КІБЕРБЕЗПЕКА: освіта, наука, техніка № 2 (30), 2025
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
Спеціальний випуск ISSN 2663 – 4023
Special issue
DOI 10.28925/2663-4023.2025.30.977
УДК 004.89

Чирун Любомир Вікторович
кандидат технічних наук, доцент,
доцент кафедри інформаційних систем та мереж
Національний університет «Львівська Політехніка», м.Львів, Україна
ORCID: 0000-0002-9448-1751
lyubomyr.v.chyrun@lpnu.ua

Назаркевич Андрій Ярославович
студент кафедри інформаційних систем та мереж
Національний університет «Львівська Політехніка», м.Львів, Україна
ORCID: 0009-0007-2078-8447
andrii.nazarkevych.ri.2023@lpnu.ua

ВИЯВЛЕННЯ НАЙПОШИРЮВАНІШИХ ФЕЙКОВИХ НОВИН ПОБУДОВОЮ КЛАСТЕРІВ ЛЕЙДЕНА

Анотація. Сьогодні значним викликом є автоматичне розпізнавання фейкових новин, а також виявлення груп повідомлень, які поширюють недостовірний або маніпулятивний контент. Одним із ефективних інструментів для розв'язання такої задачі є метод кластеризації Лейдена. Алгоритм Лейдена належить до методів виділення спільнот у графах і дає змогу об'єднувати новини у групи за принципом їхньої внутрішньої схожості. Для аналізу новин попередньо формується набір характеристик кожного повідомлення: його текстове наповнення, семантичні ознаки, джерело поширення, час публікації та інші додаткові параметри. На основі цих даних створюється граф, у якому вершини відповідають окремим новинам, а ребра описують ступінь їхньої подібності. Метод Лейдена виконує послідовне об'єднання таких новин у кластери, забезпечуючи високу якість поділу завдяки покроковому вдосконаленню структури груп. Застосування кластерів Лейдена дає змогу виявляти скупчення повідомлень, які мають ознаки фейковості або координованого поширення. Як правило, такі групи характеризуються повторюваними фразами, однаковими текстовими шаблонами, прискореною динамікою розповсюдження та великою кількістю перехресних посилань між неавторитетними джерелами. Аналіз цих структур дозволяє виявляти цілеспрямовані інформаційні кампанії, мережі ботів та масові вкиди неправдивого контенту. Таким чином, метод кластерів Лейдена є ефективним підходом для узагальнення та групування новинних повідомлень, що значно підвищує точність розпізнавання фейкових новин. Його застосування у системах моніторингу медіапростору забезпечує більш глибокий аналіз інформаційних потоків і сприяє своєчасному виявленню дезінформації.

Ключові слова: фейк, глибоке навчання, захищені дані

ВСТУП

Очікується, що у 2025 році методи розпізнавання тексту розвиватимуться завдяки штучному інтелекту та машинному навчанню, зокрема глибокому навчанню та нейронним мережам. Ці технології добре зарекомендували себе в розпізнаванні зображень та мовлення, обробці природної мови. До 2028 року ринок глибокого навчання досягне 93,8 мільярда доларів США [1], завдяки впровадженню цих передових методів.

Достовірної інформації бракує, і тоді породжується багато пліток, як стверджує BBC (2022). Як зазначає Пейт (2021) [2], ЗМІ повинні поширювати інформацію на основі

правдивості, компетентності, релевантності та динамізму. Також міністр інформації Нігерії Лай Мохаммед зазначає [3], що фейкові новини становлять особливу небезпеку.

Аналіз моделей виявлення фейкових новин на основі токенів свідчить про те, що вони мають погану узагальнюваність. Дослідження Мохаммеда [4] показує, що хоча глибокі нейронні мережі та вбудовування, такі як Glove та BERT, демонструють високу точність в межах одного набору даних (приблизно 99%), їхня продуктивність значно падає (до 30%) при тестуванні на різних наборах даних. Подібні результати були отримані [5], які зазначили, що моделі, навчені на політичних новинах з одного джерела, не змогли точно класифікувати новини з інших джерел. Використовуючи інструмент Local Interpretable Model-agnostic Explanations (LIME) [6], вони також виявили, що токенізовані репрезентації вразливі до тематичних упереджень. Це підкреслює необхідність розробки більш надійних методологій та різноманітних наборів даних.

Фейкові новини стали глобальною проблемою. Останнім часом у Твіттері поширюються численні фейкові зображення, які нібито зображують кризові події. Однак багато з них не мають жодного зв'язку з реальними подіями в цих країнах [7]. Це доводить, що дезінформація поширюється не лише в текстовій формі, а й через візуальний контент, який покликаний зробити його більш переконливим. Дослідники Пеннікук та Ренд [8] зазначають, що анонімність у цифровому просторі сприяє неконтрольованому поширенню фейків, створюючи сприятливі умови для їх ескалації. Крім того, неправдива інформація та маніпуляції, що заполонили соціальні мережі, становлять загрозу стабільності суспільства та можуть провокувати конфлікти [9]. Такі інформаційні викиди не лише дезінформують громадськість, але й перешкоджають розвитку країн.

Постановка проблеми. Таким чином, наукова задача полягає у виявленні та уникнення розповсюдження фейкових новин, неправдивої інформації, оскільки бурхливий розвиток соціальних мереж спричинив створення та поширення маніпулятивних повідомлень.

Аналіз останніх досліджень і публікацій

Термін «фейкові новини» став широко вживатися з часів президентської кампанії Дональда Трампа 2016 року [10]. Виявлення неправдивої інформації до її широкого поширення є ключовим для підтримки довіри громадськості до ЗМІ та запобігання маніпуляціям громадською думкою [11]. У демократичних країнах дезінформація може спотворювати виборчі процеси та підривати довіру до державних установ. Крім того, поширення неправдивих даних у сферах охорони здоров'я, надзвичайних ситуацій та науки може загрожувати громадській безпеці [12]. У соціальному плані такі процеси сприяють загостренню соціальної поляризації та конфліктів [13]. Таким чином, розробка ефективних методів боротьби з фейковими новинами є критично важливою для забезпечення прозорості комунікацій та підтримки соціальної стабільності.

Фахівці з інформатики активно працюють над розробкою методів виявлення фейкових новин за допомогою алгоритмів обробки природної мови (NLP) та машинного навчання (ML) [13]. Хоча ці методи продемонстрували високу ефективність у внутрішньому тестуванні, їхня здатність узагальнювати дані залишається сумнівною, особливо коли вони базуються на токенізованих представленнях, таких як Bag-of-Words [14], Term Frequency – Inverse Document Frequency (TF-IDF) [15] та BERT. Крім того, у літературі часто використовуються набори даних, позначені за джерелом, а не за допомогою ручної перевірки фактів, що може призвести до систематичних упереджень. Незважаючи на значний потенціал моделей великих мов (LLM) [16], їхня ефективність у

контексті виявлення фейкових новин все ще недостатньо вивчена. Тому ця робота зосереджена на проблемі узагальнення моделей, побудованих на токенах та LLM.

