитивні + Істинно\ негативні}{Істинно\ позитивні + Істинно\ негативні + \\ + Хибно\ позитивні + Хибно\ негативні}. \tag{11}$$

- **Влучність**: частка передбачених позитивних випадків, які дійсно є позитивними (корисно для мінімізації кількості хибно позитивних випадків).

$$Влучність = \frac{Істинно\ позитивні}{Істинно\ позитивні + Хибно\ позитивні}. \tag{12}$$

- **Повнота (чутливість)**: частка істинно позитивних результатів, яка була правильно визначена (корисно за мінімізації кількості хибно негативних результатів).

$$Повнота = \frac{Істинно\ позитивні}{Істинно\ позитивні + Хибно\ негативні}. \tag{13}$$

- **F1-міра (F1-Score)**: середнє гармонічне влучності і повноти, що забезпечує баланс між ними.

Том 25. № 1 (96). Січень–березень 2026
ISSN 1684-906X

$$F1\_міра = 2 * \frac{Влучність * Повнота}{Влучність + Повнота}. \quad (14)$$

- **ROC-AUC (площа під ROC-кривою)**: вимірює площу під кривою робочої характеристики приймача (Receiver Operating Characteristic – ROC), яка відображає співвідношення частки істинно позитивних результатів (True Positive Rate – TPR) до частки хибно позитивних результатів (False Positive Rate – FPR) при різних порогових значеннях. Чим ближче значення площі під кривою (AUC) до 1,0, тим вища здатність моделі до дискримінації між класами.
- **ROC-крива:** графічний інструмент, який ілюструє компроміс між чутливістю та специфічністю за різних порогів прийняття рішення.

Результати дослідження

У цьому розділі наведено результати експериментів із виявлення фінансових фейкових новин за допомогою моделей класифікації, а потім подано детальний аналіз їхньої ефективності. З точки зору регулювання, висока влучність (0,9977) моделі випадкового лісу є більш важливою, ніж її загальна точність. На фінансових ринках «хибно позитивний результат» (помилкове позначення легітимних новин як фейкових) може бути настільки ж шкідливим, як і «хибно негативний результат», оскільки може призвести до придушення дійсних ринкових сигналів. Статистична стійкість, яку демонструють криві ROC-AUC (0,9999), свідчить про те, що ця система забезпечує надійну «мережу безпеки», яку можна розгорнути без внесення нових викривлень у процес виявлення ціни. У дослідженні порівнюються три популярні класифікатори машинного навчання, які також застосовують методи вилучення ознак TF-IDF і Word2Vec: (а) логістична регресія, (б) випадковий ліс і (в) градієнтний бустинг. Для оцінювання загальної ефективності кожної моделі використано стандартні показники, зокрема точність, влучність, повнота, F1-міра і ROC-AUC. Ціллю дослідження є визначити, яка модель працює найкраще у процесі використання різних представлень, водночас аналізуючи, як різні типи ознак впливають на показники ефективності класифікатора. Результати дослідження представлені у вигляді таблиць з кількісними даними і візуалізацій ROC-кривих, які дають змогу наочно проаналізувати ефективність підходу до виявлення фінансових фейкових новин.

Порівняння моделей Logistic Regression

Класифікатор логістичної регресії продемонстрував майже ідеальні результати в процесі використання обох типів ознак, проте версія TF-IDF дещо перевершила версію Word2Vec (табл. 1). Зокрема, показник точності TF-IDF становив 0,9895, якщо порівняти з 0,9636 у Word2Vec, а значення F1-міри зросло з 0,9624 до 0,9892. Модель TF-IDF також продемонструвала кращий баланс між влучністю та повнотою, що свідчить про вищу здатність до узагальнення. Значення ROC-AUC для обох моделей перевищували 0,99, але модель TF-IDF знову продемонструвала дещо вищий показник – 0,9992.

Таблиця 1

Результати моделей логістичної регресії

Метрика	TF-IDF	Word2Vec
Точність	0,9895	0,9636
Влучність	0,9844	0,9592
Повнота	0,9940	0,9656
F1-міра	0,9892	0,9624
ROC-AUC	0,9992	0,9942

Джерело: власні розрахунки.

