

ПРИКЛАДНА ЕКОНОМІКА

УДК 338.2; 339.13
JEL Classification M21, M31
DOI 10.33111/EE.2022.48.Oleksiuk O_ ShafalyukA

O. Oleksiuk
*Dr. Sc. (Econ.), Prof.,
Kyiv National Economic University
named after Vadym Hetman*

O.I. Олексюк
*д.е.н., проф.,
Київський національний
економічний університет
імені Вадима Гетьмана*

ORCID: <https://orcid.org/0000-0002-3407-6201>

A. Shafalyuk
*Dr. Sc. (Econ.), Prof.,
Kyiv National Economic University
named after Vadym Hetman*

O.K. Шафалюк
*д.е.н., проф.,
Київський національний
економічний університет
імені Вадима Гетьмана*

ORCID: <https://orcid.org/0000-0003-1145-7973>

ВИКОРИСТАННЯ ІНФОРМАЦІЇ З ВІДКРИТИХ ДЖЕРЕЛ ДЛЯ МАРКЕТИНГОВИХ ДОСЛІДЖЕНЬ ТОВАРНИХ РИНКІВ

АНОТАЦІЯ. Опрацьовано альтернативні способи збору і обробки даних з відкритих джерел інформації, що формують загальне розуміння ситуацій і тенденцій розвитку бізнесу на цільових ринках товарів і послуг. Розроблено рекомендації щодо удосконалення аналітичних технологій обробки інформаційних масивів з відкритих або широко доступних джерел інформації в системі маркетингових досліджень підприємств. Вивчаються і знайшли апробацію дієві способи оцінки якості та достовірності даних з відкритих або широко доступних джерел інформації. Пропонується та апробовано авторські підходи щодо виявлення недостовірної інформації (фейків) в інформаційному полі окремого товарного ринку. Запропоновано авторський варіант алгоритму організації якісного та ефективного опрацювання даних з відкритих джерел в системі інформаційного маркетингового забезпечення управління бізнесом.

КЛЮЧОВІ СЛОВА: маркетингові дослідження, товарний ринок, менеджмент, достовірність інформації, машинне навчання

USE OF OPEN SOURCES INFORMATION FOR MARKETING RESEARCH IN PRODUCTS MARKETS

ANNOTATION. Alternative methods of collecting and processing data from open sources of information that form a common understanding of situations and trends in business development in target markets for goods and services have been developed. Recommendations for the improvement of analytical technologies for processing information arrays from open or widely available sources of information in the system of marketing research of enterprises are offered. Effective ways to assess the quality and reliability of data from open or widely available sources of information are being studied and tested. Improving approaches to integration into modern marketing information systems of modern content analysis tools, which organically provides constructive synergy of applied methods and tools of neuro-fuzzy modeling and clustering of information arrays, statistical analysis of information units, identification of psycho-emotional number of potential and actual consumers. The author's variant of the algorithm of the organization of qualitative and effective processing of data from open sources in the system of information marketing support of business management is offered. It is shown that the great variability of the modern information environment (data, content) creates significant prerequisites and significant combinatorial opportunities for generating distorted information in different ways, as well as the dissemination of the latter. Possibilities for detecting inaccurate information (fakes) in the information field of a particular product market have been worked out. Comparison of the results of different models based on the confusion matrix showed that two of the four learning models, namely the neural network and the «random forest» model, did well enough to assess the reliability messages.

KEY WORDS: marketing research, product market, management, reliability of information, machine learning

Вступ. Виходячи з існуючого сьогодні контексту розвитку українських компаній, побудова планів розвитку стає непростою задачею при обмежених ресурсах компаній. Відзначимо також, що класичні підходи до аналізу місткості ринку, структуризації сегментів споживачів та прогнозування попиту мають обмежену придатність для використання за умов війни, адміністративного регулювання ринків та стресової поведінки населення. Впевнені, що спільними зусиллями науковців та практиків вдасться вибудувати ефективний підхід до дослідження ринків чи медійного простору навколо них.