Мета статті є виявлення швидкопоширюваних повідомлень на основі кластеризації та застосуванні метода Лейдена.

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Аналіз методів виявлення фейкових новин

Головною метою цього дослідження є оцінка того, чи можуть альтернативні підходи, такі як аналіз стилістичних ознак, покращити узагальнюваність моделей виявлення фейкових новин. Оскільки докази ефективності LLM у цій галузі обмежені, LLaMa також тестується як контрольна модель. Унікальною особливістю дослідження є використання набору даних URL-адрес Facebook, який був вручну перевірений незалежними фактчекерами, що робить його більш надійним еталоном порівняно з іншими відкритими наборами даних.

Існуючі методи боротьби з фейковими новинами

Поширення фейкових новин створює серйозні глобальні ризики, впливаючи на демократичні процеси та поширюючи дезінформацію [17], тому їх виявлення залишається важливим дослідницьким завданням. Існуючі підходи, засновані на NLP та навчанні з учителем, демонструють значну ефективність при тестуванні на внутрішніх даних, але їх узагальнюваність обмежена через їх залежність від наборів даних, що містять систематичні упередження. Розглянуто можливість удосконалення моделей шляхом використання альтернативних ознак, зокрема стилістичних ознак (лексичних, синтаксичних та семантичних) та атрибутів соціальної монетизації, таких як рекламний контент, зовнішні посилання та соціальні взаємодії. Використовуючи набір даних NELA 2020-21 [18] для навчання та вручну перевірений набір даних URL-адрес Facebook для зовнішнього тестування, виявлено, що традиційні методи на основі токенів мають значні обмеження. У свою чергу, поєднання стилістичних ознак та соціальних факторів дозволяє створювати більш узагальнюючі моделі, які можуть ефективніше працювати в реальних умовах. Аналіз важливості ознак шляхом тестування підтверджує, що запропоновані підходи сприяють зменшенню системних упереджень та підвищенню ефективності моделей виявлення фейкових новин. Один із семантичних підходів до виявлення фейкових новин, що підкреслює їх часте конструювання на спотворених фактах, полягає в додаванні заспокійливого контенту для підвищення правдоподібності.

Значна увага приділяється вивченню емоцій у новинних текстах. Вивчається ступінь впливу таких новин на суспільство. Значна увага цьому питанню приділяється в [19], де емоції у вигляді страху, щастя, гніву, смутку та здивування виявляються на основі технологій машинного навчання. Українська мова складна, в ній іменники та прикметники збираються в спеціальний корпус, а словникові групи створюються на основі методів TF-IDF. Потім такі групи класифікують настрої в новинних текстах на основі методів машинного навчання, включаючи MLP, SVM, випадковий ліс та інші. Використовуються стандартні метрики F1, точність та матриця плутанини. На етапі навчання використовується словникова група разом з подвійною нормалізацією, що дозволило досягти точності 0,99.

Ще однією стратегією посилення довіри до такого контенту є часті посилання авторів фейкових новин на відомі історичні або сучасні авторитети.

Крім того, такий контент часто має емоційне забарвлення, що посилює його вплив на аудиторію та підвищує рівень маніпуляції.

На етапі попередньої валідації запропонований підхід тестується на обмеженій вибірці. Однак ці дані не можуть бути безпосередньо застосовані до реальних сценаріїв. У цьому контексті дослідження аналізує зв'язок між емоційним впливом та виявленням фейкових новин. Запропонований підхід включає інтерпретований класифікатор тексту на основі глибокого навчання, TC-CNN (Text Class-activation-mappings Convolutional Neural Network) [20], який здатний оцінювати емоційне забарвлення, рівень фальсифікації та упередженість.

У нещодавніх публікаціях застосовувалися різні методи вилучення ознак, включаючи Count Vectorizer [21], Bag of Words [22], Global Vectors for Word Representation (GloVe) [23], Word to Vector (Word2Vec) [24] та Term Frequency-Inverse Document Frequency (TF-IDF) [25]. Крім того, були використані методи вибору ознак для пошуку інформації, аналіз головних компонентів (PCA) та частота документів. Цей комплексний підхід підвищує здатність моделі правильно ідентифікувати та аналізувати важливі ознаки, що дозволяє проводити точніше та ефективніше виявлення фейкових новин. Результати дослідження підтверджують важливість використання багатогранних методів для значного підвищення точності та надійності моделі. Модель адаптована до різноманітних наборів даних, що робить її корисною для реальних застосувань. Було проведено серію тестів різних варіантів для визначення оптимальної конфігурації ансамблю. Експериментальні оцінки трьох різних наборів даних про фейкові новини підтверджують високу ефективність моделі.

Рис.1. Виявлення фейкових новин, які пов'язані з технологіями машинного навчання

Для пояснення результатів моделі в дослідженні [26] використовується метод картування активації класів (CAM) у поєднанні з попередньо навченими векторними представленнями слів FastText. Підхід CAM визначає внесок кожного слова до певного класу, виділяючи найважливіші елементи. Модель була навчена на 13 000 текстах з категорії «фейк», зібраних з набору даних фейкових новин Kaggle, а також на такій самій кількості автентичних новин, випадково вибраних з набору даних Signal News Media.

Моделі та методи виявлення неавтентичної поведінки користувачів чату

Дезінформаційні втручання здебільшого зосереджувалися на публічних платформах соціальних мереж, таких як Facebook та Twitter/X, незважаючи на емпіричні дані, які свідчать про поширення дезінформації на приватних платформах обміну повідомленнями, таких як WhatsApp, по всьому світу. Дослідження в цій галузі були обмежені труднощами у вивченні приватного та зашифрованого вмісту повідомлень та впровадженні втручань у ці канали. У приватних повідомленнях часто втрачаються підказки щодо джерела, мети та часу поширення інформації, що ускладнює для користувачів розпізнавання неправдивого контенту. Приватні канали зв'язку використовуються переважно для підтримки тісних зв'язків, таких як з родиною, друзями, колегами та місцевими громадами. Особливо важко оскаржити або виправити

дезінформацію в невеликих групах приватних повідомлень з соціальних та технологічних причин. Ці приватні групові чати часто передбачають тісні зв'язки, де соціальні норми є високими.