Цей розрив у ефективності ілюструє, що для лінійних моделей, таких як логістична регресія, високорозмірні розріджені представлення, такі як TF-IDF, забезпечують більшу дискримінаційну здатність, особливо коли текстові шаблони відрізняються між класами, як це часто буває у випадку з фейковими та реальними новинами.

Порівняння моделей Random Forest

Моделі Random Forest показали найбільшу різницю в ефективності між двома способами представлення ознак. У поєднанні з TF-IDF класифікатор Random Forest продемонстрував майже ідеальні результати: точність і F1-міра становили 0,9978, а ROC-AUC – 0,9999. Натомість, варіант на основі Word2Vec показав суттєво нижчі результати: точність – 0,9548 і F1-міра – 0,9527 (табл. 2).

Table 2

Результати моделей Random Forest

Метрика	TF-IDF	Word2Vec
Точність	0,9978	0,9548
Влучність	0,9977	0,9610
Повнота	0,9977	0,9446
F1-міра	0,9977	0,9527
ROC-AUC	0,9999	0,9916

Джерело: власні розрахунки.

Цю різницю можна пояснити принципом роботи моделей Random Forest. Оскільки дерева прийняття рішень в ансамблі покладаються на різні пороги та розбиття вхідного простору ознак, розріджені та багатовимірні вектори TF-IDF дозволяють точніше розділяти простір ознак. Натомість щільні та усереднені вектори Word2Vec згладжують багато інформативних відмінностей між документами.

Порівняння моделей Gradient Boosting

У моделях Gradient Boosting також перевагу отримала TF-IDF. З TF-IDF модель досягла точності 0,9953 і F1-міри 0,9952, а в моделі з Word2Vec точність була нижчою на рівні 0,9359, а F1-міра становила 0.9337 (табл. 3). Хоча Gradient Boosting у поєднанні з Word2Vec все-таки показала хороші результати, зниження ефективності відповідало картині, яка простежувалася в інших моделях.

Таблиця 3

Результати моделей Gradient Boosting

Метрика	TF-IDF	Word2Vec
Точність	0,9953	0,9359
Влучність	0,9940	0,9301
Повнота	0,9963	0,9374
F1-міра	0,9952	0,9337
ROC-AUC	0,9987	0,9858

Джерело: власні розрахунки.

Це ще раз підтверджує, що контекстно-незалежне векторне усереднення у Word2Vec може відкидати критичні ознаки, а TF-IDF може їх зберігати, безпосередньо враховуючи частоту вживання термінів та рідкість згадування їх у документах.