Постановка завдання. Перш за все відзначимо, що проблематику маркетингових досліджень розкривали у своїх роботах багато вітчизняних та зарубіжних науковців протягом вже досить тривалого періоду часу. Варто згадати роботи таких українських вчених

як А.Ф. Павленко, А.В. Войчака, А.М. Окландера, А.В. Федорченка, С.О. Солнцева, О.В. Зозулева, Ю.І. Яковенка, В.Г. Андрєєнкова тощо. Безпосередній зв'язок з представленими у цій роботі розробками мають праці Д. Фельдмана (Dr. R. Feldman), Дж. Сангера (J. Sanger), (Б. Берелсона (B. Berelson), О. Холсті (Ole R. Holsti), К. Кріппендора (K. Krippendor), Р. Кортена (R. Corten), Х. Мюра (H. Miura), Я. Янга (Y. Yang) [1] та багатьох інших спеціалістів у сфері як статистичного аналізу, так і дослідження текстів як джерела інформації [2, 3].

Завданнями даної роботи слід вважати підготовку переліку можливих альтернативних способів збору та обробки інформації з відкритих джерел, як основи для формування базового розуміння тенденцій у певній сфері бізнесу чи на окремому товарному ринку. Також наші зусилля будуть спрямовані на побудову алгоритму чи удосконалення аналітичної технології обробки інформаційного масиву з відкритих і доступних джерел інформації у процесі проведення маркетингових досліджень. У якості емпіричної основи даної роботи виступатиме тематичний інформаційний масив сформований з українських медіа ресурсів, інформаційних Telegram-каналів, сторінок у соціальних мережах, візуального контенту з офіційних сайтів компаній тощо. Для уникнення можливих питань щодо доступності та авторства даної інформації, зазначимо, що дослідження проводиться на основі публічно оприлюднених даних, масив інформації був зібраний спеціально для підготовки даної роботи на основі доступного безкоштовного спеціалізованого програмного забезпечення. Висновки та рекомендації авторів сфокусовані на демонстрації можливостей авторських розробок і не мають на меті рекламування чи навпаки спотворення інформацію про будь-якого оператора ринку, компанії чи персону.

Особливим аспектом, який автори хотіли опрацювати у цій роботі є оцінки якості та достовірності інформації, доступної для аналізу. У цьому випадку в нагоді має стати досить розвинутий сучасний інструментарію контент-аналізу, який використовує розробки багатьох суміжних галузей знань: статистичного аналізу інформаційних одиниць, нейро-нечіткого моделювання та кластеризації інформаційних масивів, психології сприйняття інформації людиною та інші спеціалізовані предметні області в технічних та гуманітарних науках [4, 5].

Показово можна виділити окремі сфери економічної діяльності, які потенційно можуть найбільше виграти від використання розвинених методичних підходів до аналізу інформації з відкритих джерел:

- Сфера гостинності, яка об'єднує готелі, ресторани, туризм та розваги, має досить тісний зв'язок з медійним простором та інформацією у відкритих джерелах. Тут слід вести мову про двосторонній зв'язок: вплив відкритої інформації на фінансові результати діяльності окремої компанії у певному регіоні; необхідність аналізу інформаційного простору для правильної комунікації з потенційними споживачами послуг цих галузей. Додатково слід вказати, що репутаційні ризики, доступність корисної інформації, поширення відомостей про події, вивчення відношення відвідувачів та багато інших аспектом є основним фокусом роботи відділу маркетингу компаній сфери гостинності. Напевно ключовими аспектами дослідження інформаційних потоків виступає тут інструментарій *sentiment analysis*, який буде спиратися на кластеризацію масиву відповідно до ринкових сегментів.

- Сфера фінансових операцій, банківський та страховий сектори економіки використовують інформацію з відкритих джерел для скрінінгу потенційних клієнтів чи партнерів, перевірки достовірності наданої інформації, мінімізації ризиків повернення кредитів та для багатьох інших цілей. Окремим блоком тут виступає аналіз ризиків та комплайн-менеджмент, який вже спирається на розвинутий статистично-математичний інструментарій. Відповідно до цілей компаній цих сфер, дослідження інформації з відкритих джерел буде сфокусовано на побудові моделей аналізу частотності, словоформних та лексичних конструкцій, крос-аналізу інформації відкритих та закритих джерел.

