

УДК 338.2; 339.13
JEL Classification M21, M31
DOI 10.33111/sedu.2022.50.045.055

*Олексюк Олексій Іванович**
*Шафалюк Олександр Казимирович***

ОСОБЛИВОСТІ ВИКОРИСТАННЯ ВІДКРИТИХ ДАНИХ У ПІДГОТОВЦІ РІШЕНЬ МАРКЕТИНГОВОГО ТА КОМЕРЦІЙНОГО МЕНЕДЖМЕНТУ

Анотація. У статті досліджено особливості використання даних з відкритих джерел інформації, спеціальних підходів та прийомів в системі маркетингових досліджень підприємств. Розроблено та апробовано удосконалюючі підходи щодо інтеграції в існуючі маркетингові інформаційні системи сучасного інструментарію контент-аналізу. Здійснено апробацію сучасних методичних інструментів машинного навчання на базі авторських підходів щодо виявлення «фейків» (недостовірної інформації) при формуванні ситуаційної обізнаності менеджменту компаній, аналітичного опрацювання змін, що пов'язані з високими ризиками і невизначеністю перспектив для бізнесу. Конструктивно опрацьовано рекомендації щодо організації формування вихідних масивів даних з відкритих джерел інформації, підвищення якості та достовірності їх опрацювання та успішної інтеграції системах маркетингової і комерційної аналітики.

Ключові слова: відкриті дані; достовірність інформації; машинне навчання; маркетинг; комерційна діяльність; менеджмент.

Вступ. Сучасний складний етап розвитку України супроводжується високими ризиками та виключною нестабільністю змін на всіх без винятку товарних ринках. Військові дії та ряд урядових кроків обмежувального характеру стимулюють пошук нових управлінських рішень щодо розвитку національних компаній за сучасних умов. Слід відзначити, що виключність та нестандартність поточних завдань, які постали перед вітчизняними підприємцями та інвесторами, важко переоцінити. Поряд з обмеженістю доступних ресурсів, значними соціальними зрушеннями в українському суспільстві, ризиками військових дій, відтоком людського капіталу та багатьма іншими факторами, простір альтернатив для корпоративних рішень залишається вкрай незначний. Не стала виключенням і сфера маркетингових досліджень і аналізу комерційних операцій, яка за-

* **Олексюк Олексій Іванович** — доктор екон. наук, професор кафедри комерційної діяльності та логістики, ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана» (м. Київ, Україна), ORCID 0000-0002-3407-6201, oleksiuk@kneu.edu.ua

** **Шафалюк Олександр Казимирович** — доктор екон. наук, професор кафедри маркетингу, ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана» (м. Київ, Україна), ORCID 0000-0003-1145-7973, dean_marketing@kneu.edu.ua

звичай передує будь-якій розробці бізнес-планів чи початку виробництва тих чи інших товарів.

Постановка завдання. Метою даної роботи виступає розвиток методичних підходів до проведення маркетингових досліджень і практичних способів використання інформації відкритих джерел, з урахуванням і підвищення рівня її достовірності, для розробки управлінських рішень. Слід відзначити, що під відкритими джерелами інформації ми будемо розуміти у подальшому всі доступні для дослідника, незаборонені діючим законодавством, способи отримання інформації будь-якого типу (графічного, відео, текстового формату тощо). Зазвичай під відкритими джерелами інформації розуміють публічні засоби масової інформації, відкриті інформаційні агенції, оприлюдненні повідомлення тощо. Необхідність наголосити на цьому аспекті обумовлена широким переліком каналів комунікації на сучасних ринках чи у окремих професійних спільнотах.

Серед дослідників, які мають важливі напрацювання для предметного поля даної публікації, слід особливо згадати С. Хангло (S. Hangloo), Б. Аора (B. Aora), Х. Жоу (X. Zhou), Р. Зафарані (R. Zafarani), Д. Хатара (D. Khattar), М. Гупіа (M. Gupta) [1], Д. Гоуда (J. Goud) та багатьох інших, що особливу увагу сфері визначення «фейковості» інформації в медійному просторі. Проблематика статті, розвивалася і розвивається у їх працях та інших авторів, як щодо формування конструктивного бачення актуальних соціально-економічних питань і ситуацій, так і щодо генерування відповідних рішень і рекомендацій бізнесу [2, 3].

