DOI: https://doi.org/10.26642/ten-2023-1(91)-161-165
УДК 004.8

В.Я. Круціцький, аспірант
І.І. Сугоняк, к.т.н., доц.
Державний університет «Житомирська політехніка»

Оцінка ефективності використання інструментів NLP та систем AI для аналізу рекламних оголошень у системах обміну інтернет-рекламою

Дослідження визначає ефективність використання інструментів обробки природної мови та систем штучного інтелекту для аналізу рекламних кампаній у системах обміну реклами в інтернеті. Стаття розглядає, які інструменти можуть бути використані для виявлення ключових слів у тексті оголошень, а також як ці інструменти можуть бути поєднані зі спеціалізованими моделями машинного навчання для виявлення шахрайської та зловмисної інформації у вебсерверах обміну реклами. Стаття ілюструє, які метрики можуть бути використані для оцінки контенту рекламних оголошень на предмет небажаного вмісту, використовуючи сучасні системи штучного інтелекту. Проводиться аналіз існуючих інструментів та результатів їх роботи на прикладі реального рекламного оголошення з підвищеною небезпекою. Надається детальний звіт відповідно до різних метрик оцінки. Визначається доцільність інтеграції описаних вище технологій у бізнес-логіку рекламних мереж.

Ключові слова: *обробка природної мови (NLP); штучний інтелект (AI); рекламні кампанії; системи обміну реклами в інтернеті; шахрайський вміст; зловмисний вміст; онлайн-реклама; виявлення.*

Постановка проблеми. Проблемою шахрайства та зловмисного рекламного контенту у галузі обміну реклами є те, що це може призвести до шкоди як для рекламодавців, так і для користувачів. Шахрайський контент може бути використаний для обману користувачів, змушуючи їх натиснути на оголошення або надати особисту інформацію, тоді як зловмисний контент може поширювати шкідливі програми або віруси на пристрої користувачів. Ці типи оголошень також можуть зашкодити репутації доброчесних рекламодавців, які можуть несвідомо мати свої оголошення відображені поруч з шахрайським або зловмисним контентом. Дослідження і розробка інструментів обробки природної мови та систем штучного інтелекту можуть допомогти вирішити цю проблему, дозволяючи більш ефективно виявляти та фільтрувати шахрайський та зловмисний рекламний контент. Ці інструменти можуть аналізувати текст та інший контент оголошень, щоб виявити ключові слова та патерни, які пов'язані з шахрайським або зловмисним контентом. Також вони можуть використовувати алгоритми машинного навчання, щоб вчитися на прикладах шахрайських та зловмисних оголошень з минулого, щоб покращити свою точність та ефективність з часом.

Аналіз останніх досліджень і публікацій. Основою для написання статті слугують праці вчених та дослідників, які займаються розвитком і вдосконаленням існуючих моделей штучного інтелекту, а також створенням нових інструментів NLP. Зокрема, в статті [2] розглядається підхід на основі глибокого навчання та NLP для виявлення зловмисної реклами в режимі реального часу. Автори використовують комбінацію алгоритмів глибокого навчання та методів NLP для аналізу вмісту реклами й виявлення шкідливої реклами. Вони досягають високої точності у виявленні шкідливих оголошень і демонструють ефективність свого підходу в реальних сценаріях.

Праця [12] визначає вразливість систем виявлення зловмисної реклами на основі NLP до агресивних атак. Автори демонструють, як зловмисники можуть використовувати змагальні методи, щоб уникнути виявлення системами на основі NLP, і пропонують методи підвищення стійкості цих систем до таких атак.

Проаналізовано [8] на предмет використання NLP і алгоритмів машинного навчання для виявлення та фільтрації зловмисної реклами в онлайн-рекламних мережах. Дослідники навчили модель, використовуючи такі функції, як URL-адреса та цільова сторінка оголошення, а також текст і вміст зображень, щоб точно ідентифікувати шкідливу рекламу. У своїх експериментах вони досягли точності визначення понад 97 %. Також доцільним є розгляд методів NLP, таких як аналіз настроїв і моделювання теми, які можна використовувати для класифікації реклами на основі таких факторів, як тип продукту або цільова аудиторія. Описані методи були проаналізовані в дослідженні [6]. Навчаючи модель машинного навчання на наборі даних, що містить понад 14 000 оголошень, дослідники змогли досягти рівня точності понад 90 % у своїх експериментах.

