ISSN 2306-4412 (Print), ISSN 2306-4455 (CD-ROM), ISSN 2708-6070 (Online)

JEL Classification Code C88

[0000-0002-6307-925X] **А. І. Санжаровський,** *магістр,*
e-mail: anatsanzh@gmail.com
[0000-0001-7550-9213] **В. Я. Юрчишин,** *канд. техн. наук, доцент*
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
просп. Перемоги, 37, м. Київ, 03056, Україна

МОДИФІКОВАНИЙ МЕТОД ВИЯВЛЕННЯ ФЕЙКОВИХ НОВИН НА ОСНОВІ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ

Об'єктом дослідження є процес аналізу інформації в соціальних медіа для виявлення фейкових новин. Предметом дослідження є розроблення програмного забезпечення алгоритмічно-програмного методу для виявлення фейкових новин. Мета роботи полягає у підвищенні середньої точності процесу виявлення фейкових новин у соціальних медіа шляхом розробки та реалізації алгоритмічно-програмного методу виявлення фейкових новин на основі алгоритмів машинного навчання. Використано різноманітні методи наукових досліджень: аналізу – для з'ясування переваг та недоліків існуючих методів виявлення фейкових новин; порівняння – при виборі найбільш оптимальної мови програмування та середовища програмування для розробки програмного забезпечення для виявлення фейкових новин; метод огляду актуальної літератури з виявлення фейкових новин, включаючи академічні публікації, технічні звіти та онлайн-ресурси; метод експертної оцінки, за допомогою якого було отримано інформацію щодо ефективності різних методів виявлення фейкових новин. Завдяки використанню цих методів було отримано комплексне розуміння проблеми виявлення фейкових новин та розроблено ефективне програмне забезпечення для виявлення фейкових новин. Наукова новизна роботи полягає в тому, що було запропоновано модифікований алгоритмічно-програмний метод виявлення фейкових новин на основі алгоритмів машинного навчання, який відрізняється від наявних методів використанням ансамблю з трьох алгоритмів, результати кожного з яких використовуються для вибору компактніших спеціалізованих моделей для наступних алгоритмів, що в підсумку дозволяє пришвидшити процес виявлення фейкових новин у тексті на 30 %, порівнюючи з аналогами, а також зменшити середню хибність на 25 %. Практична цінність отриманих у роботі результатів полягає в тому, що розроблене програмне забезпечення алгоритмічно-програмного методу для виявлення фейкових новин сприятиме зменшенню поширення фейків та допомогатиме їх виявленню.

Ключові слова: *алгоритмічно-програмний метод, алгоритми машинного навчання, методи виявлення та розпізнавання фейків, BERT, LSTM, Passive-Aggressive Classifier.*

Вступ. Стрімке поширення інформації в соціальних медіа призводить до того, що фейкові новини стали однією з найбільших проблем у сучасному цифровому світі, зокрема вони використовуються як інструмент інформаційної війни [1]. Найбільш поширеними є підробка чи фабрикація новин, яка полягає в маніпулятивному спотворенні фактів, свідомій дезінформації. Більшість цих фейків створюються і поширюються в соціальних медіа без перевірки на відповідність та розраховані на людей, які сприймають інформацію емоційно без належного аналізу на відповідність. Це зумовлено тим, що емоційне сприйняття, особливо негативного характеру, унеможливлює критичний аналіз, а також сприяє подальшому поширенню фейкової інформації. Дослідження [2], опубліковане в Science у 2018 р., показало, що фейкові новини в Twitter на 70 % частіше ретвітять, а також вони швидше поширюються і охоплюють більше людей, ніж правдиві історії.

Згідно зі звітом Глобального індексу дезінформації [3] глобальні витрати на онлайн-дезінформацію можуть досягати 78 млрд дол. на рік, причому найбільше постраждає бізнес. У звіті підраховано, що підприємства втрачають у середньому 3,6 % свого річного доходу через кампанії з дезінформації.

Дослідження [4], проведене вченими з Університету Уоріка та Каліфорнійського університету в Сан-Дієго, показало, що вплив

DOI: 10.24025/2306-4412.2.2023.279984

фейкових новин під час президентських виборів у США 2016 р. міг вплинути на результат. Дослідження підрахувало, що фейкові новини, можливо, переконали 4,4 % населення переключити свій голос з Гілларі Клінтон на Дональда Трампа, чого було достатньо, щоб розгойдати вибори на користь Трампа.

Фейкові новини також мали значний вплив щодо України, особливо після анексії Криму Росією в 2014 р. Російські державні ЗМІ та інші джерела були звинувачені в поширенні дезінформації та пропаганди з метою створення неправдивого наративу про конфлікт і підриву довіри до українського уряду. Одним із найбільш показових прикладів впливу фейкових новин на війну в Україні стало збиття літака рейсу MH17 Малайзійських авіаліній у липні 2014 р. Незабаром після того, як літак був збитий, російські державні ЗМІ та інші джерела почали поширювати численні неправдиві наративи про інцидент, включаючи твердження про те, що літак збили українські військові літаки. Ці заяви пізніше були спростовані розслідуваннями, які дійшли висновку, що підтримувані Росією сепаратисти несуть відповідальність за напад.

Вплив фейкових новин на війну в Україні є значним, а неправдиві наративи сприяють поширенню дезінформації, загостренню напруженості та подальшій дестабілізації в країні.

Наразі зростає кількість досліджень, спрямованих на виявлення такого типу інформації та вироблення запобіжників щодо її подальшого поширення. Однак переважна більшість таких досліджень стикається з низкою проблем під час ідентифікації фейкових новин у соціальних мережах. До найбільш поширених із них можна віднести: складність збирання та визначення «вручну» належності до фейкової новини; неефективність блокування в месенджерах (Facebook, WhatsApp, Twitter та інших) можливості поширення фейкової інформації та скарг на неї; поширення дезінформації верифікованими новинними агентствами або друзями чи родичами.