Дезінформаційні втручання здебільшого зосереджувалися на публічних платформах соціальних мереж, таких як Facebook та Twitter/X. Однак існують емпіричні дані того, що дезінформація також поширюється на приватних платформах обміну повідомленнями, таких як WhatsApp, по всьому світу. Дослідження в цій галузі обмежені через труднощі, пов'язані з вивченням приватного та зашифрованого контенту, а також складність впровадження втручань на таких платформах. Під час обміну особистими повідомленнями втрачається важлива інформація про джерело, мету та час поширення інформації, що ускладнює її перевірку. Приватні канали зв'язку часто використовуються для підтримки особистих стосунків, таких як з родиною, друзями, колегами чи місцевими громадами. Виявлення або спростування дезінформації в невеликих приватних групах набагато складніше через соціальні та технологічні фактори. Ці групи часто складаються з людей з тісними зв'язками, де соціальні норми можуть сприяти поширенню неправдивої інформації.

Характеристики приватних групових чатів

Програми обміну повідомленнями з функціями групового чату, такі як WhatsApp, WeChat, Telegram, Discord, Microsoft Teams та Slack, широко використовуються для різних цілей: командної співпраці, спілкування з родиною та друзями, організації соціальних заходів, взаємодії між школою та родиною, а також для ділових контактів.

Більшість користувачів приєднуються до невеликих закритих груп, створених близькими людьми, зазвичай не більше 10 учасників. Деякі також беруть участь у спеціалізованих групах за інтересами — хобі, ігри, фінанси — або у спільнотах підтримки, таких як чати для матерів. Головною особливістю приватних групових чатів є близькість між учасниками, яка часто слугує основою їх створення.

Платформи миттєвих повідомлень, як-от WhatsApp та LINE, дозволяють взаємодіяти через номери телефонів або унікальні ідентифікатори, і такі цифрові групи фактично відображають та зміцнюють реальні соціальні зв’язки: кола друзів, родини або сусідів. Міцні зв’язки забезпечують безпечне середовище для обговорення конфліктних тем завдяки довірі та спільній близькості. Однак близькість може також спричиняти міжособистісні конфлікти, особливо коли повідомлення залишаються без відповіді або непрочитаними.

Приватні чати зазвичай характеризуються швидким та неформальним стилем спілкування. Короткі, прямі повідомлення можуть призводити до різних інтерпретацій, що іноді викликає непорозуміння. Швидкість і неформальність обговорень також впливають на теми розмов: деякі користувачі вважають ці простори менш ефективними для прийняття рішень, але безпечнішими для політичних дискусій.

Розмір групи відіграє ключову роль у ефективності спілкування. У великих групах текстове спілкування стає менш зручним для обговорень у реальному часі, а обсяг повідомлень може ускладнювати участь кожного учасника [26].

Враховуючи різноманітність групових чатів, важливо досліджувати, як фактори міцності зв’язків, розміру групи та мети спілкування впливають на ефективність взаємодії та результати комунікації.

МЕТОДИКА ДОСЛІДЖЕННЯ

Граф може представляти будь-яку складну систему: соціальну мережу, набір текстів, послідовність взаємодій або інфраструктуру. Вершини у графі позначають об'єкти, а ребра — зв'язки між ними. Поняття спільноти або кластера означає групу вершин, які між собою більш щільно пов'язані, ніж з іншими елементами графа. Такі групи відображають приховані закономірності та структури, які не завжди можна виявити без спеціальних алгоритмів.

Метод Лейдена є високоефективним алгоритмом для виявлення спільнот у мережах. Лейден вирізняється тим, що поєднує швидкість, характерну для своїх попередників, із набагато вищою стабільністю та здатністю формувати внутрішньо узгоджені кластери. Саме тому цей метод сьогодні вважається одним із найточніших підходів до поділу графів на спільноти.

Для оцінки якості розбиття графа на кластери використовується модульність Q, яка показує, наскільки щільно вершини пов'язані всередині кластерів порівняно з випадковим графом. Математично модульність визначається як:

$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{i,j} - \frac{k_i k_j}{2m}\right)\ \delta(c_i c_j) \tag{1}$$

де $A_{i,j}$ — вага ребра між вершинами i та j, k_i, k_j — ступені вершин, m — сумарна вага всіх ребер графа, а $\delta(c_i c_j)$ дорівнює 1, якщо обидві вершини належать одному кластеру, і 0 — в іншому випадку. Ця формула дозволяє оцінити, наскільки ефективно об'єкти згруповані.

Коли алгоритм розглядає переміщення вершини з одного кластеру в інший, використовується формула зміни модульності:

$$\Delta Q = \frac{1}{2m}\left[A_{i,j} - \frac{k_i \sum_{j\in C} k_j}{2m}\right] \tag{2}$$

де C — потенційний кластер для переміщення вершини. Якщо $\Delta Q > 0$, то переміщення покращує якість кластеризації, і вершину переносимо. Таким чином алгоритм поступово формує локально оптимальні групи.

Кожен кластер має внутрішню вагу, яка визначається сумою всіх ребер всередині нього:

$$e_c = \frac{1}{2}\sum_{i,j\in C} A_{i,j} \tag{3}$$

а сумарна вага всіх ребер, що інцидентні до кластеру, обчислюється як

$$k_c = \sum_{i\in C}\ k_i \tag{4}$$

Модульність окремого кластеру можна виразити через ці величини:

$$Q_c = \frac{e_c}{m} - \left(\frac{k_c}{2m}\right)^2 \tag{5}$$

Метод Лейдена вводить також перевірку внутрішньої зв'язності: кластер вважається цілісним, якщо для будь-якої його частини не існує повної ізоляції, тобто для будь-якого підкластеру $S \in C$ виконується умова:

$$\sum_{i \in S, j \in C \setminus S} A_{i,j} > 0 \quad (6)$$

Після перевірки та уточнення внутрішньої структури кластери агрегуються в новий укрупнений граф. Вершини нового графа відповідають кластерам попередньої ітерації, а ребра формуються за правилом:

$$A'_{CD} = \sum_{i \in C} \sum_{j \in D} A_{i,j} \quad (7)$$

де C і D — кластери попереднього графа. Надалі алгоритм повторює ті самі кроки на укрупненому графі, що забезпечує поступове покращення кластеризації.

Остаточна якість розбиття визначається як сума модульностей всіх кластерів:

$$Qtotal = \sum_{C \in P} Q_C \quad (8)$$

де P — множина всіх кластерів. Ця величина служить критерієм ефективності алгоритму та дозволяє порівнювати різні способи групування об'єктів.

Проблему ідентифікації фейкових новин зазвичай формулюють як задачу класифікації, метою якої є оцінка достовірності контенту. У процесі аналізу виділяють три ключові аспекти: намір створення, механізми поширення та сприйняття аудиторією. Перший аспект досліджує мотиви публікації дезінформації, другий – шляхи її розповсюдження, третій – реакцію користувачів.