На основі навіть попереднього аналізу наукової літератури з вказаної проблематики варто однозначно відзначити, що існує постійний інтерес у наукових, політичних, маркетингових, фінансових та інших користувачів до прикладного вирішення окреслених нами вище завдань [4, 5].

Результати. На наш погляд, в організації обробки та використанні даних відкритих інформаційних джерел у процесі маркетингових дослідженнях важливо відокремити ключові блоки робочого алгоритму, а також розуміти внутрішні його взаємозв'язки. Особливу увагу при цьому слід приділити обережному та конструктивному використанню значних за обсягом і, досить часто, невисоких за якістю масивів відкритою інформації, доступність якої зростає з кожним роком. Окреслимо основні положення пропонованого авторами даної публікації підходу, який за результатами чисельних апробацій доведено продуктивний щодо вищезазначеного.

З комерційної точки зору будь-який проект починається з ідеї, яка має якийсь бізнесове значення, тобто може бути реалізована для задоволення потреб споживачів, створення доданої вартості та отримання певних вигід економічного характеру (наприклад, прибутку). Уявлення про ідею бізнес розвивається відповідно до досвіду виконавця (реалізатора) та середовища її втілення. Таким чином, на наш погляд, початок обробки інформації з відкритих джерел лежить у створенні базового розуміння цілей та мети всієї цієї активності. У вже працюючому бізнесі, уявлення про ідею також змінюються відповідно до ринкової динаміки, структури власності, джерел фінансування тощо. Але для нашої роботи важливим є тільки той факт, що ми замислюється над реалізацією тої чи іншої ідеї, а також можемо

її якимось чином сформулювати. Звичайно ж початкові уявлення про проект, або картинка з бізнес-плану, майже ніколи лишається незмінними, а навпаки вони мають розвиватися. Для аналітичної роботи важливо тільки відштовхнутися від певного базового розуміння сфери діяльності, потенційних споживачів чи інших фундаментальних аспектів.

Відповідно до сформованої ідеї (бізнес-ідеї) потрібно визначитися з певними критеріями, доступними можливостями чи обмеженнями в її реалізації. Відзначимо, що структурований підхід на цьому етапі також буде складно реалізувати, але відповіді на питання «що є для нас прийнятним?» є необхідним етапом подальшого забезпечення якості аналізу інформації з відкритих джерел.

Сформоване розуміння ситуації, своїх бажань та обмежень варто формалізувати в маркетингових цілях, які передбачається реалізувати на основі аналізу інформації з відкритих джерел. Зазвичай для поставки цілей в підручниках вказуються вимоги щодо часової визначності цілей, їх конкретності формулювання, ресурсного забезпечення та інших. З позицій нашого власного досвіду таких структурований підхід можливий на етапі коли вже є достатньо інформації, проведені додаткові роботи щодо аналізу ресурсів компанії. На початкових етапах теоретичні вимоги щодо визначеності та ресурсної забезпеченості можна реалізувати тільки дуже приблизно. Тому для нашого дослідження важливо виставити широку систему цільових індикаторів чи можливо цільових параметрів, які відповідно до нашого уявлення відтворюють погляд ключові суттєві аспекти його практичного втілення.

Наступним блоком нашого підходу слід вважати розробку (розвиток) розуміння цілей. Роботи за таким напрямом ми назвали на схемі створення «тезаурусу цілей», або іншими словами формалізації власного уявлення про цілі у певних термінах чи поняттях. З позицій маркетингового дослідження цей блок важливо реалізовувати не тільки на основі внутрішньо корпоративної роботи, але й у прив'язці до поведінки споживача. Наприклад, розуміння лояльності до торгової марки може суттєво відрізнятись у компаній, у споживачів, за регіонами, відповідно до ситуації в країні. Звідси опис цілі збільшити лояльність клієнтів до марки компанії треба «розшифрувати» у додаткових ознаках: описати про яких клієнтів йде мова, на яку перспективу, яка оцінка цієї цілі нас задовольняє, у цільовій аудиторії чи на суміжних сегментах і т.д. Описати поняття «лояльність» у різних синонімах та визначеннях прийнятних для компанії та споживачів товарів. Така робота у результаті має призвести для створення списку ключових інформаційних одиниць, на дослідження яких будуть спрямовані зусилля на наступних етапах.

