УДК 004.415.2
DOI https://doi.org/10.32782/2663-5941/2025.1.2/24

Олещенко Л.М.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Ільїн М.О.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ПРОГРАМНИЙ МЕТОД АНАЛІЗУ ПОТОКОВИХ ДАНИХ СОЦІАЛЬНИХ МЕРЕЖ ЗА ДОПОМОГОЮ ГІБРИДНОЇ МОДЕЛІ НЕЙРОННОЇ МЕРЕЖІ

У статті представлено інноваційний підхід до аналізу даних соціальних мереж у реальному часі, використовуючи гібридну модель нейронної мережі довгої короткочасної пам'яті (LSTM). Зосереджуючись на аналізі дописів, пов'язаних з війною в Україні, розроблений метод забезпечує виявлення подій та аналіз тенденцій у динамічному інформаційному середовищі. Запропоноване дослідження пропонує новий рівень програмної обробки текстових даних, включаючи токенізацію, вилучення ключових ознак та використання попередньо натренованих моделей для аналізу контексту.

Реалізація розробленої програмної системи базується на симуляції потокового передавання даних із набору даних "Russia-Ukraine war – Tweets Dataset (65 days)", що імітує динамічний потік дописів з платформи X (Twitter) та містить понад 41 мільйон записів. Такий підхід не лише дозволяє уникнути обмежень, пов'язаних із використанням API X, але й створює контрольоване середовище для тестування моделей. Потік даних обробляється у кілька етапів: від попереднього очищення текстів, видалення шумів (URL-адреси, емоджі, згадки) та токенізації до розпізнавання іменованих сутностей (NER), які виділяють релевантні геополітичні терміни, ключових гравців та локації. Усе це сприяє формуванню числових наборів даних, які використовуються для навчання моделі.

Запропонована гібридна модель поєднує елементи згорткових нейронних мереж (CNN) для виявлення локальних шаблонів у тексті та шарів LSTM для відстеження часових залежностей. Архітектура включає шар вбудовування (embedding layer), який використовує попередньо натреновані векторні представлення слів, шар макс-пулінгу для зменшення розмірності та повнозв'язний шар для об'єднання вивчених шаблонів. Такий підхід забезпечує точніше визначення контексту подій та тенденцій. Результати дослідження демонструють переваги запропонованої моделі перед базовою LSTM. Середня абсолютна похибка (MAE) прогнозування за допомогою гібридної моделі на 20.4% менше, ніж з використанням базової LSTM. Це підтверджує ефективність інтеграції CNN у структуру LSTM, що дозволяє виявляти складніші шаблони текстових даних. Крім того, використання механізмів уваги дозволяє гібридній LSTM адаптуватися до динамічних змін у потоці інформації. Висновки дослідження підкреслюють масштабованість та ефективність розробленого підходу. Hybrid LSTM дозволяє не лише оперативно реагувати на зміни у соціальних настроях, але й забезпечує засоби для аналізу інформаційної безпеки. Виявлення пропагандистських наративів та розуміння поведінки користувачів соціальних мереж може стати потужним інструментом у боротьбі з дезінформацією.

Подальші дослідження можуть бути спрямовані на інтеграцію графових нейронних мереж (GNN) для виявлення зв'язків між обліковими записами та аналізу поширення інформації. Такий підхід має потенціал для вдосконалення системи, забезпечуючи більш точне виявлення наративів та забезпечення оперативного моніторингу подій.

Ключові слова: *програмний метод, потокові дані, соціальні мережі, гібридна модель, довга короткочасна пам'ять (LSTM), геополітичні терміни, розпізнавання іменованих сутностей (NER), шаблони тексту, токенізація, пропагандистські наративи, дезінформація.*

Постановка проблеми. У сучасному цифровому світі, де обмін інформацією відбувається в реальному часі, соціальна мережа X.com (раніше Twitter) виступає однією з ключових платформ для швидкої передачі новин та думок. Потоки даних із соціальних мереж, зокрема, дописи (твіти) стають важливим джерелом актуальної інформації, випереджаючи традиційні засоби масової інфор-

мації. Проте значний обсяг і швидкість генерації таких даних створюють труднощі для їх обробки та аналізу. Існуючі підходи часто не справляються з динамічним характером потоків інформації, що підкреслює необхідність створення ефективних технологій та програмних систем для роботи з великими обсягами текстових даних у реальному часі.