Автоматизовані методи виявлення фейкових новин можуть допомогти відфільтрувати неправдиву інформацію, що сприятиме зменшенню впливу фейкових новин на окремих осіб, організації та уряди. Ці методи також сприяють забезпеченню відповідними засобами для позначення та видалення фейкових новин із платформ соціальних мереж, пошукових систем та інших онлайн-каналів. Серед найбільш цікавих досліджень можна виокремити наступні:

1. «Виявлення фейкових новин у соціальних мережах: перспектива інтелектуального аналізу даних» (2017) [5] за авторством Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, Huan Liu, в якому розглядається можливість виявлення фейкових новин у соціальних мережах за допомогою поєднання алгоритмів машинного навчання та методів інженерії функцій. Автори взяли за основу аналізу набір даних твітів, пов'язаних з президентськими виборами в США 2016 р., та з високою точністю виявили фейкові новини.

2. «Боротьба з фейковими новинами: опитування щодо методів ідентифікації та пом'якшення наслідків» (2019) за участю Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, Yan Liu [6], в якому здійснено огляд різних методів виявлення та пом'якшення фейкових новин, включаючи перевірку фактів, алгоритми машинного навчання та аналіз соціальних мереж.

3. «Лінгвістичні особливості фейкових новин та їх наслідки для автоматизованого виявлення» Хорна та Адалі (2017) [7], в якому аналізуються лінгвістичні особливості фейкових новин та визначається кілька характеристик, які можна використовувати для розрізнення справжніх та фейкових новин.

Наведені дослідження являють собою низку найбільш поширених підходів до виявлення фейкових новин, включаючи алгоритми машинного навчання, лінгвістичний аналіз та аналіз соціальних мереж. Вони підкреслюють важливість міждисциплінарних досліджень у цій сфері та припускають, що комбінація підходів може бути найбільш ефективною у виявленні фейкових новин.

Саме тому наше дослідження присвячено розробленню програмного забезпечення (ПЗ) модифікованого методу для виявлення фейкових новин, яке сприятиме зменшенню поширення фейків і допомогатиме їх виявляти. Об'єктом дослідження є процес аналізу інформації в соціальних медіа для виявлення фейкових новин. Предметом дослідження є розроблення ПЗ модифікованого методу для виявлення фейкових новин. Під час проведення дослідження були використані різноманітні методи наукових досліджень: аналізу – для з'ясування переваг та недоліків існуючих методів виявлення фейкових новин;

DOI: 10.24025/2306-4412.2.2023.279984

порівняння – при виборі найбільш оптимальної мови програмування та середовища програмування для розробки ПЗ для виявлення фейкових новин. Також було застосовано метод огляду наявної актуальної літератури з виявлення фейкових новин, включаючи академічні публікації, технічні звіти та онлайн-ресурси. Метод експертної оцінки було використано для пошуку інформації від експертів у галузі виявлення фейкових новин. Завдяки використанню цих методів було отримано комплексне розуміння проблеми виявлення фейкових новин та розроблено ефективне ПЗ для виявлення фейкових новин.

Мета та задачі дослідження. Мета дослідження полягає в підвищенні середньої точності процесу виявлення фейкових новин у соціальних медіа шляхом розробки та реалізації алгоритмічно-програмного методу виявлення фейкових новин на основі алгоритмів машинного навчання. Для досягнення мети виконані такі завдання: досліджено існуючі методи виявлення фейкових новин; виявлено переваги та недоліки використання наявних методів; розроблено програмний модуль для тестування ефективності алгоритмів наявних методів та сформованих наукових гіпотез; запропоновано новий модифікований метод виявлення фейкових новин зі збільшеною точністю, порівнюючи з існуючими методами; створено ПЗ, яке реалізує запропонований метод; експериментально досліджено ефективність розробленого методу шляхом оцінки якості роботи ПЗ за встановленими критеріями.

Виклад основного матеріалу

Аналіз наявних аналогів. Існує кілька програмних методів виявлення фейкових новин, кожен зі своїми перевагами та обмеженнями, а саме:

Обробка природної мови (НЛП) [8] – це техніка, яка використовує алгоритми машинного навчання для аналізу тексту новинних статей і виявлення закономірностей та особливостей, які вказують на ймовірність того, що стаття буде підробленою. Цей метод може бути ефективним, але він вимагає великої кількості навчальних даних і може погано працювати з нетекстовим вмістом, таким як зображення та відео.

Метод аналізу джерел передбачає аналіз достовірності джерела новинної статті. Це можна зробити, дослідивши репутацію новинного видання чи журналіста, точність їхніх попередніх репортажів та будь-які упередження, які вони можуть мати. Хоча цей метод може бути корисним, він може бути неефективним для виявлення більш тонких форм упередженості або маніпуляцій.

Перевірка фактів включає перевірку точності тверджень, зроблених у новинній статті, на відповідність базі даних перевірених фактів. Це може бути ефективним для виявлення відвертої брехні, але може бути не в змозі виявити більш тонкі форми дезінформації або пропаганди.

Метод аналізу соціальних мереж передбачає аналіз соціальної мережі, в якій поширюється новина. Досліджуючи користувачів, які діляться статтею, та їхні зв'язки з іншими користувачами, можна виявити моделі скоординованої пропаганди чи маніпуляції. Однак цей метод може бути неефективним для виявлення більш поодиноких випадків фейкових новин.

Загалом кожен із цих методів має свої сильні та слабкі сторони, а поєднання кількох методів може бути найбільш ефективним для виявлення фейкових новин.

Двосторонні рекурентні нейронні мережі LSTM (Long Short-Time Memory) (RNN) успішно використовуються для виявлення фейкових новин [9; 10; 11]. Цей підхід передбачає навчання нейронної мережі на великому наборі даних новинних статей, де кожна стаття позначається або як «справжня», або як «фейкова». Нейромережа вчиться виявляти закономірності в тексті, які пов'язані або з реальними, або з фейковими новинами, а потім може бути використана для класифікації нових статей як справжніх або фейкових.

Перевага використання двосторонньої LSTM-RNN полягає в тому, що вона може враховувати як прямий, так і зворотний контекст кожного слова в тексті, дозволяючи йому фіксувати більш складні відносини між словами та фразами. Крім того, LSTM здатні «запам'ятовувати» інформацію протягом більш тривалих періодів часу, що може бути корисно для виявлення тонких шаблонів у тексті.