- Медичні установи та фармацевтичні компанії також можуть отримати достатньо багато переваг на основі якісного аналізу інформації відкритих джерел та більш ефективно реалізувати власні корпоративні стратегії. Наприклад, тут можна говорити про вивчення сприйняття державного регулювання та змін ринкової поведінки операторів, статистично встановлювати зв'язок між запровадженням того чи іншого підходу до охорони здоров'я та якістю донесення його суті до населення, аналізувати достовірність інформаційних повідомлень фармацевтичних компаній тощо

- Маркетингові, PR та рекламні агенції фактично є ключовими споживачами і генераторами інформації у відкритих каналах комунікацій як на національному, галузевому чи регіональному рівнях. Сюди також слід віднести внутрішні комунікації компаній та їх об'єднань, що зараз є частиною соціальної відповідальності у різних сферах бізнесу. Фактично компанії цих сфер використовують найбільш широкий спектр спеціальних підходів та моделей у сфері контент-аналізу чи text-mining.

- Роздрібна торгівля товарами та послугами, на наш погляд, також сильно залежить від впливу інформаційного простору. Більшість споживачів (покупців) отримують інформацію з відкритих джерел, планують свої покупки та час їх здійснення також на цій основі. Відкрита інформація суттєво визначає поведінки людей у сучасному інформаційно насиченому комунікаційними технологіями світі [5, 6].

Результати дослідження. Спробуємо представити власне розуміння процесу організації якісної та ефективної обробки інформації з відкритих джерел у відповідних підсистемах маркетингової діяльності у вигляді певної послідовності кроків (рис. 1).

Представлена на рисунку логіка має в основному ілюстративний характер та покликана показати ключові блоки робіт. Зазвичай багато окремих блоків, представлених на цій схемі, можна ефективно об'єднати в межах одного або декількох програмних продуктів. На рис. 1 ми спробували показати логіку список окремих блоків від підготовки об'єктів та інструментів пошуку інформації в базі даних, формування цільових наборів даних до оцінки якості виконаних робіт, уточнення пошукових підходів та інше. Фактично фінальним етапом таких аналітичних операцій стане підготовка та розробка моделі аналізу, як структурованого і визначного підходу до аналізу цільової інформації в базі даних. Сформована модель має відповідати і цілям аналізу інформації з відкритих джерел і характеру атрибутованих доступних даних, витратам часу та ресурсів на її використання. Беззаперечно основним параметром вибору тої чи іншої моделі аналізу виступатиме якість отриманих на її основі результатів для розробки маркетингових рішень, прогнозування змін тощо.



Рис. 1. Організація обробки та використання інформації з відкритих джерел у маркетингових дослідженнях

Джерело: власна розробка

Останні два рядки моделі рис. 1 ми присвятили опису етапів інтерпретації, візуалізації та інтеграції отриманих результатів в маркетинговий менеджмент. Керівник вищого рівня не бачить всі попередні роботи, але спирається на отримані результати аналізу інформації з відкритих джерел. На цьому етапі починає працювати логіка розробки управлінського рішення – планування, організація, мотивація та контроль, з наступним необхідним коригуванням. Так от фаза планування, яку часто ще розділяють на прогнозування і планування, вже буде використовувати якісні та зрозумілі результати аналізу.

Прикладні можливості описаного вище підходу ми не будемо розкривати у розрізі всіх етапів, які досить легко можна об'єднати у окремі більш загальні блоки і локалізувати за окремим робочим місцем маркетингового аналітика в компанії. Натомість спробуємо представити результати нашого підходу до виявлення «фейкової» інформації в інформаційному полі окремого товарного ринку. В якості прикладу нами сформовано базу з інформаційних повідомлень про діяльність компанії Shell на ринках країн пост-радянського простору за квітень-травень 2022 р. Зрозуміло, що поточна ситуація у нашому регіоні є нестабільною, що призвело до значної кількості повідомлень різного типу – «правдивих» і «неправдивих». Взагалі, інформаційні потоки були сформовані з новин інформаційних агенцій, сторінок у соціальних мережах та інформаційних каналів месенджерів. Загальна чисельність інформаційних повідомлень на момент підготовки цієї роботи складала понад 80 тисяч англomовних, 250 тисяч російськомовних і біля 100 тисяч україномовних новин.