Для вирішення цього завдання виникає потреба в інноваційних моделях, які здатні одночасно обробляти структурну складність тексту та враховувати часові залежності. Одним із перспективних рішень є використання гібридних моделей глибокого навчання, зокрема гібридної моделі довгої короткочасної пам'яті LSTM (Long Short-Term Memory). Цей підхід дозволяє виявляти закономірності в тексті та динаміці потоку даних, що відкриває нові можливості для аналізу інформаційних потоків у різноманітних контекстах.

Аналіз останніх досліджень і публікацій. Останнім часом у медіапросторі спостерігається стрімке збільшення обсягів даних, що генеруються користувачами соціальних мереж, науковці відзначають критичну роль аналітики потокових даних для аналізу трендів і виявлення інформаційних загроз. Спостерігається значне зростання кількості публікацій, присвячених аналізу даних соціальних мереж у реальному часі, що зумовлено зростанням інформаційних потоків, актуальністю геополітичних конфліктів і необхідністю протидії дезінформації. Наприклад, дослідження [1] фокусується на аналізі якості даних, що надходять з потоків повідомлень соціальної мережі Twitter. Авторами висвітлюються ключові аспекти, що впливають на корисність таких даних, зокрема, дублікатність, наявність спаму та недостатню семантичну релевантність. Представлені методи для попередньої обробки даних та оцінки їх якості можуть бути корисними для побудови більш точних алгоритмів очищення текстових потоків. Автори статті [2] розглядають методи лексичної нормалізації тексту з соціальної мережі Twitter, що є важливим етапом для покращення точності обробки текстових даних. Запропоновані підходи, зокрема використання словників, моделей послідовностей і сучасних технік глибокого навчання, допомагають краще працювати з неструктурованими, скороченими або стилістично варіативними даними. У статті [3] представлено систематичний огляд методів розпізнавання іменованих сутностей (Named Entity Recognition), а також висвітлюються класичні підходи, такі як CRF (Conditional Random Fields) та сучасні підходи на основі глибокого навчання. Особлива увага приділяється застосуванню цих методів для аналітики потокових даних із соціальних медіа.

У статті [4] представлено фреймворк для аналізу соціальних мереж SMAFED (Social Media Analysis Framework for Event Detection), який покращує семантичний аналіз даних, репрезентацію контенту та кластеризацію подій. Для цього використано інтегровану базу знань, семантичну репрезентацію та інкрементну кластеризацію на основі семантичної спорідненості. Експерименти показали, що SMAFED перевершує інші фреймворки з мінімальною втратою функції (0.15 і 0.05 на двох наборах даних) та демонструє високу точність у виявленні подій із метриками Precision (0.922), Recall (0.793) і F-Measure (0.853). SMAFED підтвердив свою ефективність як інструмент для аналізу подій у соціальних мережах.

Хоча дослідження [5] присвячено візуальним даним, його методи поєднання нейронних мереж CNN (Convolutional Neural Network) та LSTM адаптуються для текстового аналізу. У статті представлено модель, яка здатна генерувати природні мовні описи для зображень та їх окремих регіонів. Автори використовують набори даних із зображеннями та відповідними текстовими описами, щоб навчити модель встановлювати взаємозв'язок між візуальними та мовними даними. Запропонована модель базується на комбінації згорткових нейронних мереж для аналізу регіонів зображень, двонаправлених рекурентних нейронних мереж для роботи з текстами, а також структурованого об'єктиву, що з'єднує обидві модальності через мультимодальне вбудовування. На основі отриманих вирівнювань запропоновано мультимодальну рекурентну нейронну мережу, яка здатна генерувати нові описи для регіонів зображень. Експерименти на наборах даних Flickr8K, Flickr30K і MSCOCO показали, що модель досягає найкращих результатів у завданнях пошуку. Крім того, згенеровані моделлю описи суттєво перевершують результати базових моделей як для повних зображень, так і для нового набору даних із анотаціями на рівні регіонів.