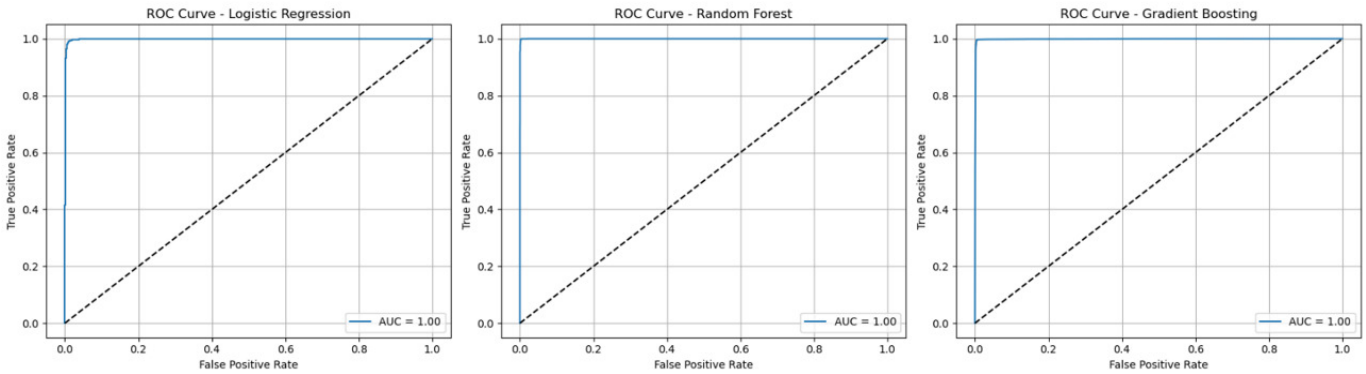
ROC-криві для TF-IDF

ROC-криві класифікаторів логістичної регресії, випадкового лісу та градієнтного бустингу з використанням представлення ознак TF-IDF демонструють виняткову ефективність класифікації в усіх трьох моделях (рис. 1). Криві ідеально прилягають до верхнього лівого кута ROC-простору, оскільки моделі досягають оптимального балансу між часткою істинно позитивних результатів (чутливістю) та часткою хибно позитивних прогнозів. Форма кривих сама вказує на те, що класифікатори мають високу здатність розрізняти фейкові та справжні новини.

Як бачимо, моделі демонструють показники AUC (площа під кривою), які наближаються до 1,00, що свідчить про їхню виняткову ефективність. Модель Random Forest показує майже ідеальний вертикальний підйом з горизонталлю на вершині, що вказує на здатність забезпечити високий рівень істинно позитивних прогнозів без хибних спрацьовувань на всіх порогах. Модель логістичної регресії формує криву, яка є одночасно плавною та крутою, завдяки генерації надійних ймовірнісних результатів. Нарешті, модель Gradient Boosting демонструє стабільні результати, проте початковий сегмент кривої має більш пологий нахил, ніж інші моделі.

Рисунок 1

ROC-криві для TF-IDF



Примітка: True Positive Rate – частка істинно позитивних спрацьовувань, False Positive Rate – частка хибно позитивних спрацьовувань, ROC-curve – Logistic Regression – ROC-крива для моделі Logistic Regression, ROC-curve – Random Forest – ROC-крива для моделі Random Forest, ROC-curve – Gradient Boosting – ROC-крива для моделі Gradient Boosting.

Джерело: власні розрахунки.

Наведені результати підтверджують, що інструмент TF-IDF забезпечує стійкий дискримінативний простір ознак для всіх трьох класифікаторів. Майже ідеальні ROC-криві свідчать про те, що моделі, використані в цих експериментах, не лише точні у своїх кінцевих прогнозах, а й добре відкалібровані для ранжування ймовірностей справжності новин. ROC-аналіз підтверджує висновок, що класичні моделі, зокрема Random Forest і Gradient Boosting у поєднанні з TF-IDF, забезпечують високоефективні рішення для виявлення фейкових новин.

ROC-криві для Word2Vec

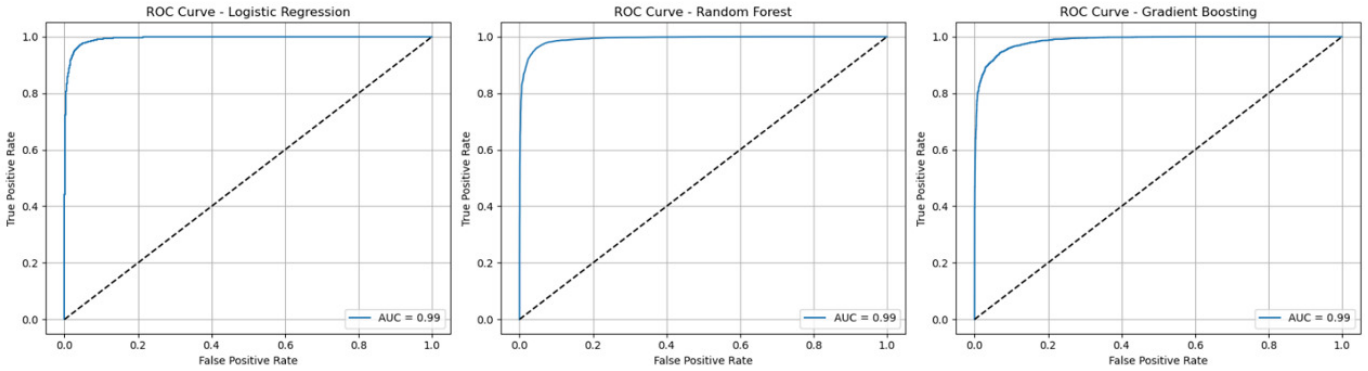
ROC-криві для класифікаторів Logistic Regression, Random Forest і Gradient Boosting з використанням вбудовувань Word2Vec показують, що всі три моделі досягають високої ефективності класифікації (рис. 2). Криві демонструють крутий висхідний підйом до верхнього лівого кута ROC-простору, що свідчить про високий рівень істинно позитивних прогнозів за різних порогових значень. Усі три моделі успішно розрізняють справжні і сфабриковані новини завдяки низькому рівню хибно позитивних результатів.