Відповідно до розуміння цілей та сформованого тезаурусу для нашого аналізу, слід приділити увагу де ми можемо отримати необхідну інформацію, де знаходяться потрібні нам дані та у якому вигляді вони там існують. Блок «селекція інформаційних каналів» наповнений роботами з підбору найбільш якісних чи достовірних джерел інформації. Хотіли б також відзначити, що мова йде не тільки про якість інформації в тому чи іншому каналі, але й про необхідне ресурсне забезпечення чи її доступність. Сьогодні перелік інформаційних агенцій, платформ та агрегаторів новин, різноманітність соціальних мереж та форматів представ-

лення там інформації однозначно піднімає питання чи вигідно обробляти весь цей потік. Тільки частина цього колосального інформаційного потоку матиме суттєве значення для вирішення завдання, а досить значна частина генерується автоматично, перерозподіляється з інших каналів (перепости, дублювання новин та інше). Доступ до частини інформаційних джерел вимагатиме певних витрат ресурсів (оплати підписок чи аккаунтів). Отже, відбір інформаційних каналів під вирішення конкретного маркетингового чи комерційного завдання дійсно стає арифметичною головоломкою – яким чином досягнути максимальної якості інформації за мінімальних витрат, у тому числі і витрат часу.

Наступним блоком є робота з побудови каталогу інформаційних джерел, що по суті представляє створення та дотримання певної логіки структурування всіх доступних чи наявних каналів інформації. Перш за все, структурований підхід, або так к ми назвали на схемі «каталог», дозволить забезпечити відслідковування інформаційної одиниці до джерела її входу в корпоративну систему моніторингу. З іншого боку, групування інформаційних каналів може бути досить корисним інструментом вирішення маркетингових завдань на наступних етапах. Наприклад, при використанні регіональної ознаки до групування дозволить відслідковувати якість комунікацій в такому розрізі і відповідним чином планувати рекламний бюджет компанії. Додатково слід розуміти, що однозначно «правильного» чи «неправильного» рішення не існує – групування визначатиметься завданнями інформаційного аналізу у маркетинговому процесі.

Наступні два блоки на нашій схемі акцентують увагу на необхідність первинної підготовки інформації – визначення способів обробки та атрибуція даних. Фактично на цих етапах відбувається перевірка консистентності інформації в каналі, очищення даних, приведення різнорідних за природою даних до одного визначеного цифрового формату. З технічної точки зору дослідники даних та аналітики розробляють систему фільтрів чи процедур підготовки бази даних (пре-процесинг даних). Метою всіх цих процедур виступає максимально якісне наповнення бази даних інформацією, без помилок, без спотворень та відповідно до інших критеріїв. Якщо говорити про вирішення одного окремого завдання аналізу інформації в при розробці певного маркетингового рішення, то його вагомість та значення для компанії мають бути дійсно вагомими, так як якісна підготовка інформації досить часто вимагає значних трудовитрат. Коли ж мова йде про налаштування системи аналізу інформації з відкритих джерел на тривалий період часу, тобто відслідковування проблематики на регулярній основі, то якісна розробка моделі підготовки даних на початку позбавить значної частини ризиків у подальшому. Відзначимо також, що досить часто у бізнесмена може не бути достатньо часу, ресурсів чи знань для організації якісної підготовки інформації, а тому ми б рекомендували концентруватися не стільки на якості отриманих з відкритих джерел даних, скільки на правильному розумінні обмежень чи недоліків, які недопрацьовані на момент прийняття рішення. Такий підхід є більш гнучким і у багатьох випадках є практично придатним, за умови що висока точність результатів не вимагається, а швидкість є основним чинником забезпечення конкурентоспроможності.

Після завершення аналітичних робіт сфокусованих на інформаційних каналах і підготовці даних, слід перейти до формування логіки та алгоритмів зберігання та накопичення даних, створення процедур перевірки та контролю інформації в корпоративній базі даних. Фактично тут відбувається вже втілення рішень на попередніх етапах, їх розвиток у системі накопичення цільової корпоративної інформації. У найбільш простому форматі, якщо було прийнято рішення слідкувати, наприклад, за змінами компанії конкурента, було організовано доступ до необхідної інформації, проведено її підготовка і вона внесена у «якусь табличку» в компанії, то залишається питання щодо її придатності для подальшого використання. Досить часто аналітику стає легше знайти інформацію поза межами компанії, ніж оперативно отримати доступ до корпоративної бази даних з дотриманням всіх необхідних процедур. Тут варто згадати атрибуцію інформації і її представлення – відповідно до сформованого тезаурусу і каталогу каналів.