Хоча двосторонні LSTM-RNN показали перспективність у виявленні фейкових новин, вони не позбавлені обмежень. Наприклад, їм може бути важко виявити більш тонкі форми маніпуляції або пропаганди, які не покладаються на очевидні фактичні неточності. Крім

DOI: 10.24025/2306-4412.2.2023.279984

того, точність моделі сильно залежить від якості та розміру навчального набору даних і може змінюватися залежно від конкретних джерел новин і тем, що аналізуються.

Двосторонні представлення кодувальників на основі перетворювачів (BERT) (Bidirectional Encoder Representations from Transformers) [12; 13] – це попередньо навчена мовна модель, розроблена Google, яка ефективна в широкому спектрі завдань обробки природної мови, включаючи виявлення фейкових новин.

BERT показав свою ефективність у виявленні фейкових новин, оскільки він може вловлювати тонкі нюанси мови, які часто використовуються у фейкових новинах для маніпулювання або введення читачів в оману. Однак, як і в інших моделях машинного навчання, його точність сильно залежить від якості та розміру навчального набору даних, і йому може бути важко виявити більш складні форми пропаганди або дезінформації. Крім того, обчислювальні витрати на точне налаштування BERT можуть бути досить високими, що ускладнює масштабування моделі до більших наборів даних або додатків у реальному часі.

Пасивно-агресивний класифікатор (Passive-Aggressive Classifier) (PA) [14] – алгоритм є одним із популярних лінійних класифікаторів, який використовувався для виявлення фейкових новин. Перевага використання алгоритму PA для виявлення фейкових новин полягає в тому, що він відносно простий і обчислювально ефективний, що робить його добре придатним для додатків у режимі реального часу. Крім того, він може обробляти високовимірні простори функцій, що корисно для обробки текстових даних.

Однак алгоритм PA може мати труднощі з виявленням більш складних зв'язків між словами та фразами. Тому він часто використовується в поєднанні з іншими моделями машинного навчання або методами для підвищення його продуктивності.

Отже, Двостороння LSTM, BERT та Пасивно-агресивний класифікатор є найбільш поширеними програмними методами виявлення фейкових новин. Кожен із цих способів має свої переваги і недоліки. Вибір методу, який буде використовуватися для виявлення фейкових новин, залежить від конкретних вимог і обмежень поставленого завдання. Двосторонні LSTM і BERT можуть бути більш придатними для завдань, які вимагають високої точності і можуть вмістити більші набори даних, тоді як пасивно-агресивні класифікатори можуть бути більш придатними для завдань з обмеженими маркованими даними або додатків у реальному часі.

Постановка завдання. Проаналізувавши наявні аналоги методу для виявлення фейкових новин, можна виділити наступні функціональні вимоги до ПЗ для виявлення фейкових новин і запобігання їх поширенню:

1. ПЗ повинно мати можливість збирати дані з різних джерел, таких як платформи соціальних мереж, новинні веб-сайти та інші онлайн-джерела.
2. ПЗ повинно мати можливість очищати, фільтрувати та попередньо обробляти зібрані дані, щоб переконатися, що вони точні та придатні для аналізу.
3. ПЗ повинно мати можливість витягувати відповідні функції з попередньо оброблених даних, такі як використовувана мова, тон автора та настрій тексту.
4. ПЗ повинно мати можливість створювати маркований навчальний набір даних, який включає як фейкові, так і реальні новини для розробки моделей машинного навчання.
5. ПЗ має використовувати алгоритми машинного навчання, такі як Bi-directional LSTM, BERT або Passive-Aggressive Classifier, для аналізу попередньо оброблених даних і виявлення фейкових новин.
6. ПЗ повинно мати можливість оцінювати точність і продуктивність моделей машинного навчання на тестовому наборі даних.
7. ПЗ повинно мати можливість виявляти фейкові новини в режимі реального часу та надавати сповіщення користувачам.
8. ПЗ повинно мати простий у використанні та інтуїтивно зрозумілий користувальницький інтерфейс, який дозволяє користувачам отримувати доступ до результатів аналізу та вживати відповідних дій.
9. ПЗ повинно бути налаштовуваним, щоб дозволити користувачам налаштовувати параметри моделей машинного навчання та джерела даних відповідно до своїх конкретних потреб.
10. ПЗ має забезпечувати конфіденційність і безпеку даних користувачів і бути стійким до атак, таких як ворожі атаки, які намагаються обійти алгоритм виявлення фейкових новин.

DOI: 10.24025/2306-4412.2.2023.279984

Обгрунтування мови та середовища програмування. Для розроблення ПЗ методу виявлення фейкових новин було проаналізовано наявні мови програмування, їх сильні та слабкі сторони і вибрано найбільш оптимальну мову програмування для досягнення належної роботи ПЗ відповідно до поставлених завдань. Було встановлено:

Java – широко використовувана мова програмування для розробки додатків корпоративного рівня, включаючи програми машинного навчання. Наявність бібліотек і фреймворків, таких як Apache Mahout і Weka, робить Java хорошим вибором для ПЗ для виявлення підроблених новин на основі машинного навчання.

R є популярною мовою програмування для статистичних обчислень і широко використовується при розробці додатків машинного навчання. Наявність бібліотек і фреймворків, таких як Caret і MLR, робить R хорошим вибором для ПЗ для виявлення підроблених новин на основі машинного навчання.

C++ є високопродуктивною мовою програмування і зазвичай використовується для розробки інтенсивних обчислювальних додатків, включаючи програми машинного навчання. Наявність бібліотек і фреймворків, таких як TensorFlow і Caffe, робить C++ хорошим вибором для ПЗ для виявлення підроблених новин на основі машинного навчання.

JavaScript є популярною мовою програмування для веб-розробки і все частіше використовується для розробки додатків машинного навчання. Наявність бібліотек та фреймворків, таких як TensorFlow.js та Brain.js, робить JavaScript хорошим вибором для веб-програмного забезпечення для виявлення підроблених новин.

Python є популярною та потужною мовою для розробки ПЗ для виявлення фейкових новин, особливо для підходів, що ґрунтуються на машинному навчанні. Хоча він має деякі обмеження, його сильні сторони, включаючи простоту використання, гнучкість і багатий набір бібліотек та фреймворків, роблять його гарним вибором для розробки складних і точних алгоритмів виявлення фейкових новин. На підставі проведеного аналізу мови програмування Python було прийнято рішення щодо використання саме цієї мови для написання коду ПЗ для модифікованого методу виявлення фейкових новин.