Крива для Logistic Regression є плавною, але демонструє різке зростання, оскільки ця модель є лінійною та генерує ймовірнісні значення. З іншого боку, крива для Random Forest швидко зростає на початку, оскільки модель забезпечує хороші показники, ефективно розбиваючи простір ознак у міру зростання щільності вхідних векторів. Крива для Gradient Boosting стабільно зростає, що відображає намагання моделі поліпшити показник повноти, утримуючи мінімальний рівень хибно позитивних результатів. Це вказує на її здатність підвищувати точність класифікації шляхом поетапної корекції помилок.

Три моделі мають значення AUC 0,99; це означає, що вони працюють на відмінному, майже ідеальному рівні. Форма та розташування кривих підтверджують, що векторні представлення Word2Vec ефективно захоплюють релевантні семантичні ознаки для бінарної класифікації в цих експериментах. Обидва класифікатори демонструють надійні результати завдяки щільним векторним представленням, попри використання різних стратегій навчання.

Рисунок 2

ROC-криві для Word2Vec



Примітка: True Positive Rate – частка істинно позитивних результатів, False Positive Rate – частка хибно позитивних результатів, ROC-curve – Logistic Regression – ROC-крива для моделі Logistic Regression, ROC-curve – Random Forest – ROC-крива для моделі Random Forest, ROC-curve – Gradient Boosting – ROC-крива для моделі Gradient Boosting.

Джерело: власні розрахунки.

Порівняння ROC-кривих

У цьому дослідженні ROC-криві для трьох класифікаторів на основі використання векторних представлень Word2Vec показують, що всі три моделі досягають високої дискримінаційної здатності, хоча їхня продуктивність дещо нижча, ніж у класифікаторів з використанням TF-IDF. Криві демонструють високу якість класифікації, оскільки вони розташовані вище базової лінії випадкового вгадування, представленої діагональною лінією. Водночас простежується помітне пом'якшення форми кривих порівняно з моделями TF-IDF, що вказує на дещо знижену здатність розрізняти класи за усіх порогових значень.

Як зазначалось вище, усі три моделі генерують значення AUC, які наближаються до 0,99, що свідчить про їхню відмінну ефективність. Модель логістичної регресії дає плавно зростаючу криву завдяки тому, що у ній для отримання результатів використовується ймовірнісний скоринг. Моделі Random Forest та Gradient Boosting демонструють високі рівні істинно позитивних результатів за низьких значень хибно позитивних прогнозів, але їхні криві ефективності мають більш пологий нахил та округлішу форму на початку, ніж моделі TF-IDF.

Така різниця в показниках ефективності, ймовірно, пов'язана з усередненим характером представлень Word2Vec, який може розмивати специфічні лексичні ознаки, важливі для розрізнення фейкових і справжніх новин. Хоча Word2Vec добре фіксує семантичні зв'язки, втрата специфіки на рівні окремих слів, схоже, знижує його здатність підтримувати деревоподібні моделі, зокрема Random Forest і Gradient Boosting, порівняно з TF-IDF. Загалом хоча ROC-криві підтверджують, що Word2Vec забезпечує достатньо ефективну класифікацію, вони також узгоджуються з попереднім висновком: TF-IDF демонструє кращі результати в цьому завданні завдяки своїй здатності зберігати інформацію про частотність і важливість слів, що є критично важливим для виявлення ледь помітних стилістичних і лексичних патернів, притаманних контенту фейкових новин.