Пройшовши етап запуску інформаційного потоку і базу даних, тестування роботи самої бази, уточнення чи коригування атрибуції даних в ній, наступним кроком доцільно сфокусуватися на завданнях аналізу. Всі попередні етапи у тій чи іншій мірі мають технічний характер, тобто забезпечують якісну організацію отримання, але ніяким чином не дають відповіді на завдання аналізу даних з відкритих джерел. Таким чином, важливо на основі тезаурусу цілей, розробленої атрибуції даних слід перейти до підготовки списку ключових слів, формування семантичних одиниць та формування семантичного ядра аналізу. Важливо виділити і визначити різноманітність та достатність представлення цілей у семантиці атрибутованих даних. Зазвичай такого роду роботи проводяться спеціалістами з оптимізації пошукових систем в Інтернет та просування сайтів у них, оптимізації контенту сайтів. Але дана технічна логіка має значення і при аналізі будь-якої інформаційної сукупності, так як дозволяє виділити та описати форми представлення семантичних одиниць в доступному інформаційному масиві. З практичної точки зору, в маркетингових дослідженнях товаром може вважатися не сам продукт, а комплекс додаткових послуг та сервісів необхідних для його споживання. Фактично «товар» у маркетинговому дослідженні, маючи декілька складових чи шарів опису, має бути виражений не тільки назвою продукту, але й списком додаткових семантичних формулювань, часто «сленгового» характеру. Таким чином, дослідження інформаційного потоку тільки на основі назви товару не буде містити комерційної цінності через відсутність повноти відображення сприйняття.

На базі виділених семантичних одиниць та розробленого ядра аналізу далі організовуються аналітичні операції на основі сформованої бази даних, проектується операції класифікації та структурного представлення інформаційної сукупності.

З використанням описаного вище підходу та відштовхуючись від сформованої бази інформаційних повідомлень щодо діяльності компанії Shell на ринках країн колишнього радянського простору за квітень-травень 2022 р. (біля 100 тисяч україномовних новин, понад 80 тисяч англомовних та 250 тисяч російськомовних) нами було відібрано для демонстрації прикладних аспектів підходу тестову вибірку з 41 публікації. Загальна описова статистика за вибіркою базується на то-

кенізації всіх інформаційних повідомлень та їх очищення, але з посиланнями на оригінальні публікації, представлена у табл. 1.