Для кластеризації графів застосовується метод Лейдена, що забезпечує багаторівневу оптимізацію та поступове об'єднання вузлів у кластери. Це дозволяє виявляти групи джерел новин та гарантує внутрішню зв'язність спільнот. Алгоритм швидко обробляє великі обсяги даних і ефективно працює з різними типами медіаконтенту: текстом, графікою та відео.

Експериментальні результати показали, що метод Лейдена дозволяє виділити логічно пов'язані кластери новин, виявити фейкові повідомлення та ідентифікувати найбільш швидко поширювані новини. Це забезпечує своєчасне реагування на дезінформацію, підвищує точність аналізу медіапотоків та сприяє зміцненню інформаційної безпеки.

Таким чином, метод Лейдена є ефективним інструментом для системного виявлення та аналізу фейкових новин, що робить його важливим засобом у дослідженні медіаконтенту та протидії дезінформації.

Послідовніст роботи алгоритму включає такі етапи:

1. Локальне переміщення вузлів: кожна новина оцінюється на основі зв'язків з іншими новинами та джерелами.

2. Агрегація спільнот: новини, що мають спільні джерела або теми, об'єднуються в кластери.

3. Повторення процесу: у новій, агрегованій мережі алгоритм повторює ці кроки, що дозволяє виявити стабільні спільноти дезінформації.

Застосування алгоритму Лейдена до аналізу новин дозволяє:

- Визначити групи джерел, які поширюють фейкову інформацію.
- Виявити закономірності поширення дезінформації в соціальних мережах.
- Аналізувати зв'язки між медіаплатформами та їхнім контентом.
- Відстежувати походження фейкових новин.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Виявлення спільнот часто використовується для розуміння структури великих і складних мереж. Одним з найпопулярніших алгоритмів для виявлення структури спільноти є так званий алгоритм Лувена. Однак цей алгоритм має серйозний недолік: алгоритм Лувена може давати як завгодно погано пов'язані спільноти. У найгіршому випадку, спільноти можуть навіть бути роз'єднаними, особливо при ітеративному виконанні алгоритму. У нашому експериментальному аналізі ми спостерігаємо, що до 25% спільнот погано пов'язані, а до 16% – роз'єднані. Щоб вирішити цю проблему, ми вводимо алгоритм Лейдена. Ми доводимо, що алгоритм Лейдена дає спільноти, які гарантовано пов'язані. Крім того, ми доводимо, що коли алгоритм Лейдена застосовується ітеративно, він сходиться до розбиття, в якому всі підмножини всіх спільнот локально оптимально розподілені. Крім того, спираючись на підхід швидкого локального переміщення, алгоритм Лейдена працює швидше, ніж алгоритм Лувена. Ми демонструємо продуктивність алгоритму Лейдена для кількох еталонних та реальних мереж. Ми виявляємо, що алгоритм Лейдена швидший за алгоритм Лувена та показує кращі розбиття, крім того, що надає явні гарантії.

Оптимізація модулярності є складною за NP5, і було запропоновано багато евристичних алгоритмів, таких як ієрархічна агломерація, екстремальна оптимізація, імітація відпалу та спектральні алгоритми. Одним з найпопулярніших алгоритмів оптимізації модульності є так званий алгоритм Лувена, названий на честь місця проживання його авторів. Було визнано, що це один із найшвидших та найефективніших алгоритмів у бенчмаркінгу, і це одна з найбільш цитованих робіт у літературі з виявлення спільнот.

У багатьох складних мережах вузли об'єднуються, утворюючи відносно щільні групи, які часто називають спільнотами. Один з найвідоміших методів виявлення спільнот називається модульністю. Неможливо заздалегідь передбачити модульну структуру. Тому виявлення спільнот у соціальній мережі є важливим завданням, яке необхідно вирішити. Рішення цієї проблеми полягає в максимізації різниці між фактичною кількістю ребер у спільноті та очікуваною кількістю таких ребер. Результатом цього методу є виявлення малих мереж шляхом оптимізації модульності, яка є мірою якості поділу вузлів на спільноти. Метод добре працює при використанні модульності та роботі з простим неорієнтованим, незваженим графом.

Завдяки поєднанню локальної оптимізації, рефінування внутрішньої структури кластерів і агрегації графа метод Лейдена дозволяє отримувати стабільні, якісні та логічні спільноти у складних мережах. Він відзначається високою швидкістю роботи навіть на великих графах, надійністю результатів та здатністю виявляти приховані закономірності у даних. Саме тому алгоритм широко використовується у соціальних мережах, біоінформатиці, текстовій аналітиці, а також у задачах виявлення фейкових новин та інформаційних кампаній.

Метод Лейдена базується на ідеї покрокової оптимізації структури об’єднань, де кожен етап спрямований на покращення якості кластерів і забезпечення їхньої внутрішньої зв’язності. Робота алгоритму починається з того, що всі вершини графа вважаються окремими кластерами. Це робиться для того, щоб алгоритм не був обмежений наперед визначеною структурою, а мав змогу вільно формувати оптимальні групи, спираючись лише на зв’язки в самій мережі. Далі алгоритм поступово проходить по всіх вершинах і оцінює, як зміниться якість кластеризації, якщо певну вершину

перенести з одного кластера в інший. На цьому етапі Лейден працює подібно до Лувена, однак це лише перша частина алгоритму.

Після початкового формування кластерів починається друга, ключова фаза — фаза уточнення структури або рефінування. Саме вона робить метод Лейдена принципово відмінним від попередніх рішень. На цьому етапі кожен сформований кластер перевіряється на внутрішню зв'язність, тобто аналізується, чи є він достатньо цілісним. Дуже часто трапляється, що група, яка загалом виглядає правильною, містить сегменти, між якими зв'язки надто слабкі. У традиційних підходах такі слабкі частини залишаються всередині кластерів, що призводить до зниження точності кластеризації. Лейден же автоматично розбиває такі внутрішньо неоднорідні кластери і створює більш логічні та цілісні структури. У результаті кожен отриманий кластер стає не просто групою щільно пов'язаних елементів, а саме структурною одиницею, яка не містить внутрішніх розривів.

Після етапу рефінування відбувається побудова укрупненого графа. Кожен кластер перетворюється на вершину нового графа, а ребра між кластерами — на агреговані зв'язки, ваги яких відповідають сумарним зв'язкам між вершинами у попередньому графі. Цей новий спрощений граф знову стає вхідними даними для алгоритму, і процес повторюється. Така ітеративність дає змогу поступово покращувати структуру кластерів і забезпечує збіжність до якісного рішення.