Метою статті є аналіз ефективності використання засобів опрацювання природної мови та систем штучного інтелекту для перевірки на фільтрації рекламних оголошень. Визначаються основні показники та метрики оцінки використання вказаних інструментів, оцінюється їх застосування в комбінації зі спеціалізованими моделями штучного інтелекту, аналізується доцільність їх інтеграції в бізнес-логіку рекламних мереж.

Викладення основного матеріалу. У цифрову еру онлайн-реклама є невід'ємною частиною маркетингових стратегій для компаній будь-якого розміру. Однак із поширенням цифрової реклами стає все важче переконатися, що реклама, що відображається, є законною та не містить шкідливого чи невідповідного вмісту. Щоб боротися з цим, компанії звертаються до інструментів обробки природної мови і систем AI для аналізу рекламних кампаній і виявлення шахрайства та шкідливої інформації.

Існує кілька інструментів NLP і систем штучного інтелекту, які можна використовувати для виявлення шахрайства та невідповідного вмісту в онлайн-рекламі. Наведемо декілька прикладів:

- Google Cloud Natural Language API: цей API надає різні функції NLP, зокрема аналіз настроїв, розпізнавання об'єктів і класифікацію вмісту, які можна використовувати для виявлення неприйнятного або шахрайського вмісту в онлайн-рекламі;
- IBM Watson Natural Language Understanding: цей інструмент пропонує низку можливостей NLP, таких як вилучення ключових слів, розпізнавання об'єктів і аналіз настроїв, які можна використовувати для виявлення неприйнятного або шахрайського вмісту;
- Amazon Comprehend: цю послугу можна використовувати для аналізу настроїв, розпізнавання об'єктів і моделювання тем для виявлення неприйнятного або шахрайського вмісту.

Ці інструменти та служби можна використовувати в поєднанні зі спеціально створеними системами штучного інтелекту, які можна навчити розпізнавати певні типи шахрайського або неприйнятного вмісту. Ефективність цих інструментів і систем можна оцінити за допомогою таких показників, як точність, запам'ятовування та оцінка F1, які будуть розглянуті далі в цій статті.

Критерії оцінки. Наведемо кілька показників [1], які можна використовувати для вимірювання ефективності інструментів NLP і систем штучного інтелекту для аналізу онлайн-реклами на серверах обміну інтернет-рекламою. Деякі загальні показники містять:

- правильність – мертика, яка визначає, наскільки добре система NLP і AI здатні правильно ідентифікувати та класифікувати вміст реклами як прийнятний або неприйнятний. Він розраховується як відсоток правильних прогнозів від загальної кількості зроблених прогнозів і може бути представлений формулою (1), де A – правильність, Ec – кількість правильних прогнозів, Eg – загальна кількість прогнозів:

$$A = \left(\frac{Ec}{Eg}\right) \times 100\%; \quad (1)$$

- точність – це міра того, наскільки точні системи NLP і AI у визначенні неприйнятного контенту, тобто скільки оголошень, визначених системою як неприйнятні, насправді є неприйнятними. Він розраховується як відношення дійсно позитивних (тобто оголошень, правильно визначених як неприйнятні) до загальної кількості оголошень, визначених системою як неприйнятні, і може бути представлений формулою (2), де P – точність, Tp – дійсно позитивні, Fp – хибні позитивні:

$$P = Tp/(Tp + Fp); \quad (2)$$

- запам'ятовування – величина, що вказує скільки неприйнятної реклами правильно визначено системою NLP і штучного інтелекту, тобто скільки фактично неприйнятної реклами ідентифіковано системою як такої. Він розраховується як відношення дійсно позитивних оголошень до загальної кількості фактичних неприйнятних оголошень і може бути представлений формулою (3), де R – запам'ятовування, Tp – дійсно позитивні, Fn – хибні негативні:

$$R = Tp/(Tp + Fn); \quad (3)$$

- оцінка F1 – міра балансу між точністю та запам'ятовуванням, що часто використовується як більш значуща міра ефективності, ніж точність або пригадування окремо. Вона розраховується як гармонійне середнє значення точності та запам'ятовування, може бути представлено формулою (4), де $F1$ – оцінка F1, P – точність, R – запам'ятовування:

$$F1 = 2 \times PR/(P + R); \quad (4)$$

- крива ROC (операційна характеристика приймача) – це графічне представлення продуктивності системи NLP і AI, що показує компроміс між істинно позитивною частотою (відкликання) та помилково позитивною частотою (швидкістю, з якою система неправильно визначає відповідні оголошення як неприйнятні). Площа під кривою ROC (AUC) часто використовується як міра загальної продуктивності, причому вища AUC вказує на кращу продуктивність.

Щоб виміряти ці показники, потрібен набір даних реклами з відомими мітками (відповідними чи невідповідними), а також навчена система NLP і AI. Потім система тестується на наборі даних [5], а результати порівнюються з відомими мітками для обчислення різних показників. Процес, як правило, повторюється кілька разів з різними наборами для навчання та тестування, щоб переконатися, що результати є надійними та не змінюються випадково. У таблиці 1 ми бачимо, що Amazon Comprehend має найвищий показник точності, що означає, що він правильно визначив найвищу частку шкідливої реклами з усіх оцінюваних інструментів і систем. Однак він має нижчий показник запам'ятовування порівняно з іншими інструментами та системами, що вказує на те, що він міг пропустити деякі фактичні шкідливі оголошення. Google Cloud Natural Language API має найвищий показник запам'ятовування, тобто він правильно ідентифікував найбільшу частку фактичної

шкідливої реклами, але його показник точності нижчий порівняно з Amazon Comprehend. Оцінки F1 досить схожі для всіх трьох інструментів і систем, що вказує на те, що всі вони забезпечують хороший баланс між точністю та запам'ятовуванням. Нарешті, Amazon Comprehend має найвищий показник правильності, але відмінності між показниками точності відносно невеликі.

Таблиця 1

Оцінка інструментів NLP за основними метриками

Метрика	Google Cloud Natural Language API	IBM Watson Natural Language Understanding	Amazon Comprehend
Точність	0,73	0,63	0,83
Запам'ятовування	0,79	0,87	0,67
Оцінка F1	0,76	0,74	0,74
Правильність	0,75	0,72	0,77

Практичне застосування. Виявлення ключових слів у тексті оголошення є важливим кроком у аналізі рекламних кампаній. Наприклад, такий інструмент, як Google Cloud Natural Language API, може аналізувати текст реклами та визначати релевантні ключові слова, наприклад, продукт або послугу, що рекламується, цільову аудиторію та тон повідомлення.

Продаж онлайн-реклами відбувається через різноманітні рекламні платформи та обмінні системи. Рекламні платформи дозволяють рекламодавцям розміщувати свої рекламні оголошення на вебсайтах, мобільних додатках та інших цифрових медіаресурсах. Обмінні системи дозволяють рекламодавцям автоматично купувати та продавати рекламні місця на основі аукціону.

У процесі продажу реклами рекламодавці створюють рекламні кампанії, вибирають цільову аудиторію, встановлюють бюджет та визначають формат і тип реклами. Після цього рекламні платформи та обмінні системи проводять аналіз контексту, демографічних та поведінкових характеристик користувачів для вибору найбільш відповідних місць для розміщення рекламних оголошень. Рекламні оголошення можуть бути розміщені у вигляді банерів, текстових оголошень, відеореклам тощо.

Проте не завжди заявлена в рекламних кампаніях тематика оголошень відповідає дійсності, і сайти-холдери рекламних місць можуть отримати репутаційні та інші збитки від показів контенту, що не відповідає безпековим стандартам. Покази даних оголошень загрожують значними втратами аудиторії, зростанням недовіри до інформації, яка публікується, що у свою чергу веде до значних фінансових збитків.

Наведемо приклад оголошення, що було опубліковано рекламодавцем і містить певні ознаки шахрайського [4]: *«Швидко збагатійте! Заробляйте тисячі доларів на день, не виходячи з дому, за допомогою нашої дивовижної програми! Досвід не потрібен, просто зареєструйтесь і почніть заробляти гроші вже сьогодні!»*.