Обговорення

Практичне впровадження та наслідки для економічної політики

Перехід від теоретичної концепції виявлення фейкових новин до функціонального інструменту ринкового нагляду вимагає чіткої стратегії впровадження, яка відповідає європейському порядку денному цифрового розвитку. Для європейської економіки практична цінність запропонованої моделі обробки природної мови (NLP) полягає в її здатності зменшити системний «шум», який перешкоджає ефективній алокації капіталу і порушує механізм виявлення цін.

Інтеграція в європейську систему ринкового нагляду

Висока статистична надійність конвеєра Random Forest і TF-IDF уможливлює його пряму інтеграцію в інфраструктуру нагляду національних фінансових органів і органів рівня ЄС, таких як Європейське управління з цінних паперів і ринків (ESMA). Використання цієї моделі як автоматизованого превентивного фільтра для високочастотних новинних стрічок і потоків даних у соціальних мережах дозволяє регуляторним органам зайняти більш проактивну позицію у протидії ринковим маніпуляціям. Крім того, притаманна архітектурі Random Forest інтерпретованість забезпечує можливість документування метрик важливості ознак, таких як специфічні лінгвістичні маркери або сфабриковані тригери настроїв. Така прозорість має критично важливе значення для забезпечення дотримання законодавства та вимог Регламенту про зловживання на ринку (Market Abuse Regulation – MAR) (European Parliament & Council of the European Union, 2014), створюючи слід, що підлягає аудиту, якого бракує моделям глибокого навчання типу «чорних скриньок».

Інституційні захисні бар'єри і захист механізму виявлення цін

Окрім використання у цілях регулювання, запропонований підхід здатний забезпечити захист як для інституційних, так і для роздрібних інвесторів. Інформаційна асиметрія зазвичай ставить у невигідне становище учасників, які не мають ресурсів для миттєвої перевірки фактів. Впровадження такого підходу дасть можливість агрегаторам фінансових новин і трейдинговим терміналам присвоювати цифровому контенту «рейтинг лінгвістичної достовірності», що дозволятиме учасникам ринку розрізняти повідомлення, що базуються на фактах, і спекулятивну дезінформацію. Також у середовищі високочастотного трейдингу (high-frequency trading – HFT) вихідні дані моделі можуть стати критично важливим параметром у ризик-менеджменті. Якщо но-

вина отримує позначку повідомлення з високою ймовірністю дезінформації, автоматизовані системи можуть бути запрограмовані на зниження чутливості до відповідного потоку новин. У такому разі цей механізм фактично виконуватиме роль «цифрового вимикача», який дасть змогу запобігати каскадним розпродажам активів, що спричинені поширенням сфабрикованих даних.

Операціоналізація Закону про цифрові послуги (Digital Services Act – DSA) в ЄС

Згідно з чинним європейським законодавством, системно важливі платформи несуть дедалі більшу відповідальність за фінансові та суспільні ризики, що виникають у їхніх мережах. Це дослідження пропонує технічний орієнтир для «проактивного зменшення системних ризиків», як це передбачено статтею 35 Закону про цифрові послуги (DSA) (European Parliament & Council of the European Union, 2022). У статті надано стандартизовану методологію, яка дозволяє платформам виявляти контент, спрямований на маніпулювання ринком, за допомогою процесу, який є прозорим для незалежних аудиторів і акредитованих дослідників. Імплементація цих NLP-бенчмарків дасть змогу європейським ЗМІ та соціальним платформам сприяти створенню більш прозорої цифрової екосистеми, гарантуючи, що швидкість поширення цифрової інформації не ставитиме під загрозу її точність, необхідну для забезпечення стабільної та резильєнтної економіки.