Під час роботи над дослідженням було проаналізовано декілька середовищ розробки з метою вибору найбільш оптимального для розробки ПЗ методу виявлення фейкових новин. До найбільш поширених можна віднести:

Anaconda – це програмне середовище, яке має різноманітні інструменти і бібліотеки для машинного навчання та аналізу даних. Anaconda підтримує декілька мов програмування, таких як Python, R та Java.

Visual Studio Code – це легкий редактор відкритого коду, який підтримує кілька мов програмування. Він включає такі функції, як налагодження, контроль версій і розширення для машинного навчання та аналізу даних.

PyCharm – це потужна IDE Python, яка надає такі функції, як налагодження, завершення коду та контроль версій. Він також включає інтеграцію з бібліотеками машинного навчання, такими як TensorFlow і Keras.

RStudio – інтегроване середовище розробки (IDE) з відкритим вихідним кодом для мови програмування R. Він включає такі функції, як завершення коду, налагодження та візуалізація інструментів для аналізу даних.

Eclipse – популярна IDE з відкритим вихідним кодом, яка підтримує кілька мов програмування, включаючи Java, C++ та Python. Вона включає такі функції, як налагодження, завершення коду та контроль версій.

При виборі середовища розробки ПЗ для розробки ПЗ для виявлення фейкових новин нами було враховано такі фактори, як використовувана мова програмування, доступність бібліотек машинного навчання та аналізу даних, а також простота використання для розробників. Насамперед, вибір середовища розробки ПЗ залежав від конкретних потреб і вимог проєкту. Саме тому для розробки модифікованого методу виявлення фейкових новин нами було обрано середовище Python IDLE. Це зумовлено тим, що він поставляється з такими функціями, як оболонка Python, яка є інтерактивним інтерпретатором. Він має широкі функції, такі як автодоповнення, підсвічування синтаксису, розумний відступ і базовий інтегрований відладчик.

Результати досліджень, отримані в роботі: розроблено ПЗ, в якому використовується алгоритмічно-програмний метод для виявлення фейкових новин на основі алгоритмів машинного навчання – це застосунок, що містить користувацький інтерфейс. Застосу-

DOI: 10.24025/2306-4412.2.2023.279984

нок для виявлення фейкових новин – це програмне рішення, призначене для боротьби з поширенням фейкової інформації за допомогою алгоритмічно-програмних методів та алгоритмів машинного навчання. Основною метою цього додатка є швидке та точне виявлення фейкових новин, запобігання їх поширенню та пом'якшення потенційного негативного впливу на суспільне життя. Спеціально орієнтоване на користувачів соціальних мереж, це ПЗ забезпечує зручний інтерфейс, щоб надати людям можливість приймати обґрунтовані рішення щодо автентичності новин, з якими вони стикаються.

Ключові особливості:

1. Алгоритми машинного навчання: ПЗ включає алгоритми машинного навчання, які були навчені на великих наборах даних перевірених новинних статей та зразків фейкових новин. Ці алгоритми аналізують різні лінгвістичні та контекстуальні особливості новинних статей, щоб відрізнити справжню та фейкову інформацію.

2. Аналіз тексту та обробка природної мови: додаток використовує методи аналізу тексту та обробку природної мови (NLP) для вилучення значущої інформації з новинних статей. Він аналізує такі фактори, як мовні моделі, настрої, стиль письма та лексичні сигнали, щоб оцінити достовірність вмісту.

3. Джерела даних та агрегація: ПЗ інтегрується з кількома надійними джерелами даних, включаючи авторитетні новинні агентства, організації з перевірки фактів та звіти, створені користувачами. Він агрегує дані з цих джерел для підвищення точності процесу виявлення.

4. Виявлення в режимі реального часу: додаток працює в режимі реального часу, дозволяючи користувачам оперативно перевіряти справжність новинних статей, перш ніж ділитися ними на платформах соціальних мереж. Він аналізує надану статтю та генерує оцінку довіри або мітку, що вказує на ймовірність того, що вміст є фальшивим.

5. Інтерфейс користувача: ПЗ має інтуїтивно зрозумілий та зручний інтерфейс, доступний на персональних компютерах або мобільних пристроях. Інтерфейс забезпечує простий і зрозумілий досвід, дозволяючи користувачам аналізувати статті та переглядати результати виявлення.

Загалом розроблений нами застосунок для виявлення фейкових новин використовує модифікований метод та алгоритми машинного навчання, щоб забезпечити зручне рішення для виявлення та запобігання поширенню фейкової інформації. Розроблення графічного інтерфейсу, як правило, передбачає використання певного набору віджетів – елементів керування, які є зручними як для користувача, оскільки використовуються в більшості застосунків і мають приблизно однаковий вигляд і поведінку в рамках однієї платформи, так і для розробника, оскільки вони вже реалізовані в численних бібліотеках, які реалізують самі елементи керування, а їх властивості та поведінка швидко та просто налаштовуються відповідно до потреб проєкту. Для реалізації інтерфейсу в проєкті використана бібліотека tkinter, що є вбудованою бібліотекою Python.

Дизайн інтерфейсу відповідає функціональним вимогам проєкту. Враховуючи вимоги, інтерфейс містить секцію зміни налаштувань, що керують поведінкою ПЗ. Головна функція цього ПЗ – це повідомлення користувача про наявність фейкової інформації в тестованому тексті та, відповідно, виявлення фейкових новин.

Для вирішення проблеми поширення фейкових новин нами було розроблено модифікований метод виявлення фейкових новин на основі алгоритмів машинного навчання, який за вхідні дані приймає текст змісту новин, що потребують перевірки на наявність фейкової інформації.

Загалом цей метод складається з трьох етапів:

1. Класифікація вхідних даних за мовою тексту.

2. Кластеризація вхідних даних за вмістом.

3. Виявлення наявності фейкової інформації у вхідних даних.

На ***першому етапі*** методу після отримання вхідних даних відбувається визначення мови, до якої належить текст у вхідних даних. З цією метою використовується алгоритм, який навчений працювати з певним сталим наперед заданим набором мов. Доки використовується конкретна мова, в цьому алгоритмі «літерами» визначаються лише символи її алфавіту.

Під час навчання алгоритму:

1. Як вхідні дані використовується набір текстів, які написані різними мовами.