Перш ніж продовжити прикладні дослідження вважаємо приділити окрему увагу аспекту ідентифікації «фейковості» або недостовірності інформації, а також похідних продуктів підготовлених на цій основі – «фейкових новин». Говорячи про недостовірність або «фейковість» інформації природньо постає питання щодо критеріїв її перевірки чи порівняння. Велика варіативність сучасного контенту (даних) дозволяє також говорити про значні комбінаторні можливості генерування спотворених тим чи іншим чином інформаційних повідомлень. Використовуючи інформацію з відкритих джерел для наповнення корпора-

тивних баз даних і відслідковування ключових подій, маркетингові аналітики повинні також вміти виділяти недостовірну або спотворену інформацію. Традиційно під «фейками» розуміють неправдиві повідомлення, які розповсюджуються по каналах комунікації як правдиві і впливають на суспільну думку. Таке досить узагальнення визначення дозволить включити у нього значну різноманітність реальних сучасних видів неправдивих повідомлень. Відзначимо також, що проблема «фейкизації» інформаційного повідомлення не є чимось новим, чи феноменом який виникнув у діджиталізації. Варто згадати поняття «жовта преса», яка і сьогодні широко використовуються і в Україні, і в цілому у світі. Але розвиток Інтернет і комунікаційних технологій дійсно став потужним поштовхом досліджень у цій сфері, розробки базису легітимізації боротьби з «фейковими» повідомлення та розбудови юридичної відповідальності за порушення певних суспільних норм.

У найбільш загальному вигляді варто виділити наступні два основні класи «фейкових» повідомлень: 1) повідомлення з недостатньою інформацією; 2) повідомлення зі спотвореною інформацією тим чи іншим чином. Далі можна виділити більш «міксовані види фейків»: пропаганда, карикатури, чутки, спам, клікбейти, персональні оцінки, меми та багато іншого. Наявність такого широкого спектру різновидів «фейків» вимагатиме окремого дослідження та розробки підходу до їх класифікації, але з позицій маркетингового дослідження слід вказати на суттєві аспекти:

- здійснення впливу на суспільну думку (конкуренти, партнери та інші стейкхолдери);
- здійснення вплив на ринкову поведінки особи;
- спотвореність інформації в повідомленні, без різниці яким чином і з якою ціллю.

Фактично інформаційна політика компаній на цільових ринках спрямована на досягнення певних маркетингових цілей, а це призводить до того, що використання достовірних та недостовірних даних у інформаційних повідомленнях є звичайною частиною реалізації комунікацій. У найбільш загальному вигляді можна рекомендувати використання приблизно такого дерева рішень при ідентифікації або оцінюванні «фейковості» повідомлення (рис. 2).