Ефективність використання ресурсів та транскордонна масштабованість

В умовах цифрового ринку, де новини поширюються за мілісекунди, традиційної перевірки фактів вручну недостатньо. Обчислювальна ефективність підходу на основі TF-IDF забезпечує можливість одночасної обробки величезних обсягів даних у всіх 27 країнах-членах ЄС. Така придатність до масштабування має критично важливе значення для запобігання інформаційному арбітражу, коли дезінформація використовується для отримання вигоди від різниці в цінах між фрагментованими національними ринками ще до того, як втрутиться нагляд за участі людини. Зрештою, розгортання цього підходу сприятиме реалізації мети Європейського Союзу щодо формування об'єднаного, безпечного та прозорого Єдиного цифрового ринку.

Напрями майбутніх досліджень

Наші результати свідчать про те, що стандартні моделі машинного навчання з ретельно вибудуваним NLP-конвеєром здатні з високою точністю виявляти неправдиві фінансові статті на тестових наборах даних. Проте для створення практичної системи, придатної для використання редакціями новин або регуляторними органам, однієї лише точності на еталонному наборі даних недостатньо. Така система повинна працювати в різних медіа та мовних середовищах, зберігати надійність у разі зміни тематики та надавати пояснення щодо своїх позначок про недостовірні новини у такий спосіб, щоб людина могла їх перевірити. Для просування в цьому напрямі подальші дослідження можна зосередити на п'яти напрямках. По-перше, створення більш репрезентативних, специфічних для сфери фінансів наборів даних, що відображають реальні паттерни публікацій. По-друге, використання мовних моделей, спеціально налаштованих на фінансові тексти так, щоб система коректно розуміла стилі журналістських повідомлень та назви суб'єктів. По-третє, поєднання текстових даних з базовими ринковими сигналами (ціни, обсяги, волатильність) для зменшення кількості помилкових спрацьовувань. По-четверте, можливість керувати таймінгом і механізмами поширення інформації: новини переміщуються між джерелами і платформами, а моделі повинні враховувати це і залишатися стабільними в умовах таких змін. По-п'яте, впровадження захисних бар'єрів, таких як аналіз людиною, моніторинг дрейфу і належне ведення журналів, що забезпечить можливість аудиту і поступового вдосконалення системи (Hu et al., 2022; Theodorakopoulos et al., 2025). У табл. 4 узагальнено ці п'ять можливих напрямків майбутніх досліджень.

Таблиця 4

Напрями майбутніх досліджень

Категорія	Фокус дослідження	Необхідні дані/сигнали	Методи (коротко)	Ризики & етика
Збагачення і диверсифікація даних	Створення спеціалізованих фінансових репрезентативних датасетів для виявлення дезінформації.	Анотація новин і звітності, верифіковані фактчекінги, соціальні потоки (X/Reddit/StockTwits), багатомовні джерела.	Чіткі інструкції щодо анотації; слабкий контроль для масштабування; публікація описів датасетів.	Упереджена розмітка, питання конфіденційності і умов використання, нерівномірне покриття ринків.

Категорія	Фокус дослідження	Необхідні дані/сигнали	Методи (коротко)	Ризики & етика
Доменні моделі і адаптація	Використання мовних моделей, налаштованих на фінансову сферу з короткими і обґрунтованими виведеннями.	Ваш власний корпус + фінансовий банк фраз/стенограми звітів про доходи, словники сутностей.	Донавчання FinBERT/FinancialBERT або використання адаптерів, вилучення обґрунтувань.	Галюциновані обґрунтування, витік чутливого тексту.
Мультимодальний зв'язок з ринком	Об'єднання тексту з ринковими даними для зниження кількості хибно позитивних / хибно негативних прогнозів.	Текст новин плюс OHLCV, волатильність, активність на ринку опціонів, часові мітки.	Пізнє злиття / механізм уваги; зв'язування подій між статтями і цінами.	Помилкове відстеження нормальних цінових рухів; ще більше посилення коливань як наслідок реагування на них.
Темпоральна / графова стійкість	Моделювання розтягнуте в часі і стійкість до змін	Зв'язки джерело – стаття – сутність з часовими мітками, графи взаємодії платформ.	Темпоральні GNN/процеси Хоукса, тести на адверсаріальні атаки / перефразування.	Перенавчання під конкретну платформу, крихкість за відсутності зв'язків.
Розгортання, моніторинг & управління	Експлуатація системи в реальному часі з можливістю її аудиту і участю людини в циклі.	Потокове завантаження даних, зворотний зв'язок від аналітиків, сигнали дрейфу, журнали пояснень.	Панель тріажу, виявлення дрейфу / перекалібрування; SHAP / візуалізація уваги.	Надлишкове / недостатнє блокування, справедливість щодо джерел, аудитні журнали і картки моделей.