2. Для кожної мови відбуваються такі дії:

DOI: 10.24025/2306-4412.2.2023.279984

2.1. Створюється пустий словник з назвою «словник частот».

2.2. Проходячи кожний текст для певної мови від початку до кінця для кожної послідовної пари літер у тексті:

2.2.1. Якщо в «словнику частот» є запис, в якому ключ має таке саме значення, як певна послідовна пара літер, тоді значення лічильника запису з цим ключем збільшується на 1.

2.2.2. Якщо в «словнику частот» немає запису, в якому ключ має таке саме значення, як певна послідовна пара літер, тоді в «словник частот» додається запис з ключем, рівним цій послідовній парі літер, та обнуленим лічильником як значення.

3. Після побудови «словників частот» для мов, використовуючи всі вхідні дані, «словники частот» перетворюються на «словники ймовірностей» шляхом отримання суми всіх значень для кожного «словника частот» і ділення всіх значень елементів кожного «словника частот» на відповідну суму.

4. Для кожного отриманого «словника ймовірностей» береться натуральний логарифм від кожного значення ймовірності і присвоюється на місце того значення ймовірності, таким чином утворюючи «словники логарифмів ймовірностей».

Під час використання алгоритму:

1. Алгоритм отримує на вхід один текст;

2. Для кожного «словника логарифмів ймовірностей» виконуються такі кроки:

2.1. Створюються змінні «лічильник пар літер» і «лічильник логарифмів» та призначається їм значення 0.

2.2. Для кожної пари послідовних літер у тексті, що наявна в цьому «словнику логарифмів ймовірностей», збільшується значення «лічильника логарифмів» на відповідне значення запису в «словнику логарифмів ймовірностей» та збільшується значення «лічильника пар літер» на 1.

2.3. Ділиться значення «лічильника логарифмів» на значення «лічильника пар літер» та зберігається результат як «остаточне значення» для цього «словника логарифмів ймовірностей».

3. Мова, словник якої має найбільше отримане значення, є передбаченою мовою тексту.

Після цього виконується ***другий етап*** методу, в якому відбувається визначення групи, до якої належить текст у вхідних даних за змістом. При визначенні групи, до якої належить текст у вхідних даних, використовується наступний алгоритм, який використовує отримані результати першого етапу. Перед виконанням цього алгоритму спочатку відбувається обробка і зміна тексту вхідних даних:

1. Видаляються з вхідних даних тексту всі символи, які не є літерами визначеної мови тексту, та замінюються на пробіли.

2. Переносяться всі літери тексту до нижнього реєстру.

3. Текст розбивається пробілами на список слів, при цьому прибираючи пусті слова.

4. Вилучаються «стоп-слова» зі списку слів.

5. Проводиться стемінг кожного слова.

Під час навчання алгоритму:

1. При навчанні певний алгоритм приймає за вхідні дані лише тексти тією самою мовою для конкретної моделі (до попередньої обробки тексту вхідних даних).

2. Створює список «унікальних слів» і робить його пустим.

3. Для кожного навчального тексту спочатку після попередньої обробки тексту відбувається додавання всіх слів зі списку слів цього тексту до списку «унікальних слів», яких там немає.

4. Після цього для кожного навчального тексту:

4.1. Використовуючи список «унікальних слів», визначаються значення TF-IDF для кожного «унікального слова», використовуючи список слів цього тексту як документ.

4.2. Створюється вектор зі значеннь TF-IDF для даного тексту.

5. Викоритовуючи алгоритм k-means++, кластеризуються тексти за їхніми векторами.

Під час використання алгоритму:

1. Для кожного слова зі списку «унікальних слів» визначаються значення TF-IDF в цьому списку слів і, таким чином, створюється вектор значень.

2. Кластеризується цей вектор, використовуючи алгоритм k-means++ [15].

Насамкінець виконується ***третій етап*** методу, в якому виявляється наявність або відсутність фейкової інформації у вхідних даних. На цьому етапі використовується оброблений текст вхідних даних з другого етапу за допомогою алгоритму BERT для класифікації вхідних даних на наявність фейкової інформації. Для кожної унікальної комбінації

DOI: 10.24025/2306-4412.2.2023.279984

мови в першому етапі та кластеру в другому етапі створюється своя модель алгоритму BERT при навчанні. Після цього при використанні цього методу вибирається відповідна модель залежно від результатів першого та другого етапів. Таким чином можна зробити моделі алгоритму BERT меншого розміру та покращити точність їх результатів. У цьому алгоритмі застосовується модель «bert-large-uncased».

Цей алгоритм має чотири види використання:

1. Задача класифікації пари речень.
2. Задача класифікації одного речення.
3. Задача отримання відповіді на запитання.
4. Задача тегування одного речення.

В алгоритмічно-програмному методі виявлення фейкових новин алгоритм BERT застосовується таким чином:

1. Спочатку вхідні дані (оброблені на другому етапі) будуть перетворені так:

1.1. Обробляються всі слова у вхідних даних, використовуючи спеціальний вбудований токенізатор BERT.

1.2. Якщо довжина списку слів більша, ніж певне значення (128), то розбивається список слів на набір списків слів таким чином, щоб кожний список у наборі був меншим за вищезазначене значення.

1.3. Для кожного списку слів у наборі:

1.3.1. Додається спеціальне слово «[CLS]» перед початком списку слів.

1.3.2. Додається спеціальне слово «[SEP]» після кінця списку слів.

1.3.3. Якщо цей список слів досі менший ніж вищезазначене значення, то додається спеціальне слово «[PAD]», доки довжина цього списку не буде дорівнювати вищезазначеному значенню.

1.3.4. Створюються «маски уваги» для всіх слів «[PAD]» у списку слів та додаються ці «маски уваги» до відповідного списку слів.

1.3.5. Перетворюються всі слова у списку слів на відповідні індекси, використовуючи спеціальний вбудований токенізатор BERT.

2. Для кожного списку індексів у наборі:

2.1. Використовуючи 2-й вид роботи алгоритму BERT і цю модель, обробляється список індексів та відповідна «маска уваги» і отримується результат (значення від 0 до 1).

2.2. Результат свідчить про наявність фейкової інформації, якщо значення більше чи дорівнює 0.5, або про її відсутність – якщо менше 0.5.