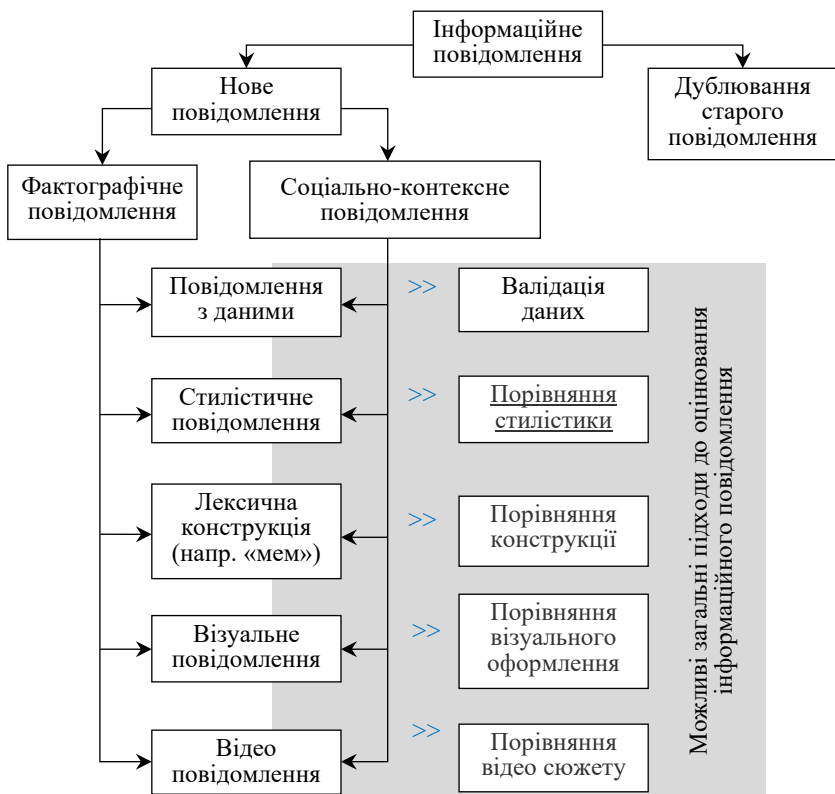


Рис. 2. Дерево рішень щодо оцінювання «фейковості» повідомлення з відкритих інформаційних джерел

Джерело: власна розробка

На основі такого базового розуміння рис. 2 можна рекомендувати, з використанням блоків рис. 1, певні етапи організації оцінювання інформаційних повідомлень у процесі маркетингових досліджень:

1. Ідентифікація повідомлення – блоки 1-8
2. Парсинг даних – блоки 7-11, 18-20
3. Перевірка якості даних – блоки 14-17, 22, 25
4. Аналіз даних – блоки 4, 12-13, 21-24
5. Використання даних – блоки 26-28.

На основі описаної вище вибірки публікацій було проведено підготовку бази даних, як це було описано раніше у нашій роботі. Для скорочення опису цих технічних операцій та візуалізації отриманих результатів наведемо тільки на рис. 3 загальну «хмару» ключових слів виділених на основі їх частотності у нашому масиві:

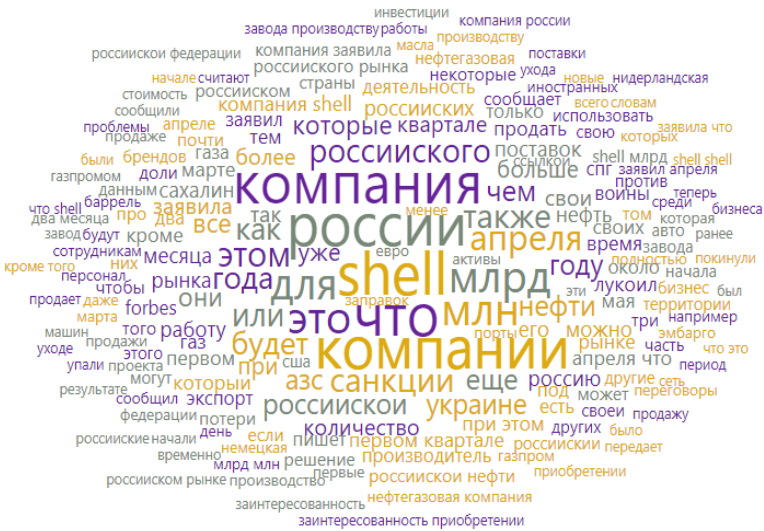


Рис. 3. Хмара ключових слів за їх частотністю в вибірці

Джерело: власна розробка

На перший погляд, з представленої на рис. 3 інформації видно, що у нашому масиві домінують характерні для ринку компанії Shell поняття та терміни, а також найбільш актуальні події останнього періоду, а саме її вихід з російського ринку під дією санкцій світової спільноти. Наступним кроком став також поглиблений базовий аналіз текстового масиву шляхом очищення від «несуттєвих» слів (слів-зв'язок, скорочень тощо). Все це створило необхідні передумови для дієвого машинного навчання з використанням різних його алгоритмів та категоріальної змінної-маркера фейковості новин.