Однією з найважливіших переваг методу Лейдена є його здатність гарантувати внутрішню зв'язність кластерів. У практичних задачах це означає, що кожний кластер є логічною групою, а не випадковим набором елементів, які об'єдналися лише через частково сильні зв'язки. Внутрішня зв'язність забезпечує високу інтерпретованість кластерів, що є критично важливим у таких сферах, як аналіз соціальних мереж, кластеризація текстів, моделювання інформаційних потоків та виявлення фейкових новин.

Ще однією суттєвою перевагою Лейдена є його стійкість до випадкового вибору початкових умов. Алгоритм Лувена часто давав змінні результати залежно від порядку перебору вершин, тоді як Лейден завдяки phase-refinement працює набагато стабільніше. Це робить його корисним у дослідженнях, де потрібна висока повторюваність та надійність результатів.

Метод Лейдена виявився надзвичайно ефективним у роботі з великими мережами, що містять сотні тисяч або навіть мільйони вершин. Він швидко виконує обчислення і добре масштабується, не втрачаючи якості кластеризації. Саме тому його широко застосовують у задачах аналізу поведінкових моделей користувачів, побудови тематичних груп новин, виявлення інформаційних кампаній та цілеспрямованих вкидів недостовірної інформації. Його можливість виявляти приховані структури у потоках даних робить метод незамінним у галузі кібербезпеки, де необхідно швидко розпізнавати аномальні патерни, такі як синхронне поширення фейкових повідомлень або координація бот-мереж.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Зібраний нами датасет складався з півторатисячі повідомлень, у яких мвістилися фейкові і правдиві новини. Граф фейкових новин представлено на рис. 2. Новини було розподілено на 27 кластерів, кожен з яких був окреслений окремим кольором. На рисунку 3 показано інтерактивний граф, побудований методом Лейдена найвпливовіших новин, який розділений на 7 кластерів, а на рис. 4 – на 5 кластерів.

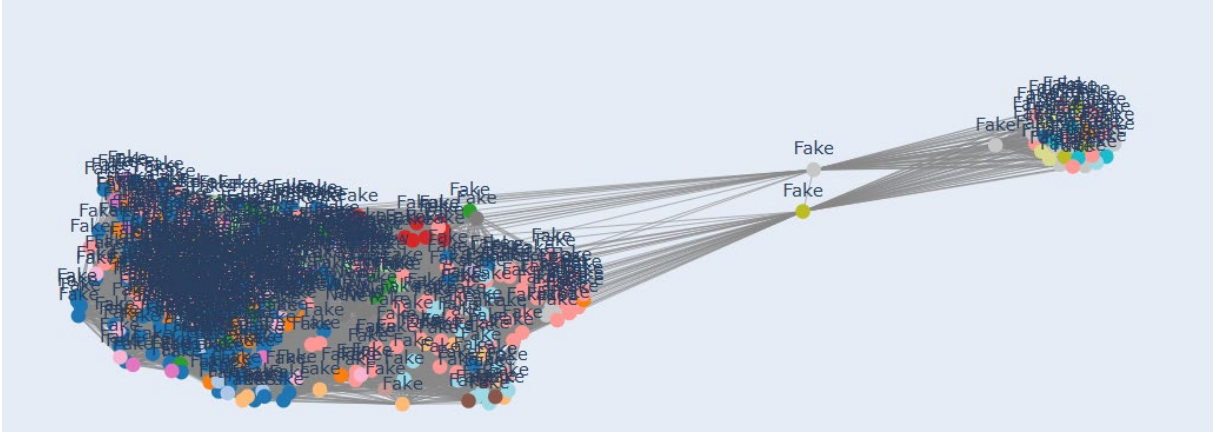

Рис. 2 Граф фейкових новин на 27 кластерів

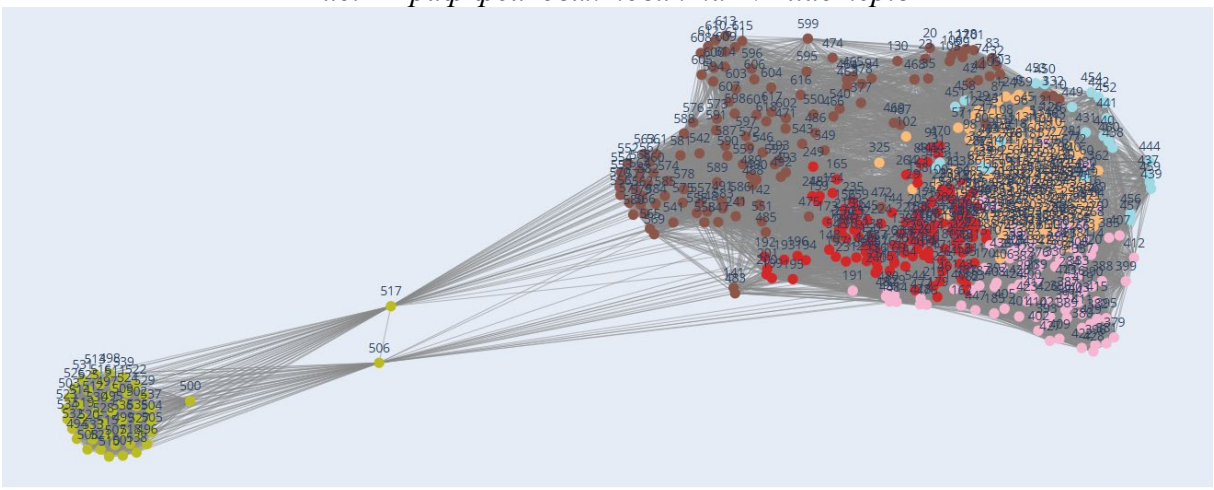

Рис. 3 Граф фейкових новин на 7 кластерів

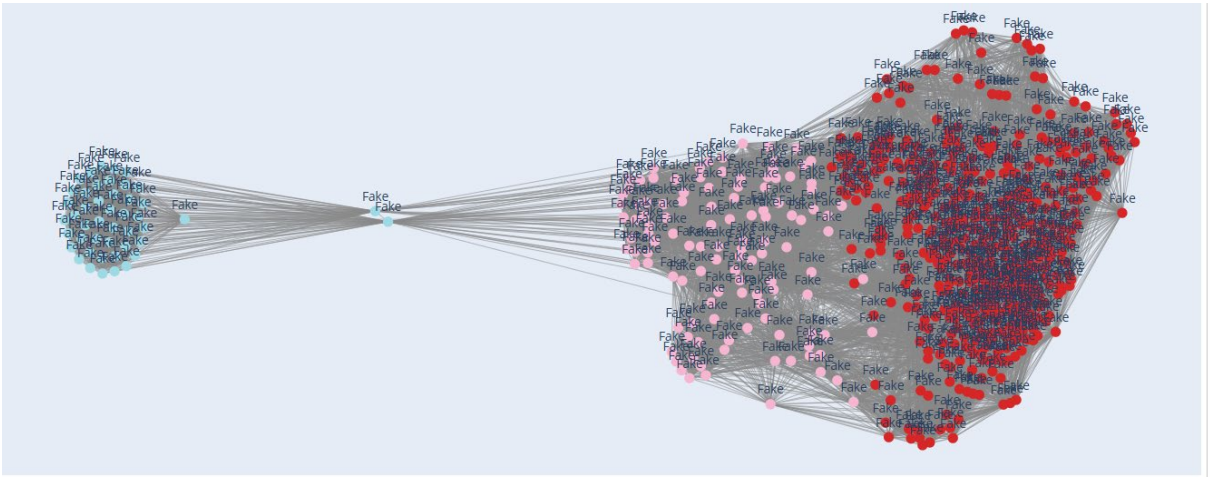

Рис. 4 Граф фейкових новин на 5 кластерів

Наступний експеримент полягав у виокремленні одного кластера новин, вигляд якого приводимо на рисунку 5.