Таблиця 1

ТЕСТОВА ВИБІРКА ІНФОРМАЦІЙНИХ ПОВІДОМЛЕНЬ

| ID | Fake indicator | Word count | Character count | Average word length | LIX index | Title | Text | Source |
|------------|----------------|------------|-----------------|---------------------|------------|--------------------------|---------------------------|---|
| continuous | 0/1 | continuous | continuous | continuous | continuous | string | string | string |
| 1 | 0 | 126 | 732 | 5,8 | 53,2 | Бегство из России доро | Наибольшие потери от у | https://t.me/karalunny_account |
| 2 | 0 | 82 | 411 | 5,0 | 31,9 | Моторные масла для ав | «У брендов Idemitsu и Liq | https://t.me/RUSanctions/486 |
| 3 | 1 | 20 | 151 | 7,6 | 49,5 | Компания Shell собирае | Компания Shell собирает | https://t.me/karalunny_account |
| 4 | 1 | 22 | 153 | 7,0 | 43,3 | Shell в ближайшие дни | Shell в ближайшие дни в | https://t.me/vedomosti/2138 |
| 5 | 1 | 18 | 120 | 6,7 | 43,7 | Сеть заправок Shell вре | Сеть заправок Shell врем | https://t.me/rian_crimea/919 |
| 6 | 1 | 30 | 219 | 7,3 | 58,2 | Работа заправок Shell в | Работа заправок Shell в Р | https://t.me/varlamov_news/ |
| 7 | 0 | 20 | 153 | 7,7 | 57,8 | Shell приняла решение | Покупателя сотрудникам | https://t.me/moscowtoplive/4 |
| 8 | 0 | 865 | 4917 | 5,7 | 54,4 | «Крутое пике». В России | Что происходит состояни | http://rkm.kiev.ua/v-mire/211 |
| 9 | 1 | 130 | 816 | 6,3 | 59,4 | «Газета.Ру» следит за с | Первый за сегодняшний | https://t.me/gazetaru/5053 |
| 10 | 0 | 95 | 548 | 5,8 | 44,2 | Сети АЗС в Украине пов | В период с 5 по 6 мая сет | https://dengi.ua/finance/6295 |
| 11 | 1 | 16 | 129 | 8,1 | 55,0 | Нефтегазовая компания | Покупателя сотрудникам | https://t.me/crimea_chp/2065 |
| 12 | 1 | 110 | 711 | 6,5 | 61,5 | Военная операция Росс | Санкции ЕС в отношении | https://t.me/gazetaru/5038 |
| 13 | 0 | 147 | 912 | 6,2 | 54,5 | Компания Shell продает | Нефтегазовая компания | https://dengi.ua/business/629 |
| 14 | 1 | 80 | 482 | 6,0 | 50,9 | Газ из России невозмож | Для увеличения поставо | https://t.me/talksrussia/1862 |
| 15 | 0 | 270 | 1650 | 6,1 | 56,2 | Нафтогігант Shell продає | Нафтова компанія Shell з | https://www.epravda.com.ua |
| 16 | 0 | 7 | 69 | 9,9 | 51,0 | Shell официально вых | Руководство и техническ | https://t.me/novosti_voyna/5 |
| 17 | 1 | 629 | 4128 | 6,6 | 68,7 | Ушли чтобы остаться: з | В след за объявлением С | https://privet-rostov.ru/gorod |
| 18 | 0 | 212 | 1312 | 6,2 | 60,9 | Сотрудники московской | Сотрудники московской | https://t.me/bazabazon/1148 |
| 19 | 1 | 179 | 1170 | 6,5 | 64,2 | Нефтяной гигант Shell п | Европейский нефтегазов | http://rkm.kiev.ua/obschestvo |
| 20 | 1 | 193 | 1190 | 6,2 | 60,5 | Западным компаниям т | Западным энергетически | https://t.me/karalunny_account |
| 21 | 1 | 135 | 820 | 6,1 | 51,5 | Индийские компании г | Место западного бизнес | https://t.me/emphasises/4807 |
| 22 | 1 | 508 | 2782 | 5,5 | 52,8 | Хлопают дверью и теря | Зарубежные компании, с | https://t.me/stalin_gulag/3575 |
| 23 | 0 | 13 | 130 | 10,0 | 43,8 | Shell больше не будет п | Нефтегигант Shell больше | https://minfin.com.ua/2022/0 |
| 24 | 0 | 13 | 128 | 9,8 | 53,4 | Shell відмовилася від су | Нафтогігант Shell більше | http://www.epravda.com.ua/ |
| 25 | 1 | 85 | 562 | 6,6 | 62,2 | Россия – Украина / Что | «Экспорт ископаемого т | https://vot-tak.tv/provekt/ros |
| 26 | 1 | 119 | 728 | 6,1 | 59,5 | The Guardian: с начала | За два месяца войны Рос | https://vot-tak.tv/novosti/28 |
| 27 | 0 | 860 | 4996 | 5,8 | 53,3 | «Крутое пике». Россия с | Что происходит состояни | http://rkm.kiev.ua/v-mire/211 |
| 28 | 0 | 22 | 200 | 9,1 | 59,8 | Япония занепокоена пер | Переговоры британськог | http://www.epravda.com.ua/ |
| 29 | 0 | 93 | 494 | 5,3 | 51,9 | Компания Shell присту | Компания Shell присту | https://t.me/stalin_gulag/3508 |
| 30 | 0 | 27 | 206 | 7,6 | 65,0 | Британско-нидерландс | Британско-нидерландск | https://t.me/tele_eve/4461 |
| 31 | 1 | 108 | 713 | 6,6 | 59,4 | Спецоперация ВС Росси | Спецоперация ВС Росси | https://t.me/gazetaru/4758 |
| 32 | 0 | 282 | 1529 | 5,4 | 45,4 | В какое время лучше е | В какое время суток опре | http://e-news.com.ua/show/5 |
| 33 | 0 | 297 | 1864 | 6,3 | 60,5 | 5 лучших продуктов для | Существуют особые прод | http://e-news.com.ua/show/5 |
| 34 | 0 | 158 | 1014 | 6,4 | 63,2 | В УКРАИНЕ ОГРАНИЧИЛ | Кабинет министров прин | https://dilova.com.ua/ukraina |
| 35 | 0 | 195 | 1164 | 6,0 | 51,5 | Водка, хлеб, консервы" | Российские чиновники о | http://internetua.com/vodka-l |
| 36 | 0 | 183 | 1110 | 6,1 | 60,2 | Керівники і техперсона | Енергетичний гігант Shell | https://www.epravda.com.ua |
| 37 | 0 | 252 | 1527 | 6,1 | 59,9 | Shell больше не будет п | Нефтегигант Shell больше | https://minfin.com.ua/2022/0 |
| 38 | 0 | 197 | 1326 | 6,7 | 68,2 | Компания Nikola гарант | В марте конкурирующая | http://internetua.com/kompa |
| 39 | 0 | 287 | 2030 | 7,1 | 67,1 | В апреле количество ос | По состоянию на конец а | https://bin.ua/news/economic |
| 40 | 0 | 148 | 1042 | 7,0 | 63,6 | Правительство РФ сост | Советник мэра Мариупол | https://ukrudprom.com/news |
| 41 | 1 | 189 | 1215 | 6,4 | 65,6 | Жалобы игнорируют, а | Рассказываем, почему ук | https://economics.segodnya.u |