У статті [6] запропоновано підхід для виявлення подій у часі за допомогою глибокого навчання та мульти-векторних вбудовувань на текстових даних із соціальних мереж. Перш за все, для попередньої обробки текстових даних використовується згорткова нейронна мережа, доповнена різними архітектурами вбудовувань слів. Потім модель виявлення подій на основі рекурентної нейронної мережі застосовується для вивчення

часових рядів і вилучення часової інформації. У статті показано, що поєднання вбудовувальних характеристик на рівні вбудовування та їх подача на згорткову нейронну мережу дозволяє досягти високої ефективності без обмеження за розміром і з додатковими вбудовуваннями порівняно зі стандартними багатоканальними підходами. Для виявлення подій використовуються рекурентні нейронні мережі, що дозволяють прогнозувати майбутні значення, а також моделі для оцінки аномалій та виявлення подій через віконний метод. У дослідженні [7] пропонується техніка машинного навчання для виявлення неправдивих новин шляхом фільтрації даних соціальних мереж, класифікації попередньо оброблених даних за допомогою алгоритму машинного навчання, оцінки розробленої системи та оцінки результатів. Основна проблематика, висвітлена в зазначених статтях, зосереджується на проблемах обробки, аналізу та використання даних соціальних мереж для виявлення подій, трендів та неправдивих новин у реальному часі. У попередніх дослідженнях наголошується на необхідності врахування високої динамічності та шумності потоків даних [1], що вимагає розробки спеціалізованих алгоритмів для забезпечення якості інформації. Значна увага приділяється методам попередньої обробки тексту, зокрема лексичній нормалізації [2] та розпізнаванню іменованих сутностей [3], які є критичними для вилучення релевантної інформації в умовах великого обсягу неструктурованих даних. Використання сучасних підходів, таких як рекурентні нейронні мережі [6] та машинне навчання [7], підкреслює необхідність побудови моделей, здатних виявляти закономірності та наративи в потоках даних у режимі реального часу. Також розглядаються питання інтеграції семантичного аналізу [4] та візуально-семантичного вирівнювання для покращення інтерпретації даних [5], що дозволяє ефективніше виявляти як події, так і дезінформацію. У статті зосереджується увага на актуальності створення автоматизованих систем, які могли б забезпечити точність і масштабованість аналізу для задоволення зростаючих потреб у моніторингу потокових даних соціальних мереж у глобальному контексті.

Постановка завдання. У рамках цього дослідження пропонується застосування гібридної моделі довгої короткочасної пам’яті для аналізу потокових даних, що імітують реальні умови програмних платформ соціальних мереж. Основна мета дослідження полягає у створенні програмного методу, який забезпечить ефективне виявлення подій, аналіз, прогнозування тенденцій з високою точністю та адаптацію до змін у динамічному інформаційному середовищі.

Виклад основного матеріалу. Розглянемо ключові особливості запропонованого програмного методу на основі гібридної довгої короткочасної пам'яті, спеціально розробленої для обробки великих обсягів даних із соціальних мереж (зокрема, з Twitter), з метою виявлення значущих подій і трендів у контексті війни в Україні. Запропоноване програмне рішення функціонує в режимі реального часу, що дозволяє оперативно аналізувати часову динаміку взаємодій користувачів у соціальних мережах. Завдяки цьому система може ідентифікувати зміни у шаблонах взаємодії, які можуть свідчити про важливі події або нові тренди, що стосуються військових, соціальних або інформаційних аспектів конфлікту.

Однією з ключових проблем при роботі з даними Twitter є залежність від API цієї платформи, яка накладає суттєві обмеження, такі як обмеження за кількістю запитів, висока вартість доступу до великих обсягів даних, а також затримки через політику доступу до API. Щоб уникнути цих обмежень, у нашій роботі було реалізовано потоковий симулятор, який дозволяє імітувати динаміку даних Twitter в реальному часі.

Для цієї мети використано набір даних "Russia-Ukraine war – Tweets Dataset (65 days)". Досліджуваний набір даних містить дописи соціальної мережі Twitter, пов’язані з соціальними та геополітичними подіями, і пропонує унікальний матеріал для тестування моделей машинного навчання, його структура та наявність контекстуального контенту дозволяють імітувати реальні сценарії потокового передавання даних (Рис. 1).