Для візуалізації створеної моделі загальної підсистеми аналізу інформації нашої вибірки повідомлень представляємо наступну схему (рис. 4). Фактично на цьому рисунку ми віалізували тільки вузли, що відповідають блокам 20-25 рисунку 1, або змістовну частину нашого

аналізу. Відзначимо, що всі інші блоки винесені за межі цього фрагменту і представлені тільки двома потоками даних – один потік даних для навчання моделей, а інших фактично аналіз тестової вибірки. Як легко можна помітити на схемі всі результати зводяться до блоку «передбачення», тобто розрахункового модулю оцінювання повідомлень на основі коефіцієнтів навчання різних моделей.

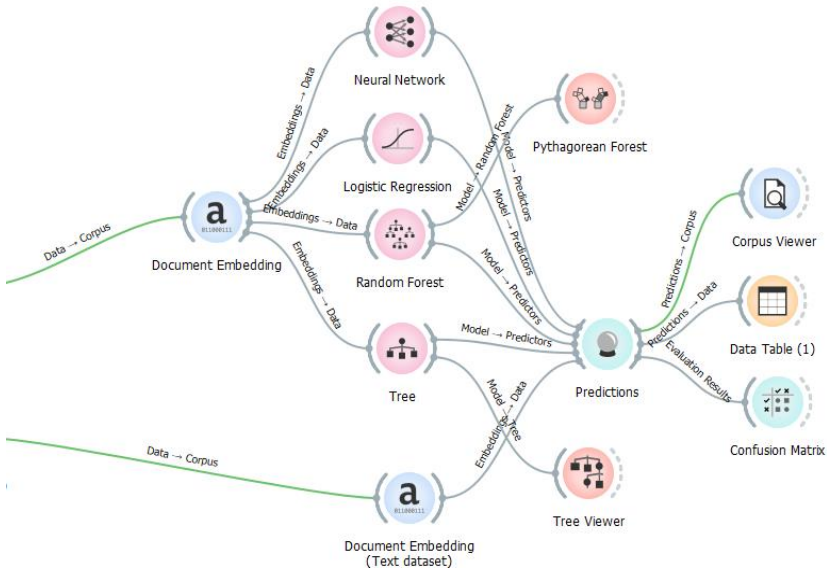


Рис. 4. Фрагмент тестової моделі оцінювання інформації з відкритих джерел за повідомленнями щодо діяльності компанії Shell та з використанням Orange Data Mining Software

Джерело: власна розробка

Співставлення результатів різних моделей на базі матриці розбіжностей (confusion matrix) показали, що дві (нейромережа і модель «випадкового лісу») з чотирьох моделей навчання, представлених на рис. 4, достатньо добре впоралися з оцінюванням достовірності («фейковості») повідомлень. Ці моделі забезпечили повну відповідність прогнозованих та реальних оцінок достовірних/фейкових повідомлень, відповідно, 17/24.

Звичайно розрахункові значення нашої тестової вибірки потребують більшого статистичного наповнення і можливо не можуть бути прямо інтерпретовані у процесі маркетингових досліджень,

але вони створюють основу для результативного порівняння моделей, забезпечують достатньо показову апробацію запропонованих підходів.

Висновки. Формування ефективної ситуаційної обізнаності менеджменту за умов складної кон'юнктури ринків, станів високої невизначеності перспектив розвитку бізнесу або форс-мажорних і глобальних за масштабами впливу обставин є складним, але надзвичайно актуальним, завданням підсистем маркетингових досліджень в управлінні компаніями. При цьому, за критеріями економії та доступності інформації у забезпеченні результативності сучасного проактивного менеджменту, особливої уваги науковців і практиків набувають відкриті джерела даних для цілей управління бізнесом.