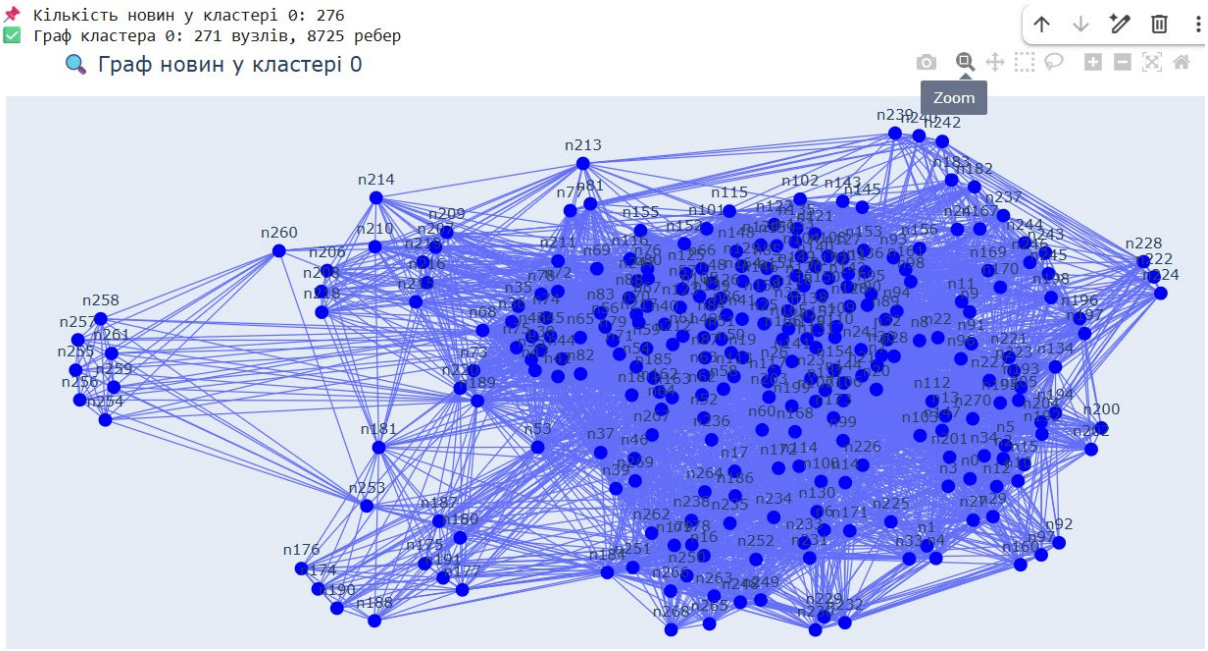

Рис. 5 Граф фейкових новин в одному кластері

Виокремлення найвпливовіших новин та обчислення їх центральності показано у таблиці 1.

Таблиця 1.

Виокремлення найвпливовіших новин та обчислення їх центральності

№ Кластера	Найвпливовіша новина	Кількість зв'язків (ребер)	Швидкість поширення	Центральність, *Q*
1	Заборона експорту української зброї дуже серйозно шкодить нашій оборонній індустрії.	124	15/год	0.35
2	Спецпризначенці ГУР у акваторії Чорного моря збили російський Су-30СМ із ПЗРК	98	12/год	0.29
3	Фінляндія оголосила про новий пакет військової допомоги Україні на €118 млн.	210	22/год	0.42
4	США виділяють Україні $700 мільйонів гуманітарної та енергетичної допомоги	87	9/год	0.25
5	В Омську палає танковий завод "ОмскТрансМаш"	143	18/год	0.38

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Лейденський алгоритм – ефективний інструмент для аналізу новинних мереж та виявлення джерел дезінформації. Завдяки своїй здатності швидко та точно кластеризувати великі обсяги даних, він допомагає медіааналітикам, журналістам та

дослідникам виявляти неправдиву інформацію та запобігати її поширенню. Використання цього методу сприяє покращенню якості інформаційного простору та боротьбі з фейковими новинами.

Лейденський алгоритм довів свою ефективність як метод виявлення спільнот у великих мережах новин та аналізу поширення дезінформації. Використання цього алгоритму дозволяє виявляти щільно пов'язані групи новин, що містять ознаки фейковості, та відокремлювати їх від достовірних повідомлень. Ідентифікувати найшвидше поширювані новини, що є ключовим для своєчасного реагування на дезінформацію. Аналізувати структуру мережі джерел новин, що дає змогу виявляти координовані кампанії та мережі ботів. Підвищувати надійність кластеризації завдяки внутрішній зв'язності кластера та повторюваності результатів незалежно від початкових умов.

Експериментальні результати показали, що застосування методу Лейдена на датасеті з 1500 новин дозволило виділити 27 чітко структурованих кластерів та визначити ключові фейкові повідомлення. Виявлені закономірності поширення інформації підтверджують, що метод забезпечує високу точність та масштабованість аналізу новинних потоків.

Перспективи подальших досліджень включають поєднання методу Лейдена з сучасними моделями глибокого навчання для автоматичного визначення тематики та емоційного забарвлення фейкових новин; інтеграцію алгоритму у системи моніторингу соціальних мереж у режимі реального часу; дослідження ефективності методу у виявленні дезінформації у мультимедійному контенті (зображення, відео); розширення аналізу на приватні платформи обміну повідомленнями, такі як WhatsApp, Telegram та Discord, для виявлення закритих мереж поширення фейків.