Набір даних містить повідомлення, які мають точні часові позначки, що забезпечує можливість їх хронологічного відтворення. Симулятор працює так, що дані з набору послідовно генеруються, імітуючи реальний потік нових повідомлень. Один із ключових аспектів цього підходу полягає в тому, що швидкість передачі даних у симуляторі можна регулювати залежно від наявних обчислювальних ресурсів та поточних аналітичних завдань. Наприклад, при тестуванні алгоритму в умовах з обмеженими ресурсами швидкість потоку можна зменшити, тоді як при виконанні масштабних експериментів її можна збільшити. Таким чином, цей підхід забезпечує реалістичну імітацію роботи з великими обсягами потокових даних, що дозволяє тестувати ефектив-

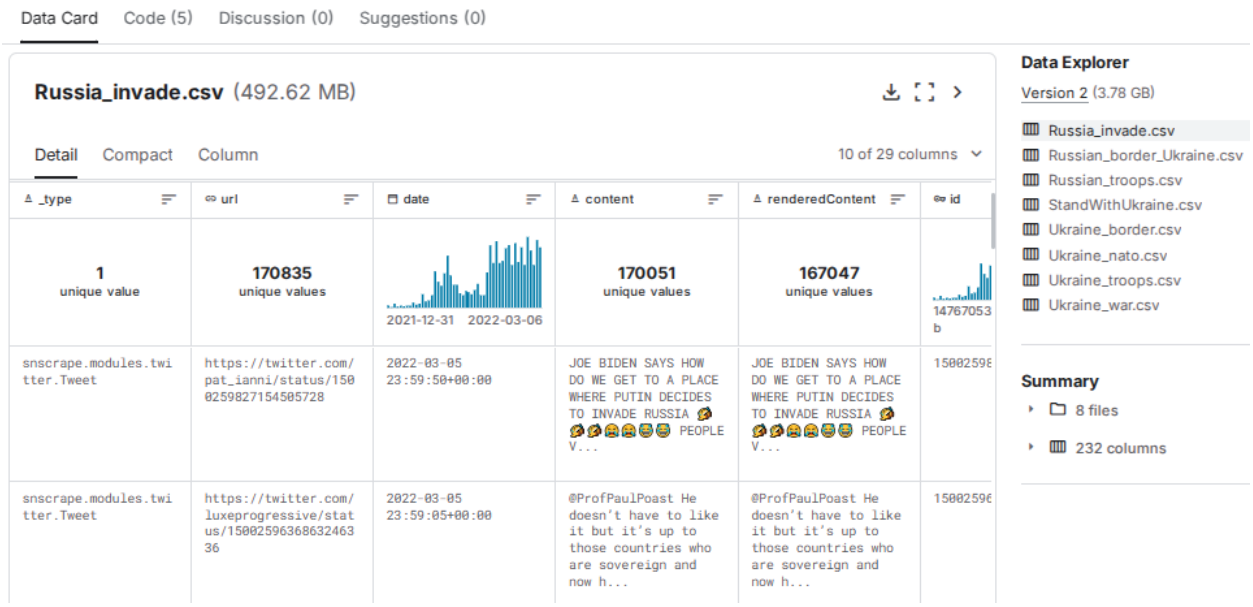

Рис. 1. Досліджуваний датасет "Russia-Ukraine war – Tweets Dataset (65 days)" [8]

```python
import pandas as pd
import time

# Завантаження набору даних
dataset = pd.read_csv('Russia_invade.csv')

# Функція симуляції потоку
def stream_simulator():
    for index, row in dataset.iterrows():
        tweet = row['content']
        timestamp = row['date']
        print(f"Потік дописів о {timestamp}: {tweet}")
        time.sleep(0.01)

# Запуск симулятора
stream_simulator()
```

Рис. 2

ність і продуктивність запропонованого рішення. Завдяки створенню контрольованого середовища для експериментів система отримує змогу адаптуватися до реальних сценаріїв, що є важливим для аналізу складних інформаційних потоків у соціальних мережах.

Програмна система симуляції потокового передавання у реальному часі побудована на основі Python та бібліотеки Pandas. В її основі лежить набір даних із файлу *Russia_invade.csv*, що містить історичні дані дописів, які імітують динаміку їх надходження в реальному часі.

Набір даних завантажується у DataFrame за допомогою функції *read_csv*.

Функція *stream_simulator()* ітерується по кожному рядку набору даних, отримуючи зміст допису (*content*) та час публікації (*date*).

Кожен допис виводиться на екран разом із міткою часу, яка вказує момент, коли його слід обробляти. Для імітації реального часу використовується функція *time.sleep(0.01)*, яка задає невеликий часовий проміжок між передаванням дописів (Рис. 2).