3. Фінальним результатом у попередньому кроці буде те значення, яке трапилося більше разів, або наявність фейкової інформації, якщо кількість обох значень є однаковою.

3.1. Якщо відбувається навчання моделі: використовуючи фінальний результат, порівнюють його з очікуваним і при нерівності оптимізують модель, використовуючи оптимізатор Adam.

Алгоритм адаптивної оцінки моменту (Adam) [16] як алгоритм оптимізації використано замість класичного градієнтного спуску для ітераційного оновлення ваг мережі на основі навчальних даних. Це зумовлено його головною перевагою над подібними алгоритмами, а саме: можливість адаптивно змінювати «швидкість навчання» моделі під час його використання, а також швидкість виконання та мала ресурсоємкість.

Отже, під час першого етапу цього методу відбувається навчання на великих наборах текстів відповідними мовами. На другому і третьому етапах методу навчання відбувається на тому самому наборі даних, що містять фейкові новини.

Кожна мова має власний набір кластерів k-means++ після навчання. Кожний кластер має власну модель BERT після навчання. Через це набори кластерів та відповідні моделі BERT для різних мов використовують набори даних, що містять фейкові новини відповідною мовою. Навчання кожної частини відбувається відповідно до описаної послідовності дій.

Обговорення результатів. Відповідно до сформованих функціональних вимог до ПЗ було розроблено застосунок, що містить програму, яку можна запустити, використовуючи програму Python. Ця програма при запуску відкриває вікно інтерфейсу застосунку. Користувач може використати це вікно, щоб ввести текст новини та перевірити її на наявність фейкової інформації. Розроблене ПЗ для виявлення фейкових новин відповідає нефункціональним вимогам і є сумісним з варіаціями операційних систем Windows та Linux. Також було розроблено зручний для користувача інтерфейс.

DOI: 10.24025/2306-4412.2.2023.279984

У процесі тестування ПЗ методу виявлення фейкових новин на основі алгоритмів машинного навчання було перевірено таке:

1. Функціональна працездатність можливості зміни налаштувань роботи застосунку.
2. Функціональна працездатність застосунку обробляти поданий текст і визначати наявність фейкових новин у ньому з достатньою точністю.
3. Відсоток правильних позитивних реакцій на наявність фейкових новин в аналізованому тексті.
4. Коректна робота мобільного застосунку на пристроях з операційними системами Windows та Linux.
5. Відповідність дизайну вимогам технічного завдання.

Тестування було виконано методом Black Box Testing. Зокрема було протестовано:

1. Виявлення швидкості та точності визначення фейкових новин.
2. Користувацький інтерфейс (зміни налаштувань, які застосовуються).

Було також проведено тестування швидкодії ПЗ на наборах даних із соціальних мереж. Результати тестування подано в таблицях 1 та 2. Після тестування було виявлено, що середня швидкодія застосунку при кількості 200 слів дорівнює 5 с.

Таблиця 1. **Результати тестування застосунку для виявлення фейкових новин за тривалістю роботи, секунд**

	Тривалість роботи (с.)			
Кількість слів у тексті	Цей метод	BERT	PAC	BiLSTM
260	5.3	5.8	4.3	5.5
500	7.1	8.7	6.3	8.2
1000	13.5	17.2	11.8	15.8
5000	20.7	28.5	18.4	25.4

Таблиця 2. **Результати тестування застосунку для виявлення фейкових новин за точністю виявлення фейків, %**

	Точність виявлення фейків (%) (TP)			
Кількість слів у тексті	Цей метод	BERT	PAC	BiLSTM
260	97.4	97.1	96.4	95.4
500	98.8	97.6	96.9	96.1
1000	99.3	98.8	98.2	97.8
5000	99.5	99.1	98.9	98.4

За результатами перевірки та оцінки розробленого методу і застосунку для визначення фейкових новин на основі алгоритмів машинного навчання за допомогою цього тестування було встановлено відсутність відхилень поведінки розробленого ПЗ від вимог, зазначених у технічному завданні (рисунки 1 та 2).

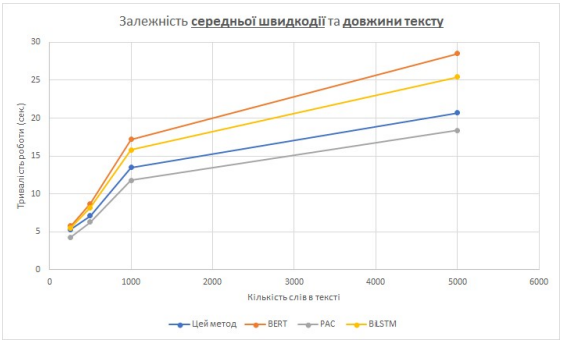

Рисунок 1. Графік залежності тривалості роботи методів від довжини тексту

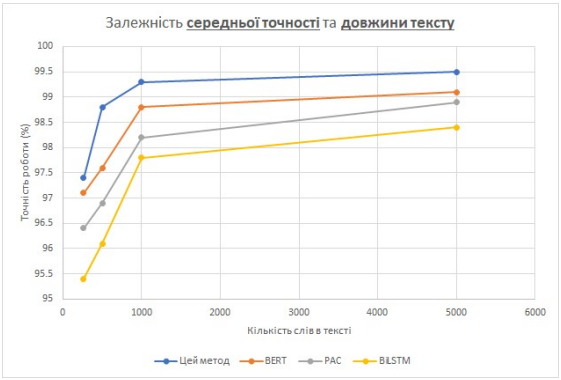

Рисунок 2. Графік залежності точності роботи методів від довжини тексту

Висновки. Розроблено програмне забезпечення модифікованого методу виявлення фейкових новин, який дозволяє уникнути небезпек поширення фейків та запобігти їм завдяки тому, що він аналізує текст новин та повідомлень, визначає і повідомляє користувача про наявність фейкової інформації.

Експериментальним шляхом підібрано оптимальний варіант принципу дії методу та відповідного програмного забезпечення, а саме застосування стандартних алгоритмів нейронних мереж машинного навчання, які забезпечують коректне функціонування методу та ПЗ. Під час роботи застосунок аналізує наданий текст на наявність фейкової інформації. Після доопрацювання результати цього

DOI: 10.24025/2306-4412.2.2023.279984

проєкту можна застосувати для створення програмних систем промислового рівня для розв'язання прикладних проблем.