Загалом, застосування методу Лейдена створює основу для побудови більш надійних та гнучких систем виявлення фейкових новин, що дозволяє значно підвищити інформаційну безпеку та довіру до медіапростору.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Shivani, D. B. (2017). Techniques of Text Detection and Recognition: A Survey. International Journal of Emerging Research in Management &Technology (IJERMT), 83-87. DOI: 10.1109/TPAMI.2014.2366765
2. Pate, C. A. (2021). Asthma surveillance - United States, 2006–2018. *MMWR. Surveillance Summaries, 70* (SS-5), 1–32. http://dx.doi.org/10.15585/mmwr.ss7005a1
3. Ndidiamaka, A. F., & Awo, I. N. M. Creating social and political capital throughstrategic information and communication management in governance: A content analysis of federal government's information and communication managers'public communication from 2019–. Gregory University, 44.
4. Mohammed, N. (2023). Bangla Plagiarism checker using Machine Learning.
5. Hoy, N., & Koulouri, T. (2021). A systematic review on the detection of fake news articles. arXiv preprint arXiv:2110.11240. https://doi.org/10.48550/arXiv.2110.11240
6. Mishra, S., Sturm, B. L., & Dixon, S. (2017, October). Local interpretable model-agnostic explanations for music content analysis. In ISMIR (Vol. 53, pp. 537-543).
7. Olan, F., Jayawickrama, U., Arakpogun, E. O., Suklan, J., & Liu, S. (2024). Fake news on social media: the impact on society. Information Systems Frontiers *26*(2), 443-458. https://doi.org/10.1007/s10796-022-10242-z
8. Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. Journal of experimental psychology: general, 149(8), 1608.
9. Mendiguren, T., Pérez Dasilva, J., & Meso Ayerdi, K. (2020). Actitud ante las Fake News: Estudio del caso de los estudiantes de la Universidad del País Vasco. Revista de comunicación, *19*(1), 171-184. http://dx.doi.org/10.26441/rc19.1-2020-a10.
10. Ross, A. S., & Rivers, D. J. (2018). Discursive deflection: Accusation of "fake news" and the spread of mis-and disinformation in the tweets of President Trump. Social media+ society. *4*(2), 2056305118776010. https://doi.org/10.1177/20563051187760

11. Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. Journal of cyber policy *3*(1), 39-43. https://doi.org/ 10.1080/23738871.2018.1462395
12. Nelson, J. L., & Taneja, H. (2018). The small, disloyal fake news audience: The role of audience availability in fake news consumption. New media & society. *20*(10), 3720-3737. https://doi.org/10.1177/14614448187587
13. Tajrian, M., Rahman, A., Kabir, M. A., & Islam, M. R. (2023). A review of methodologies for fake news analysis. IEEe Access. *11*, 73879-73893. doi: 10.1109/ACCESS.2023.3294989.
14. Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. International journal of machine learning and cybernetics. *1*, 43-52. https://doi.org/10.1007/s13042-010-0001-0
15. Triyono, A., & Faqih, A. (2025). Implementation of the Naive Bayes Method in Sentiment Analysis of Spotify Application Reviews. Journal of Artificial Intelligence and Engineering Applications (JAIEA), *4*(2), 1091-1097.
16. Yi, J., Xu, Z., Huang, T., & Yu, P. (2025). Challenges and Innovations in LLM-Powered Fake News Detection: A Synthesis of Approaches and Future Directions. arXiv preprint. https://doi.org/10.48550/arXiv.2502.00339
17. Amriza, R. N. S., Chou, T. C., & Ratnasari, W. (2025). Understanding the shifting nature of fake news research: Consumption, dissemination, and detection. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/asi.24980
18. Hoy, N., & Koulouri, T. (2025). An exploration of features to improve the generalisability of fake news detection models. Expert Systems with Applications, 126949. https://doi.org/10.1016/j.eswa.2025.126949
19. Öztürk, A. G., & Turan, M. (2024). Sentiment Analysis Using Machine Learning Methods on News Texts. International Journal of Engineering and Computer Science. https://doi.org/10.18535/ijecs/v13i11.4932.
20. Gadek, G., & Guélorget, P. (2020). An interpretable model to measure fakeness and emotion in news. Procedia Computer Science, *176*, 78-87. https://doi.org/10.1016/j.procs.2020.08.009
21. Rahman, A., Zaman, S., Parvej, S., Shill, P. C., Salim, M. S., & Das, D. (2025). Fake News Detection: Exploring the Efficiency of Soft and Hard Voting Ensemble. Procedia Computer Science.
22. Vo, T. H., Felde, I., & Ninh, K. C. (2025). Fake News Detection System, based on CBOW and BERT. Acta Polytechnica Hungarica, *252*, 748-757. https://doi.org/10.1016/j.procs.2025.01.035
23. Orebi, S. M., & Naser, A. M. (2025). Opinion Mining in Text Short by Using Word Embedding and Deep Learning. Journal of Applied Data Sciences. https://doi.org/10.47738/jads.v6i1.438
24. Chabukswar, A., Shenoy, P. D., & Venugopal, K. R. (2025). Detecting Misinformation in COVID-19 Content: A Machine Learning and Deep Learning Approach with Word Embeddings. SN Computer Science, *6*(1), 1-18. DOI: 10.1109/ACCESS.2020.3022867
25. Nguyen, H. T., Duong, K. H., Pham, L. T. T., Bui, P. H. D., Thai-Nghe, N., & Dien, T. T. (2025). Inverted Index for Similar Document Detection: A Case Study at Can Tho University Journal of Science. SN Computer Science, 6(3), 216. DOI https://doi.org/10.1007/s42979-025-03707-w
26. Jiang, P. T., Zhang, C. B., Hou, Q., Cheng, M. M., & Wei, Y. (2021). Layercam: Exploring hierarchical class activation maps for localization. IEEE Transactions on Image Processing. *30*, 5875-5888. DOI: 10.1109/TIP.2021.3089943

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 2 (30), 2025
Спеціальний випуск ISSN 2663 – 4023
Special issue

Liubomyr Chyrun
PhD in Technical Sciences, Associate Professor
Associate Professor at the Department of Information Systems and Networks
Lviv Polytechnic National University, Lviv, Ukraine
ORCID: 0000-0002-9448-1751
lyubomyr.v.chyrun@lpnu.ua

Andrii Nazarkevych
Student, Department of Information Systems and Networks
Lviv Polytechnic National University, Lviv, Ukraine
ORCID: 0009-0007-2078-8447
andrii.nazarkevych.ri.2023@lpnu.ua

DETECTION OF THE MOST COMMON FAKE NEWS USING LEIDEN CLUSTERING

Abstract. Today, a significant challenge is the automatic identification of fake news and the detection of groups of messages that spread unreliable or manipulative content. One of the effective tools for addressing this task is the Leiden clustering method. The Leiden algorithm belongs to community detection methods in graphs and enables grouping news items based on their internal similarity. For news analysis, a set of characteristics is first formed for each message: its textual content, semantic features, source of dissemination, publication time, and other additional parameters. Based on these data, a graph is created in which vertices represent individual news items, while edges describe the degree of their similarity. The Leiden method performs the step-by-step merging of such news items into clusters, ensuring high-quality partitioning due to iterative refinement of group structures. Using Leiden clusters makes it possible to identify concentrations of messages that exhibit signs of falsification or coordinated dissemination. Typically, such groups are characterized by repeated phrases, identical text templates, accelerated dissemination dynamics, and a large number of cross-references between non-authoritative sources. Analyzing these structures allows the detection of targeted information campaigns, bot networks, and mass distributions of false content. Thus, the Leiden clustering method is an effective approach for summarizing and grouping news messages, significantly improving the accuracy of fake-news detection. Its application in media monitoring systems provides deeper insight into information flows and contributes to the timely identification of disinformation.