Як тільки дописи надходять у систему, активується базовий етап попереднього фільтрування.

На цьому етапі відбувається швидке сканування вхідних дописів з метою відбору лише релевантного контенту. Основними критеріями відбору є наявність таких елементів, як хештеги, згадки користувачів або геополітично значущі ключові слова. Це забезпечує відсіювання шуму ще до основного аналізу, підвищуючи ефективність наступних етапів.

Після того, як дописи проходять через симулятор і первинне фільтрування, вони надходять до етапу глибокої попередньої обробки, яка включає очищення тексту (видалення несуттєвих елементів, таких як URL-адреси, згадки (наприклад, @username), емодзі та зайві символи), токенізацію (розбиття тексту твіту на окремі аналітичні одиниці – слова або фрази), розпізнавання іменованих сутностей та виділення ознак (Рис. 3).

Токенізація є базовим кроком для структурування даних, необхідних для подальших алгоритмів обробки. Розпізнавання іменованих сутностей (NER) допомагає виявляти конкретні сутності, такі як назви локацій, згадки ключових осіб або важливі хештеги. NER додає контекстуальний рівень аналізу, необхідний для глибокого розуміння змісту допису. На завершення попередньої обробки система виділяє релевантні ознаки з текстів дописів. Ключовими ознаками є довжина допису в символах, частота вживання ключових слів або термінів та інші метрики, що вказують на можливу значущість допису для подальшого аналізу. Ці ознаки перетворюються на числові формати, які є ідеальними для обробки моделлю LSTM. Завдяки цьому підходу система забезпечує точне й ефективне опрацювання великих обсягів даних у реальному часі.

Для аналізу дописів та передбачення подій ми порівняли дві різні архітектури нейронних мереж: базову LSTM і вдосконалену гібридну модель LSTM. Вдосконалена модель об'єднує переваги двох підходів: згорткових нейронних мереж (CNN) для вилучення локальних ознак і мереж LSTM для вивчення часових залежностей. Це порівняння допомагає оцінити, наскільки ефективнішою є інтеграція обох підходів для обробки потоку даних із соціальних мереж.

Гібридна модель була побудована на основі унікального поєднання шарів CNN і LSTM. Кожен компонент моделі виконує чітко визначену роль, забезпечуючи точність і ефективність прогнозування.

Основним завданням *шару вбудовування (Embedding layer)* є перетворення текстових даних у векторний формат. Для цього використовуються попередньо натреновані векторні представлення слів (наприклад, Word2Vec або GloVe), які зберігають семантичний зв’язок між словами. Векторне представлення дозволяє моделі працювати з тек-