Джерело: побудовано авторами.

На основі такої вибірки було проведено стандартні процедури очищення тексту, підготовки бази даних, як це було описано вище у нашій роботі. Наступним

кроком було зроблено підготовку бази даних для навчання різних моделей ідентифікації фейкових повідомлень. Технічно було введено категоріальну змінну-маркер фейковості новини, на базі якої було підготовлено масив для використання різних алгоритмів машинного навчання:

- нейромережі (neural network);
- логістичної моделі (logistic model);
- моделі «випадковий ліс» (random forest);
- «деревовидної» моделі (tree model).

Кожен з вказаних алгоритмів машинного навчання не є новітньою розробкою і використовується у різних галузях знань вже досить успішно тривалий період часу [5, 6]. Для проведення технічних розрахунків та налаштування аналізу було використано найбільш стандартні налаштування і спеціалізоване програмне забезпечення для аналізу даних Orange Data Mining Software. Відзначимо, що вказаний програмний продукт розроблений і розповсюджується на безкоштовній основі, а тому підходить нам для використання у академічних цілях без порушення будь-яких авторських прав.

Логіка нашого подальшого аналізу інформації з відкритих джерел полягає у наступному:

- 1) підготувавши базу даних спробувати навчити різні моделі машинного навчання на навчальній вибірці;
- 2) обрати одну чи декілька моделей машинного навчання для ідентифікацій (оцінювання) «фейковості» інформаційних повідомлень;
- 3) перевірити якість отриманих результатів на основі співставлення отриманих прогнозних оцінок з фактичним маркером достовірності повідомлення;
- 4) використати обрані моделі для обробки інформаційного потоку повідомлень з відкритих джерел у наступних маркетингових дослідженнях.

Для кращого розуміння відмінностей у різних моделях машинного навчання нами візуалізовано тільки результати двох з чотирьох застосованих моделей – «деревовидної» моделі і моделі «випадкового лісу» (рис. 1).

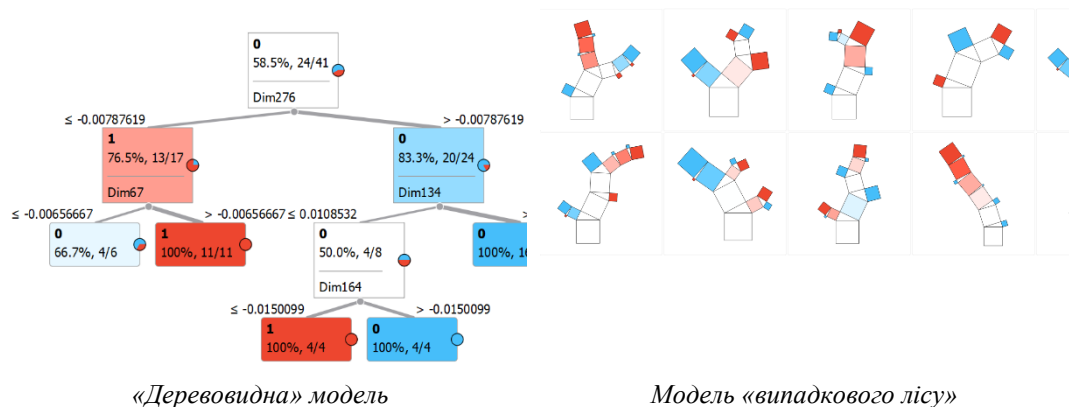


Рисунок 1. Візуалізація результатів деревовидної моделі та моделі «випадкового лісу» оцінювання інформаційних повідомлень про компанію Shell у сформованій тестовій вибірці з відкритих джерел

Джерело: побудовано авторами.