Keywords: fake news, deep learning, protected data

REFERENCES

1. Shivani, D. B. (2017). Techniques of Text Detection and Recognition: A Survey. International Journal of Emerging Research in Management &Technology (IJERMT), 83-87. DOI: 10.1109/TPAMI.2014.2366765
2. Pate, C. A. (2021). Asthma surveillance - United States, 2006–2018. *MMWR. Surveillance Summaries, 70* (SS-5), 1–32. http://dx.doi.org/10.15585/mmwr.ss7005a1
3. Ndidiamaka, A. F., & Awo, I. N. M. Creating social and political capital throughstrategic information and communication management in governance: A content analysis of federal government's information and communication managers'public communication from 2019–. Gregory University, 44.
4. Mohammed, N. (2023). Bangla Plagiarism checker using Machine Learning.
5. Hoy, N., & Koulouri, T. (2021). A systematic review on the detection of fake news articles. arXiv preprint arXiv:2110.11240. https://doi.org/10.48550/arXiv.2110.11240
6. Mishra, S., Sturm, B. L., & Dixon, S. (2017, October). Local interpretable model-agnostic explanations for music content analysis. In ISMIR (Vol. 53, pp. 537-543).
7. Olan, F., Jayawickrama, U., Arakpogun, E. O., Suklan, J., & Liu, S. (2024). Fake news on social media: the impact on society. Information Systems Frontiers *26*(2), 443-458. https://doi.org/10.1007/s10796-022-10242-z

8. Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. Journal of experimental psychology: general, 149(8), 1608.
9. Mendiguren, T., Pérez Dasilva, J., & Meso Ayerdi, K. (2020). Actitud ante las Fake News: Estudio del caso de los estudiantes de la Universidad del País Vasco. Revista de comunicación, *19*(1), 171-184. http://dx.doi.org/10.26441/rc19.1-2020-a10.
10. Ross, A. S., & Rivers, D. J. (2018). Discursive deflection: Accusation of "fake news" and the spread of mis-and disinformation in the tweets of President Trump. Social media+ society. *4*(2), 2056305118776010. https://doi.org/10.1177/20563051187760
11. Morgan, S. (2018). Fake news, disinformation, manipulation and online tactics to undermine democracy. Journal of cyber policy *3*(1), 39-43. https://doi.org/ 10.1080/23738871.2018.1462395
12. Nelson, J. L., & Taneja, H. (2018). The small, disloyal fake news audience: The role of audience availability in fake news consumption. New media & society. *20*(10), 3720-3737. https://doi.org/10.1177/14614448187587
13. Tajrian, M., Rahman, A., Kabir, M. A., & Islam, M. R. (2023). A review of methodologies for fake news analysis. IEEe Access. *11*, 73879-73893. doi: 10.1109/ACCESS.2023.3294989.
14. Zhang, Y., Jin, R., & Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. International journal of machine learning and cybernetics. *1*, 43-52. https://doi.org/10.1007/s13042-010-0001-0
15. Triyono, A., & Faqih, A. (2025). Implementation of the Naive Bayes Method in Sentiment Analysis of Spotify Application Reviews. Journal of Artificial Intelligence and Engineering Applications (JAIEA), *4*(2), 1091-1097.
16. Yi, J., Xu, Z., Huang, T., & Yu, P. (2025). Challenges and Innovations in LLM-Powered Fake News Detection: A Synthesis of Approaches and Future Directions. arXiv preprint. https://doi.org/10.48550/arXiv.2502.00339
17. Amriza, R. N. S., Chou, T. C., & Ratnasari, W. (2025). Understanding the shifting nature of fake news research: Consumption, dissemination, and detection. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/asi.24980
18. Hoy, N., & Koulouri, T. (2025). An exploration of features to improve the generalisability of fake news detection models. Expert Systems with Applications, 126949. https://doi.org/10.1016/j.eswa.2025.126949
19. Öztürk, A. G., & Turan, M. (2024). Sentiment Analysis Using Machine Learning Methods on News Texts. International Journal of Engineering and Computer Science. https://doi.org/10.18535/ijecs/v13i11.4932.
20. Gadek, G., & Guélorget, P. (2020). An interpretable model to measure fakeness and emotion in news. Procedia Computer Science, *176*, 78-87. https://doi.org/10.1016/j.procs.2020.08.009
21. Rahman, A., Zaman, S., Parvej, S., Shill, P. C., Salim, M. S., & Das, D. (2025). Fake News Detection: Exploring the Efficiency of Soft and Hard Voting Ensemble. Procedia Computer Science.
22. Vo, T. H., Felde, I., & Ninh, K. C. (2025). Fake News Detection System, based on CBOW and BERT. Acta Polytechnica Hungarica, *252*, 748-757. https://doi.org/10.1016/j.procs.2025.01.035
23. Orebi, S. M., & Naser, A. M. (2025). Opinion Mining in Text Short by Using Word Embedding and Deep Learning. Journal of Applied Data Sciences. https://doi.org/10.47738/jads.v6i1.438
24. Chabukswar, A., Shenoy, P. D., & Venugopal, K. R. (2025). Detecting Misinformation in COVID-19 Content: A Machine Learning and Deep Learning Approach with Word Embeddings. SN Computer Science, *6*(1), 1-18. DOI: 10.1109/ACCESS.2020.3022867
25. Nguyen, H. T., Duong, K. H., Pham, L. T. T., Bui, P. H. D., Thai-Nghe, N., & Dien, T. T. (2025). Inverted Index for Similar Document Detection: A Case Study at Can Tho University Journal of Science. SN Computer Science, 6(3), 216. DOI https://doi.org/10.1007/s42979-025-03707-w
26. Jiang, P. T., Zhang, C. B., Hou, Q., Cheng, M. M., & Wei, Y. (2021). Layercam: Exploring hierarchical class activation maps for localization. IEEE Transactions on Image Processing. *30*, 5875-5888. DOI: 10.1109/TIP.2021.3089943