Важливою особливістю аналізу інформації з відкритих джерел є їх неструктурованість та різноманітність. Інформаційні канали можуть надавати зовсім різні дані, які на основі атрибутування можна розподілити на нечіткі підмножини, але навіть в середині таких масивів досягнути однорідності даних майже неможливо.

Дана робота покликана продемонструвати можливості існуючих алгоритмів машинного навчання у маркетингових дослідженнях і не претендує на абсолютну достовірність з позицій розробки остаточних рішень менеджменту компаній щодо розвитку бізнесу, проактивного оцінювання ситуацій конкретного ринку (в нашому випадку нафтопродуктів). Використання досвіду роботи компанії на її цільових ринках, з використанням відповідних результатів обробки інформації відкритих джерел, дозволить їй напрацювати найбільш ефективні та якісні прийоми формування ситуаційної обізнаності менеджменту.

Література

1. Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Xiaoming Zhang, Zhoujun Li, Philip S. Yu. TI-CNN: Convolutional Neural Networks for Fake News Detection. arXiv:1806.00749v1 [cs.CL] – Vol. 3, Jun 2018. URL: <https://arxiv.org/pdf/1806.00749.pdf>
2. Eslam Amer, Kyung-Sup Kwak, Shaker El-Sappagh. Context-Based Fake News Detection Model Relying on Deep Learning Models. Electronics – Vol. 11(8), April 2022. URL: <https://doi.org/10.3390/electronics11081255>
3. Jamal Abdul Nasir, Osama Subhani Khan, Iraklis Varlamis. Fake news detection: A hybrid CNN-RNN based deep learning approach. International Journal of Information Management Data Insights – Vol. 1, Issue 1, April 2021. URL: <https://doi.org/10.1016/j.ijime.2020.100007>.

4. Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, Soumendu Sinha. FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research* – Vol. 61, June 2020, p. 32-44. URL: <https://doi.org/10.1016/j.cogsys.2019.12.005>.

5. Somya Ranjan Sahoo, B.B. Gupta. Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing* – Vol. 100, March 2021. URL: <https://doi.org/10.1016/j.asoc.2020.106983>

6. Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang. A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021)* – Vol. 2, p. 1066-1072. DOI: 10.5220/0010316010661072. URL: <https://www.scitepress.org/Papers/2021/103160/103160.pdf>

References

1. Yang Yang, Lei Zheng, Jiawei Zhang, Qingcai Cui, Xiaoming Zhang, Zhoujun Li, Philip S. Yu. TI-CNN: Convolutional Neural Networks for Fake News Detection. arXiv:1806.00749v1 [cs.CL] – Vol. 3, Jun 2018. URL: <https://arxiv.org/pdf/1806.00749.pdf> [in English].

2. Eslam Amer, Kyung-Sup Kwak, Shaker El-Sappagh. Context-Based Fake News Detection Model Relying on Deep Learning Models. *Electronics* – Vol. 11(8), April 2022. URL: <https://doi.org/10.3390/electronics11081255> [in English].

3. Jamal Abdul Nasir, Osama Subhani Khan, Iraklis Varlamis. Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights* Vol. 1, Issue 1, April 2021. URL: <https://doi.org/10.1016/j.jjime.2020.100007> [in English].

4. Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang, Soumendu Sinha FNDNet – A deep convolutional neural network for fake news detection. *Cognitive Systems Research* Vol. 61, June 2020, p. 32-44. URL: <https://doi.org/10.1016/j.cogsys.2019.12.005> [in English].

5. Somya Ranjan Sahoo, B.B. Gupta. Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing* – Vol. 100, March 2021. URL: <https://doi.org/10.1016/j.asoc.2020.106983> [in English].

6. Rohit Kumar Kaliyar, Anurag Goswami, Pratik Narang. A Hybrid Model for Effective Fake News Detection with a Novel COVID-19 Dataset. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence (ICAART 2021)* – Vol. 2, p. 1066-1072. DOI: 10.5220/0010316010661072. URL: <https://www.scitepress.org/Papers/2021/103160/103160.pdf> [in English].

Стаття надійшла до редакції 30.03.2022 р