Інструментом співставлення результатів різних моделей у нашому прикладі виступила матриця розбіжностей (confusion matrix) (рис. 2).

| | | Predicted | | Σ |
|--------|----------|-----------|----|----------|
| | | 0 | 1 | |
| Actual | 0 | 24 | 0 | 24 |
| | 1 | 2 | 15 | 17 |
| | Σ | 26 | 15 | 41 |

А) «Деревовидна» модель

| | | Predicted | | Σ |
|--------|----------|-----------|----|----------|
| | | 0 | 1 | |
| Actual | 0 | 24 | 0 | 24 |
| | 1 | 0 | 17 | 17 |
| | Σ | 24 | 17 | 41 |

Б) Модель «випадкового лісу»

| | | Predicted | | Σ |
|--------|----------|-----------|---|----------|
| | | 0 | 1 | |
| Actual | 0 | 20 | 4 | 24 |
| | 1 | 13 | 4 | 17 |
| | Σ | 33 | 8 | 41 |

В) Неймережа

| | | Predicted | | Σ |
|--------|----------|-----------|----|----------|
| | | 0 | 1 | |
| Actual | 0 | 24 | 0 | 24 |
| | 1 | 0 | 17 | 17 |
| | Σ | 24 | 17 | 41 |

Г) Логістична модель

Рисунок 2. Матриці розбіжностей оцінювання «фейковості» інформаційних повідомлень у тестовій вибірці новин з відкритих джерел щодо діяльності компанії Shell

Джерело: побудовано авторами

Відзначимо, що виходячи з представлених матриць розбіжностей, на яких видно як було оцінено те чи інше повідомлення окремою моделлю навчання та його фактичною попередньою оцінкою, можна зробити висновок що дві з чотирьох справилися «на відмінно» – неймережа і модель «випадкового лісу». Обидві показали досить високий рівень точності оцінювання і фактичних значень маркеру достовірності («фейковості») повідомлень.

Висновки. За умов форс-мажорних ситуацій розвитку бізнесу, високого рівня мінливості параметрів і невизначеності перспектив кон'юнктури ринків зростає значущість нефінансових показників результативності діяльності компаній та інформації щодо впливових соціально-політичних й подібних чинників станів і змін

середовища. Особливої уваги в системах маркетингової аналітики заслуговують та набувають масиви даних з відкритих джерел інформації.

Донедавна нечувана доступність інформації для прийняття управлінських рішень у маркетингу та комерції відкриває безпрецедентні можливості для нарощування ефективності бізнесу. Проте технології машинного навчання в системах Big Data і Data Science у забезпеченні ефективності маркетингової та комерційної діяльності підприємств, окрім прогресу ототожнюють зростаючу складність середовища і процесів, розвиток актуальних для сучасної економіки феноменів, які потребують нагального конструктивного науково-практичного опрацювання.

Доводиться і акцентується увага маркетингових аналітиків та дослідників на необхідності:

- 1) перевірки достовірності інформаційних повідомлень з відкритих джерел перед їх включенням до процесів розробки управлінських рішень;
- 2) використання існуючого сучасного інструментарію машинного навчання та моделювання для спрощення рутинних операцій та економії ресурсів компанії.

Розроблено і апробовано авторські підходи щодо удосконалення аналітичних систем на базі відкритих джерел інформації, яке має за мету максимізацію ситуаційної обізнаності менеджменту та фахівців маркетингу й комерційної діяльності підприємств, потенційно забезпечує проактивне системне розуміння станів і змін бізнес-середовища на основі релевантних і достатньо достовірних даних.

Накопичення і використання досвіду відповідної аналітичної і управлінської роботи компаніями на цільових ринках, із застосуванням запропонованих підходів, зростаючих масштабів релевантних даних й результатів їх обробки - дозволять їм напрацювати максимально результативні прийоми забезпечення ситуаційної обізнаності менеджменту, підвищити ефективність бізнесу. У свою чергу, це створюватиме передумови й для прогресу подальших досліджень за напрямком, розвитку представлених вище наукових результатів.