Наукова новизна дослідження полягає в тому, що було запропоновано модифікований метод виявлення фейкових новин на основі алгоритмів машинного навчання, який відрізняється від наявних методів використанням ансамблю з трьох алгоритмів, результати кожного з яких використовуються для вибору компактніших спеціалізованих моделей для наступних алгоритмів, що в підсумку дозволяє пришвидшити процес виявлення фейкових новин у тексті на 30 %, порівняно з аналогами, а також зменшити середню хибність на 25 %.

Практична цінність дослідження полягає в тому, що за допомогою модифікованого методу та розробленого на його основі програмного забезпечення застосунку для виявлення фейкових новин можна успішно їх виявляти в соціальних мережах, а також запобігати їх поширенню. Особливо велике значення це має в умовах зростаючої деструктивної ролі мережевої війни, одним із інструментів якої є поширення фейків та дезінформації.

Перспективи подальших досліджень. У перспективі планується продовжити вдосконалювати процес визначення мови тексту та кластеризації тексту новин, а також дослідити використання інших методів для попереднього розподілу вхідних даних для прискорення та вдосконалення розробленого методу.

Детальний аналіз характеристик програмного застосунку дозволив дійти висновку, що покращення та розвиток програмного застосунку можливі у кількох різних напрямах, серед яких:

Досягнення *оптимізації та покращення ефективності* шляхом використання: більш оптимізованого алгоритму обробки тексту, інших алгоритмів для попереднього розподілу вхідних даних для прискорення та уточнення загального алгоритму тощо.

Підтримка інших операційних систем: поширення ПЗ для інших середовищ виконання, наприклад для операційної системи IOS, може значно розширити застосовність застосунку, але потребує застосування інших технологій, покращення продуктивності та зміни внутрішньої будови ПЗ для його підтримки на інших операційних системах.

Рефакторинг ПЗ можна проводити майже за будь-яких умов, оскільки він не відображається на поведінці ПЗ, а покращує стан кодової бази і для впровадження потребує тільки проведення локального тестування компонентів, які були змінені, щоб упевнитись, що поведінка ПЗ не змінилася.

Список використаних джерел

[1] А. Санжаровський, та В. Юрчишин, "Алгоритмічно-програмний метод для виявлення фейкових новин на основі алгоритмів машинного навчання", на *П'ятнадцятій наук. конф. магістрантів та аспірантів Прикладна математика та комп'ютинг (ПМК-2022)*, Київ, 16-18 листоп. 2022, с. 499-504.

[2] Study: On Twitter, false news travels faster than true stories. [Online]. Available: https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308. Accessed on: Jan. 20, 2023.

[3] Disinformation risk assessment: The online news market in the United States. [Online]. Available: https://www.disinformationindex.org/country-studies/2022-12-16-disinformation-risk-assessment-the-online-news-market-in-the-united-states/. Accessed on: Jan. 20, 2023.

[4] H. Alcott, and M. Gentzkow, "Social media and fake news in the 2016 election", *Journal of Economic Perspectives*, vol. 31 (2), pp. 211-236. doi: 10.1257/jep.31.2.211.

[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective", *ACM SIGKDD Explorations Newsletter*, vol. 19, iss. 101, pp. 22-36, Sept. 2017. doi: 10.1145/3137597.3137600.

[6] K. Sharma, F. Qian, H. Jiang, and N. Ruchansky, "Combating fake news: A survey on identification and mitigation techniques", *ACM Transactions on Intelligent Systems and Technology,* vol. 10 (3), pp. 1-42, Apr. 2019. [Online]. Available: https://www.researchgate.net/publication/332434399_Combating_Fake_News_A_Survey_on_Identification_and_Mitigation_Techniques. Accessed on: Jan. 20, 2023. doi: 10.1145/3305260.

DOI: 10.24025/2306-4412.2.2023.279984

[7] B. D. Horne, and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news", *ArXiv* abs/1703.09398, 2017. [Online]. Available: https://www.semanticscholar.org/paper/This-Just-In%3A-Fake-News-Packs-a-Lot-in-Title%2C-Uses-Horne-Adali/f8366afaf58bbb9db151a1168bb6f14b618955b4. Accessed on: Jan. 20, 2023.

[8] D. Rothman, "Transformers for natural language processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more"; Birmingham, UK: Packt Publishing Ltd. Birmingham Mumbai, 2021.

[9] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional LSTM-recurrent neural network", *Procedia Comput. Sci.,* vol. 165, pp. 74-82, 2019. [Online]. Available: https://doi.org/10.1016/j.procs.2020.01.072. Accessed on: Jan. 20, 2023.

[10] S. M. Padnekar, G. S. Kumar, and P. Deepak, "Bilstm-autoencoder architecture for stance prediction", in *Proc. 2020 Int. Conf. on Data Science and Engineering (ICDSE)*, Kochi, India, pp. 1-5, Dec. 3-5, 2020.

[11] E. Amer, K.-S. Kwak, and S. El-Sappagh, "Context-based fake news detection model relying on deep learning models". *Electronics*, vol. 11 (8), p. 1255, 2022. [Online]. Available: https://doi.org/10.3390/electronics11081255. Accessed on: Jan. 20, 2023

[12] A. Malakhov, A. Patruno, and S. Bocconi, "Fake news classification with BERT". [Online]. Available: http://ceur-ws.org/Vol-2882/paper38.pdf. Accessed on: Jan. 20, 2023.

[13] D. Jacob, C. Ming-Wei, L. Kenton, and T. Kristina, "BERT: Pre-training of deep bi-directional transformers for language understanding". [Online]. Available: https://arxiv.org/pdf/1810.04805.pdf. Accessed on: Jan. 20, 2023.

[14] Fake News Detection Using Passive-Aggressive Classifier. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-7345-3_13. Accessed on: Jan. 20, 2023.

[15] D. Arthur, and S. Vassilvitskii, "k-means++: The advantages of careful seeding". [Online]. Available: https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf. Accessed on: Jan. 20, 2023.

[16] D. P. Kingma, and J. L. Ba, "Adam: A method for stochastic optimization". [Online]. Available: https://arxiv.org/abs/1412.6980. Accessed on: Jan. 20, 2023.