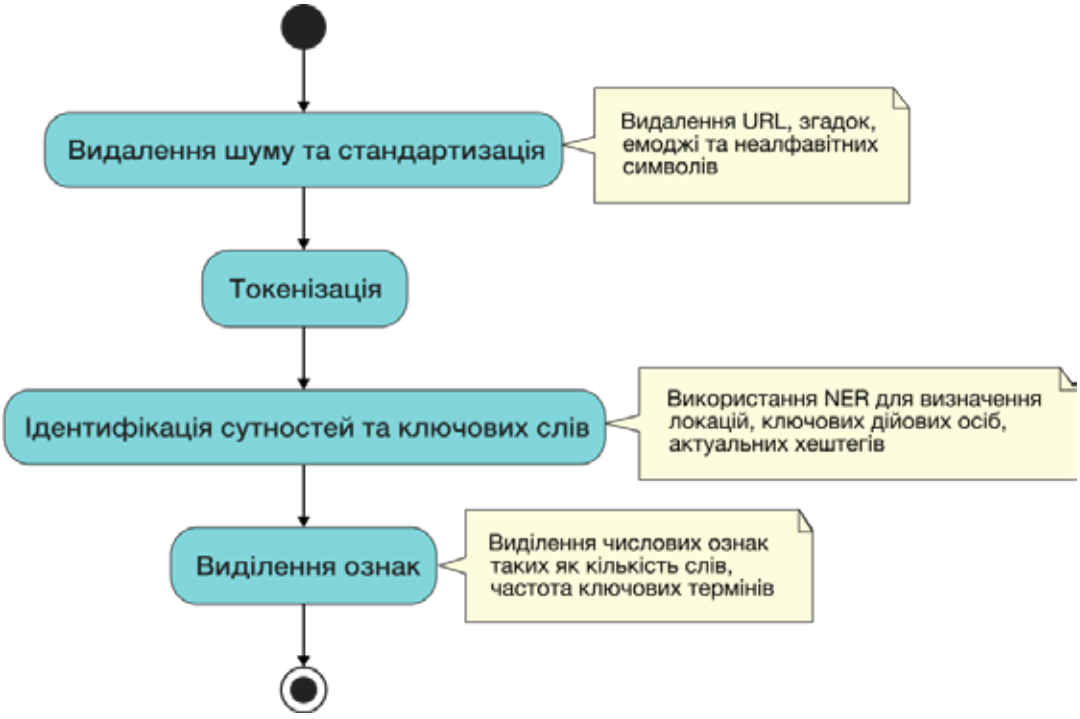

Рис. 3. Шар попередньої обробки даних

стом як із математичною сутністю, що є необхідним для наступних шарів.

Згортковий шар CNN (Convolutional layer) виконує аналіз текстових послідовностей, визначаючи локальні шаблони, наприклад, поєднання ключових слів або фраз, які часто зустрічаються разом. Він використовує кілька фільтрів, які «згорткою» проходять по текстових даних, щоб виділити локальні ознаки, які можуть бути пропущені іншими методами. Наприклад, CNN може виявити часті згадки про місця чи події в певному контексті.

Шар макс-пулінгу (Max-pooling layer) зменшує розмірність даних, залишаючи тільки найважливіші ознаки. Замість того, щоб передавати всі отримані ознаки, макс-пулінг вибирає ті, які мають найбільшу значущість. Це знижує обчислювальну складність моделі та зменшує ризик перенавчання, одночасно зберігаючи ключову інформацію.

Шар LSTM (Long Short-Term Memory layer) відповідає за аналіз часових залежностей у текстових даних. Завдяки здатності LSTM зберігати інформацію про попередні контексти, модель може враховувати послідовність дописів і визначати їхній взаємозв'язок. Наприклад, якщо кілька твітів протягом короткого часу містять схожі ключові слова, LSTM допомагає моделі розпізнати, що це може свідчити про значущу подію.

Повнозв'язний шар (Fully-connected layer) – це завершальний шар, що агрегує всі ознаки, які були виділені попередніми шарами та формує вихідний прогноз. Це може бути ймовірність певної події, наприклад, військової ескалації або нового тренду. Повнозв'язний шар забезпечує інтерпретацію всіх вхідних даних і перетворює їх на конкретний результат.

Інтеграція CNN і LSTM дозволяє моделі не лише аналізувати локальні текстові шаблони, але й враховувати їх у часовому контексті. Це особливо важливо для аналізу дописів, де послідовність публікацій та їхній контекст мають вирішальне значення для виявлення подій.

Порівняльний аналіз показав, що гібридна модель перевершує базову LSTM у точності виявлення тенденцій і подій завдяки своїй здатності враховувати як локальні, так і глобальні залежності в текстових даних (Рис. 4).

Оцінка ефективності гібридної моделі проводилась на основі метрики середньої абсолютної похибки (Mean Absolute Eirror, MAE), що є одним із ключових показників якості прогнозування в задачах аналізу даних. Результати свідчать про те, що гібридна модель LSTM демонструє значно менше значення MAE (0.2263) порівняно з базовою LSTM (0.2843). Такий показник підтверджує здатність гібридної моделі краще вловлювати складні текстові шаблони та часові залежності, характерні для соціальних медіа.

Однією з ключових переваг удосконаленої моделі LSTM є інтеграція згорткових шарів CNN, які дозволяють виявляти локалізовані закономірності у вхідних даних. CNN може розпізнати повторювані фрази, частотні шаблони ключових слів та контекстуальні зв'язки, які часто залишаються поза увагою у випадку базової LSTM. Це