Література

1. Sahoo Somya Ranjan, Gupta B. B. Multiple features based approach for automatic fake news detection on social networks using deep learning. Applied Soft Computing. 2021. Vol. 100, March. DOI: 10.1016/j.asoc.2020.106983
2. Yang Yang, Zheng Lei, Zhang Jiawei, Cui Qingcai, Zhang Xiaoming, Li Zhoujun, Yu Philip S. TI-CNN: Convolutional Neural Networks for Fake News Detection. arXiv:1806.00749v1 [cs.CL]. 2018. Vol. 3, June. URL: <https://arxiv.org/pdf/1806.00749.pdf>
3. Amer Eslam, Kwak Kyung-Sup, El-Sappagh Shaker. Context-Based Fake News Detection Model Relying on Deep Learning Models. Electronics. 2022. Vol. 11(8), April. DOI: 10.3390/electronics11081255
4. Kaliyar Rohit Kumar, Goswami Anurag, Narang Pratik, Sinha Soumendu. FNDNet – A deep convolutional neural network for fake news detection. Cognitive Systems Research. 2020. Vol. 61, June. P. 32-44. DOI: 10.1016/j.cogsys.2019.12.005.
5. Kaliyar Rohit Kumar, Goswami Anurag, Narang Pratik. 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> (дата зверення 29.04.2022)

6. Nasir Jamal Abdul, Khan Osama Subhani, Varlamis Iraklis. Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*. 2021. Vol. 1, Issue 1, April. DOI: 10.1016/j.jjimei.2020.100007.

References

1. Sahoo, Somya Ranjan, and B. B. Gupta. "Multiple features based approach for automatic fake news detection on social networks using deep learning". *Applied Soft Computing* 100 (2021). DOI: 10.1016/j.asoc.2020.106983.

2. Yang, Y., Zheng, L., Zhang, J., Cui, Q., Zhang, X., Li, Z., and Philip S. Yu. "TI-CNN: Convolutional Neural Networks for Fake News Detection". arXiv:1806.00749v1 [cs.CL] 3, (2018). <https://arxiv.org/pdf/1806.00749.pdf> (accessed 29.04.2022).

3. Amer, E., Kwak, K.-S., and S. El-Sappagh. "Context-Based Fake News Detection Model Relying on Deep Learning Models". *Electronics* 11, no. 8 (2022). DOI: 10.3390/electronics11081255.

4. Kaliyar, R. K., Goswami, A., Narang, P., and S. Sinha. "FNDNet – A deep convolutional neural network for fake news detection". *Cognitive Systems Research* 61 (2020): 32-44. DOI: 10.1016/j.cogsys.2019.12.005.

5. Kaliyar, R. K., Goswami, A., and P. A. Narang. "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 (2021): 1066-1072. DOI: 10.5220/0010316010661072.

<https://www.scitepress.org/Papers/2021/103160/103160.pdf> (accessed 29.04.2022).

6. Nasir, J. A., Subhani, K. O., and Iraklis Varlamis. "Fake news detection: A hybrid CNN-RNN based deep learning approach". *International Journal of Information Management Data Insights* 1, no. 1 (2021). DOI: 10.1016/j.jjimei.2020.100007.

PECULIARITIES OF USING OPEN DATA IN PREPARING MARKETING AND COMMERCIAL MANAGEMENT DECISIONS

Oleksii I. Oleksiuk

Doctor of Economic Sciences, Professor,
SHEI «Kyiv National Economic University
named after Vadym Hetman» (Ukraine)
ORCID 0000-0002-3407-6201

Oleksandr K. Shafalyuk

Doctor of Economic Sciences, Professor,
SHEI «Kyiv National Economic University
named after Vadym Hetman» (Ukraine)
ORCID 0000-0003-1145-7973

Abstract. The article examines the features of the use of data from open sources of information. special approaches and techniques in the svstem of marketing research of enterprises. Improving approaches to integration into

modern marketing information systems of modern content analysis tools have been developed and tested, which organically provides constructive synergy of applied methods and means of neuro-fuzzy modeling and clustering of information arrays, statistical analysis of information units including potential and actual consumers. The modern methodological tools of machine learning were tested on the basis of the author's approaches to detecting "fakes" (unreliable information) in the formation of situational awareness of company management, identifying trends in target markets, analytical processing of changes associated with high risks and uncertainties for business. 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 ("fake") messages. The recommendations on the organization of the formation of source data from open sources of information, improving the quality and reliability of their processing and successful integration of marketing and commercial analytics systems have been constructively developed.

Keywords: open data; reliability of information; machine learning; marketing; business; management.

Стаття надійшла до редакції 05.05.2022