References

[1] A. Sanzharovsky, and V. Yurchyshyn, "Algorithmic-software method for detecting fake news based on machine learning algorithms", on *Fifteenth Sci. Conf. of Undergraduates and Graduate Students Applied Mathematics and Computing (PMK-2022)*, Kyiv, Nov. 16-18, pp. 499-504, 2022 [in Ukrainian].

[2] Study: On Twitter, false news travels faster than true stories. [Online]. Available: https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308. Accessed on: Jan. 20, 2023.

[3] Disinformation risk assessment: The online news market in the United States. [Online]. Available: https://www.disinformationindex.org/country-studies/2022-12-16-disinformation-risk-assessment-the-online-news-market-in-the-united-states/. Accessed on: Jan. 20, 2023.

[4] H. Alcott, and M. Gentzkow, "Social media and fake news in the 2016 election", *Journal of Economic Perspectives*, vol. 31 (2), pp. 211-236. doi: 10.1257/jep.31.2.211.

[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective", *ACM SIGKDD Explorations Newsletter*, vol. 19, iss. 101, pp. 22-36, Sept. 2017. doi: 10.1145/3137597.3137600.

[6] K. Sharma, F. Qian, H. Jiang, and N. Ruchansky, "Combating fake news: A survey on identification and mitigation techniques", *ACM Transactions on Intelligent Systems and Technology,* vol. 10 (3), pp. 1-42, Apr. 2019. [Online]. Available: https://www.researchgate.net/publication/332434399_Combating_Fake_News_A_Survey_on_Identification_and_Mitigation_

DOI: 10.24025/2306-4412.2.2023.279984

Techniques. Accessed on: Jan. 20, 2023. doi: 10.1145/3305260.

[7] B. D. Horne, and S. Adali, "This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news", *ArXiv* abs/1703.09398, 2017. [Online]. Available: https://www.semanticscholar.org/paper/This-Just-In%3A-Fake-News-Packs-a-Lot-in-Title%2C-Uses-Horne-Adali/f8366afaf58bbb9db151a1168bb6f14b618955b4. Accessed on: Jan. 20, 2023.

[8] D. Rothman, "Transformers for natural language processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more"; Birmingham, UK: Packt Publishing Ltd. Birmingham Mumbai, 2021.

[9] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional LSTM-recurrent neural network", *Procedia Comput. Sci.,* vol. 165, pp. 74-82, 2019. [Online]. Available: https://doi.org/10.1016/j.procs.2020.01.072. Accessed on: Jan. 20, 2023.

[10] S. M. Padnekar, G. S. Kumar, and P. Deepak, "Bilstm-autoencoder architecture for stance prediction", in *Proc. 2020 Int. Conf. on Data Science and Engineering (ICDSE)*, Kochi, India, pp. 1-5, Dec. 3-5, 2020.

[11] E. Amer, K.-S. Kwak, and S. El-Sappagh, "Context-based fake news detection model relying on deep learning models". *Electronics*, vol. 11 (8), p. 1255, 2022. [Online]. Available: https://doi.org/10.3390/electronics11081255. Accessed on: Jan. 20, 2023

[12] A. Malakhov, A. Patruno, and S. Bocconi, "Fake news classification with BERT". [Online]. Available: http://ceur-ws.org/Vol-2882/paper38.pdf. Accessed on: Jan. 20, 2023.

[13] D. Jacob, C. Ming-Wei, L. Kenton, and T. Kristina, "BERT: Pre-training of deep bi-directional transformers for language understanding". [Online]. Available: https://arxiv.org/pdf/1810.04805.pdf. Accessed on: Jan. 20, 2023.

[14] Fake News Detection Using Passive-Aggressive Classifier. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-15-7345-3_13. Accessed on: Jan. 20, 2023.

[15] D. Arthur, and S. Vassilvitskii, "k-means++: The advantages of careful seeding". [Online]. Available: https://theory.stanford.edu/~sergei/papers/kMeansPP-soda.pdf. Accessed on: Jan. 20, 2023.

[16] D. P. Kingma, and J. L. Ba, "Adam: A method for stochastic optimization". [Online]. Available: https://arxiv.org/abs/1412.6980. Accessed on: Jan. 20, 2023.

A. I. Sanzharovskyi, *Master,*
e-mail: anatsanzh@gmail.com
V. Ya. Yurchyshyn, *Cand. Tech. Sc., Associate Professor*
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
Peremohy ave., 37, Kyiv, 03056, Ukraine

MODIFIED METHOD FOR DETECTING FAKE NEWS BASED ON MACHINE LEARNING ALGORITHMS

The object of research is the process of analyzing information in social media to identify fake news. The subject of research is the development of software of algorithmic-software method for detecting fake news. The aim of the work is to increase the average accuracy of the process of detecting fake news in social media by developing and implementing an algorithmic-software method for detecting fake news based on machine learning algorithms. Various methods of scientific research: analysis – to find out the advantages and disadvantages of existing methods for detecting fake news; comparison – when choosing the most optimal programming language and programming environment

DOI: 10.24025/2306-4412.2.2023.279984

for developing software to detect fake news; a method of reviewing existing literature to detect fake news, including academic publications, technical reports, and online resources; peer review method, which obtained information on the effectiveness of various methods for detecting fake news have been used. Through the use of these methods, a comprehensive understanding of the problem of detecting fake news has been obtained and effective software for detecting fake news has been developed. The scientific novelty of the work lies in the fact that a modified algorithmic-software method for detecting fake news based on machine learning algorithms has been proposed, which differs from the existing methods by using an ensemble of three algorithms, the results of each of which are used to select more compact specialized models for subsequent algorithms, which ultimately allows to speed up the process of detecting fake news in the text by 30% compared to analogs, and reduce the average falsehood by 25%. The practical value of the results obtained in the work lies in the fact that the developed software of the algorithmic-software method for detecting fake news will help to reduce the spread of fakes and detect them.

Keywords: *algorithmic-software method, machine learning algorithms, methods of fake detection and recognition, BERT, LSTM, Passive-Aggressive Classifier.*

Стаття надійшла 23.04.2023
Прийнято 10.05.2023

DOI: 10.24025/2306-4412.2.2023.279984