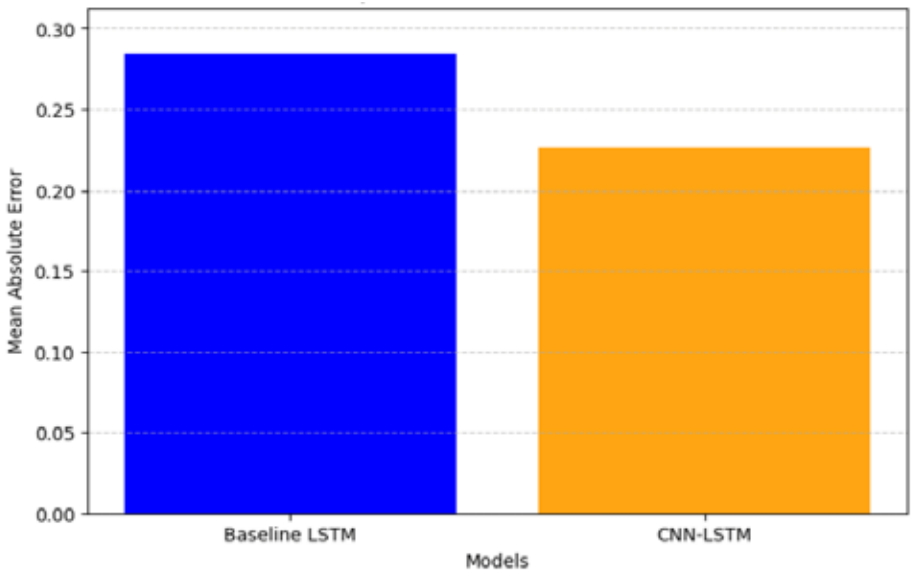

Рис. 4. Порівняння MAE між базовою та гібридною моделлю LSTM

забезпечує більш глибоке розуміння структури та динаміки даних, підвищуючи загальну точність прогнозів. Запропонований підхід сприяє розширенню можливостей цифрової медіа-аналітики та відкриває перспективи для його застосування в різних сферах, від моніторингу соціальних настроїв до аналізу інформаційної безпеки.

Висновки. Запропонована гібридна модель LSTM підтвердила свою ефективність у задачах аналітики потокових даних у реальному часі, зокрема для виявлення подій та аналізу тенденцій у соціальних мережах. Гібридна модель досягла MAE на 20.4% менше порівняно з показником базової LSTM. Це свідчить про вищу здатність моделі працювати з динамічними потоками даних, властивими платформам соціальних мереж. Використання згорткових шарів дозволяє ефективніше витягувати релевантні ознаки з текстових даних, тоді як механізми уваги покращують здатність обробляти складні часові залежності.

Гібридна модель LSTM довела свою здатність працювати у режимі реального часу навіть при обробці великих обсягів потокових даних. Це забезпечує можливість постійного моніторингу соціальних настроїв та динаміки інформаційного простору.

Подальші дослідження авторів спрямовані на вдосконалення архітектури гібридної моделі LSTM для більшої адаптації до змін у динаміці потоків повідомлень соціальних мереж та оптимізації точності моделі, зокрема, шляхом інтеграції нових методів, які дозволять краще обробляти нетипові дані, такі як меми, зображення чи відео. Це розширить функціональні можливості моделі та забезпечить її адаптацію до різноманітних форматів контенту, характерних для сучасних соціальних мереж. Окрему увагу слід приділити використанню графових нейронних мереж (GNN). Додавання GNN до архітектури моделі дозволить аналізувати зв'язки між обліковими записами користувачів, розпізнавати схеми поширення інформації, а також виявляти інформаційні атаки. Це стане важливим кроком у напрямку глибшого розуміння структур соціальних мереж і їхньої динаміки. Важливим напрямом удосконалення є розширення функціональності моделі для раннього виявлення пропаганди. Завдяки цьому можна буде ідентифікувати пропагандистські наративи та оцінювати їхній вплив на масову аудиторію. Такий підхід сприятиме більш ефективному реагуванню на дезінформаційні кампанії. Крім того, перспективним є розроблення адаптивних систем, здатних реагувати на зміни поведінки користувачів і появу нових трендів у реальному часі. Це особливо важливо в контексті забезпечення інформаційної безпеки.

Запропонований програмний метод аналізу потокових даних соціальних мереж за допомогою гібридної довгої короткочасної пам'яті демонструє не лише високі результати у своїй поточній конфігурації, але й має великий потенціал для подальшого вдосконалення та інтеграції в складніші програмні системи моніторингу інформаційного простору.