Джерело: складено авторами.

Висновки

У дослідженні запропоновано стійкий рамковий підхід до виявлення неправдивих фінансових наративів, спрямований на вирішення ключової проблеми забезпечення стабільності та ефективності сучасної цифрової економіки. Переносячи аналітичний фокус із суто технічного завдання класифікації на ширший механізм захисту цілісності ринку, автори демонструють, що обробка природної мови (NLP) є критично важливим інструментом для зменшення інформаційної асиметрії, яка традиційно призводить до неоптимальної алокації капіталу та системних втрат багатства.

Емпіричні результати дослідження підтверджують, що класичні ансамблеві методи, зокрема Random Forest у поєднанні з TF-IDF векторизацією, забезпечують достатній рівень точності та інтерпретованості, необхідний для інституційного впровадження. З регуляторної точки зору ці висновки є вагомими, оскільки пропонують масштабовану методологію, придатну для реалізації вимог європейського Закону про цифрові послуги (DSA). Забезпечуючи прозорі та придатні до аудиту механізми фільтрації інформаційного «шуму» у фінансових потоках даних, такі моделі сприяють збереженню фундаментальних засад процесу ціноутворення. Це гарантує, що ринкова оцінка активів ґрунтується на базових економічних показниках, а не на штучно сформованих настроях.

Крім того, у дослідженні наголошується, що зі зростанням швидкості поширення цифрової інформації «вартість перевірки» стає ключовим обмеженням ефективності ринку. Запропонований NLP-конвеєр суттєво знижує цей бар'єр, виконуючи функцію суспільного блага та створюючи цінність як для роздрібних, так і для інституційних інвесторів. Зменшуючи вплив маніпулятивних наративів на ринкову динаміку, цей підхід сприяє формуванню більш стійкої (резильєнтної) фінансової екосистеми, здатної адаптуватися до швидких змін цифрової епохи.

У перспективі інтеграція цих моделей у системи регуляторного нагляду в режимі реального часу залишається пріоритетом для єдиного цифрового ринку Європейського Союзу. Подальші дослідження мають розширити отримані результати, зокрема шляхом аналізу міжмовних можливостей моделей на фрагментованих ринках, а також вивчення рекурсивного взаємозв'язку між автоматизованою перевіркою новин і волатильністю алгоритмічної торгівлі.

Зрештою, впровадження інтерпретованих і високоточних систем виявлення дезінформації є не лише технічним досягненням, а й необхідним наступним етапом розвитку управління сучасними фінансовими ринками.

Список використаної літератури

Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. O. (2020). Fake news detection using machine learning ensemble methods. *Complexity*, Article 8885861. https://doi.org/10.1155/2020/8885861

Alghamdi, J., Luo, S., & Lin, Y. (2024). A comprehensive survey on machine learning approaches for fake news detection. *Multimedia Tools and Applications, 83*, 51009–51067. https://doi.org/10.1007/s11042-023-17470-8

Alshuwaier, F. A., & Alsulaiman, F. A. (2025). Fake news detection using machine learning and deep learning algorithms: A comprehensive review and future perspectives. *Computers*, *14*(9), Article 394. https://doi.org/10.3390/computers14090394

Cheng, L.-C., Lu, W.-T. & Yeo, B. (2023). Predicting abnormal trading behavior from internet rumor propagation: A machine learning approach. *Financial Innovation, 9*, Article 3. https://doi.org/10.1186/s40854-022-00423-9

Du, K., Xing, F., Mao, R., & Cambria, E. (2024). Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys*, *56*(9), 1–42. https://doi.org/10.1145/3649451

D'Ulizia, A., Caschera, M. C., Ferri, F., & Grifoni, P. (2021). Fake news detection: A survey of evaluation datasets. *PeerJ Computer Science 7*, Article e518. https://doi.org/10.7717/peerj-cs.518

European Parliament & Council of the European Union. (2014). Regulation (EU) No. 596/2014 of the European Parliament and of the Council of 16 April 2014 on market abuse (market abuse regulation) and repealing Directive 2003/6/EC of the European Parliament and of the Council and Commission Directives 2003/124/EC, 2003/125/EC and 2004/72/EC and Commission Decision 2004/693/EC. *Official Journal of the European Union, L 173,* 1–142. https://eur-lex.europa.eu/eli/reg/2014/596/oj

European Parliament & Council of the European Union. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act). *Official Journal of the European Union, L 277*, 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Hu, L., Wei, S., Zhao, Z., & Wu, B. (2022). Deep learning for fake news detection: A comprehensive survey. *AI Open*, *3*, 133–155. https://doi.org/10.1016/j.aiopen.2022.09.001

kaggle.com. (n.d.). *Fake and real news dataset* [Data set]. Retrieved July 17, 2025, from https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research, 50*(2), 495–540. https://doi.org/10.1111/j.1475-679X.2012.00450.x

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10 - Ks. *The Journal of Finance*, *66*(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Lyzun, M., Desyatnyuk, O., Savelyev, Y., Kuryliak, V., Sachenko, S., Lishchynskyy, I. (2023). Architectonics of the European Currency Integration: Cluster and Gravity Modeling. *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Dortmund, Germany, 2023, pp. 661–664. https://doi.org/10.1109/IDAACS58523.2023.10348652

Lyzun, M., Lishchynskyy, I., Savelyev, Y., Kuryliak, V. and Kurylyak, Y. (2019). Modeling Evaluation of Dollarization Economic Efficiency. *International Conference on Advanced Computer Information Technologies (ACIT).* Ceske Budejovice, Czech Republic: 366–370.

Majumdar, A., & Bose, I. (2018). Detection of financial rumors using big data analytics: The case of the Bombay Stock Exchange. *Journal of Organizational Computing and Electronic Commerce*, 28(2), 79–97. https://doi.org/ 10.1080/10919392.2018.1444337

Mishra, S., Shukla, P., & Agarwal, R. (2022). Analyzing machine learning enabled fake news detection techniques for diversified datasets. *Wireless Communications and Mobile Computing*, Article 1575365. https://doi.org/10.1155/ 2022/1575365

Nghiem, H., Muric, G., Morstatter, F., & Ferrara, E. (2021, November). Detecting cryptocurrency pump-and-dump frauds using market and social signals. *Expert Systems with Applications, 182,* Article 115284. https://doi.org/ 10.1016/j.eswa.2021.115284

Rangapur, A., Wang, H., Jian, L., & Shu, K. (2025, April 28-May 2). Fin-Fact: A benchmark dataset for multimodal financial fact-checking and explanation generation. In *WWW Companion'25: Companion Proceedings of the ACM Web Conference 2025* (pp. 785–788). ACM. https://doi.org/10.1145/3701716.3715292

Selyukh, A. (2013, April 24). *Hackers send fake market-moving AP tweet on White House explosions*. Reuters. https://www.reuters.com/article/technology/hackers-send-fake-market-moving-ap-tweet-on-white-house-explosions-idUSBRE93M12Y/

Theodorakopoulos, L., Theodoropoulou, A., Tsimakis, A., & Halkiopoulos, C. (2025). Big Data-driven distributed machine learning for scalable credit card fraud detection using PySpark, XGBoost, and CatBoost. *Electronics*, *14*(9), Article 1754. https://doi.org/10.3390/electronics14091754

Отримано: 7 листопада 2025 р.
Рецензовано: 10 березня 2026 р.
Рекомендовано до друку: 20 березня 2026 р.