Список літератури:

1. Arolfo F., Rodriguez K.C., Vaisman A. Analyzing the Quality of Twitter Data Streams. *Information Systems Frontiers*. 2022. 24(1). P. 349-369. DOI: 10.1007/s10796-020-10072-x.

2. Bilal A. Lexical Normalisation of Twitter Data. *2015 Science and Information Conference*. 2023. DOI: 10.1109/SAI.2015.7237164.

3. Goyal A., Gupta V., Kumar M. Recent Named Entity Recognition and Classification techniques: A systematic review. 2018. *Computer Science Review*. 29:21-43. DOI: 10.1016/j.cosrev.2018.06.001.

4. Kolajo T., Daramola O., Adebiyi A.A. Real-time event detection in social media streams through semantic analysis of noisy terms. *Big Data*. 2022. 9(90). https://doi.org/10.1186/s40537-022-00642-y.

5. Karpathy A., Fei-Fei L. Deep Visual-Semantic Alignments for Generating Image Descriptions. *2015 IEEE Conference on Computer Vision and Pattern Recognition*. 2015. DOI: 10.1109/CVPR.2015.7298932.

6. Nguyen V.Q., Anh T.N., Yang H.-J. Real-time event detection using recurrent neural network in social sensors. *International Journal of Distributed Sensor Networks*. 2019.15(6). DOI:10.1177/1550147719856492.

7. Micheal Arowolo, Sanjay Misra, Roseline Ogundokun. A Machine Learning Technique for Detection of Social Media Fake News. *International Journal on Semantic Web and Information Systems*. 2023. 19(1):1-25. DOI: 10.4018/IJSWIS.326120.

8. Russia-Ukraine war – Tweets Dataset (65 days). 2022. Kaggle. URL: https://surl.li/wkdrin (дата звернення 10.11.2024).

Oleshchenko L.M., Ilin M.O. SOFTWARE METHOD FOR ANALYZING SOCIAL MEDIA STREAMING DATA USING NEURAL NETWORK HYBRID MODEL

The article presents an innovative approach to real-time social media data analysis using a hybrid Long Short-Term Memory (LSTM) neural network. Focusing on the analysis of tweets related to the war in Ukraine, the proposed method enables event detection and trend analysis in a dynamic information environment. The study offers a new level of text data processing, including tokenization, feature extraction, and the use of pre-trained models for context analysis.

The developed system implementation is based on simulating data streaming from the Kaggle dataset "Russia-Ukraine war – Tweets Dataset (65 days)," which emulates the dynamic information flow from the X (Twitter) platform and contains over 41 million records. This approach not only avoids the limitations associated with using the X API but also creates a controlled environment for model testing. The data stream is processed in several stages: from initial text cleaning, noise removal (URLs, emojis, mentions), and tokenization to named entity recognition (NER) that identifies relevant geopolitical terms, key players, and locations. This facilitates the creation of numerical datasets used for model training.

The proposed hybrid model combines elements of Convolutional Neural Networks (CNN) for detecting local patterns in the text and LSTM layers for capturing temporal dependencies. The architecture includes an embedding layer that uses pre-trained word vector representations, a max-pooling layer to reduce dimensionality, and a fully connected layer to integrate the learned patterns. This approach enables more accurate detection of event contexts and trends.

The research results demonstrate the advantages of the proposed model over the baseline LSTM. The Mean Absolute Error (MAE) of forecasting using the hybrid model is 20.4% lower compared to the baseline LSTM model. This confirms the effectiveness of integrating CNN into the LSTM structure, allowing the detection of more complex text data patterns. Additionally, the use of attention mechanisms enables the hybrid LSTM to adapt to dynamic changes in the information flow. The study's conclusions highlight the scalability and efficiency of the developed approach. The hybrid LSTM not only provides tools for responding promptly to shifts in social sentiments but also ensures capabilities for information security analysis. Identifying propaganda narratives and understanding the behavior of social media users can become a powerful tool in combating disinformation.

Future research could focus on integrating Graph Neural Networks (GNN) to detect connections between accounts and analyze information dissemination. This approach has the potential to improve the system further, ensuring more precise narrative detection and real-time event monitoring.

Key words: *software method, streaming data, social media, hybrid model, Long Short-Term Memory (LSTM), geopolitical terms, Named Entity Recognition (NER), text patterns, tokenization, propaganda narratives, disinformation.*