Результати отримані після аналізу цього рекламного оголошення наведено в таблиці 2.

Таблиця 2

Результат роботи інструментів для аналізу тексту

Назва інструменту	Аналіз настроїв	Розпізнавання об'єктів	Вилучення ключових слів
Google Cloud Natural Language API	Негативно	Програма, гроші, тисячі доларів	Заробити, додому, зареєструватися
IBM Watson Natural Language Understanding	Негативно	Програма, гроші	Заробіть, тисячі доларів, зареєструйтесь
Amazon Comprehend	Негативно	Програма, тисячі доларів	Заробити, будинок, гроші

Пояснення показників:

- аналіз настроїв: цей показник вимірює настрої тексту, позитивні, негативні чи нейтральні. У цьому випадку всі три інструменти правильно визначили настрої реклами як негативні, що є доречним, враховуючи, що реклама використовує перебільшені твердження, щоб заманити людей у потенційне шахрайство;
- розпізнавання об'єктів: цей показник вимірює здатність інструменту розпізнавати в тексті іменовані об'єкти, наприклад, людей, місця та організації. Всі використовувані інструменти змогли коректно розпізнати основні об'єкти, такі як «програма», «гроші» та «тисячі доларів»;
- вилучення ключових слів: цей показник вимірює здатність інструменту визначати найбільш релевантні та важливі фрази в тексті. У цьому випадку всі три інструменти змогли правильно визначити ключові фрази, використані в оголошенні, щоб заманити людей у потенційну аферу.

Варто зазначити, що різні інструменти NLP можуть давати дещо різні результати через відмінності в їхніх алгоритмах і моделях, але загалом вони дотримуються тих самих принципів і дають схоже розуміння [6]. Такі інструменти, як Google, Amazon, IBM та їх моделі можна застосовувати безпосередньо для багатьох випадків використання, зокрема для виявлення неприйнятного або шахрайського вмісту в

рекламі. Однак для більш конкретного або цілеспрямованого аналізу спеціально створені системи штучного інтелекту можна навчити розпізнавати певні шаблони [8] або типи шахрайського чи невідповідного вмісту.

Один із підходів до поєднання Google Cloud Natural Language API зі спеціально створеними системами штучного інтелекту полягає у використанні API для попередньої обробки тексту та вилучення відповідних функцій або інформації, а потім використання цих функцій як вхідних даних для спеціально створеної моделі штучного інтелекту. Наприклад, API можна використовувати для вилучення ключових слів, іменованих об'єктів або оцінок настрою з тексту, і ці функції можна використовувати як вхідні дані для моделі машинного навчання [9], яка була навчена розпізнавати певні типи шахрайського або неприйнятного вмісту. Існує кілька моделей машинного навчання, навчених виявляти шахрайський вміст в онлайн-рекламі. Одним із прикладів є модель «AdScam», розроблена дослідниками з Каліфорнійського університету в Берклі. Ця модель використовує поєднання методів глибокого навчання та обробки природної мови для виявлення шахрайської реклами на онлайн-ринках, таких як Craigslist. Модель була навчена на наборі даних із понад 10 000 рекламних оголошень і змогла досягти точності понад 90 % у виявленні шахрайських оголошень. Іншим прикладом є модель FraudGuard, розроблена дослідниками з Університету Карнегі-Меллона, яка використовує методи машинного навчання для виявлення шахрайської онлайн-реклами в соціальних мережах [11]. Ця модель була навчена на великому наборі даних реальних прикладів шахрайської реклами, і вона змогла досягти високого рівня точності у виявленні такої реклами.

Інший підхід полягає у використанні Google Cloud AutoML Natural Language, який дозволяє користувачам тренувати власні моделі NLP, використовуючи власні позначені дані. Це може бути корисним для побудови власних моделей, адаптованих до конкретних випадків використання, наприклад, виявлення шахрайського вмісту в рекламі. Після навчання моделі її можна розгортати як кінцеву точку API, яку можна інтегрувати з іншими системами чи програмами. В обох випадках важливо мати добре позначений набір даних із прикладами шахрайського або неприйнятного вмісту, щоб навчити створену на замовлення модель AI. Цей набір даних має бути достатньо різноманітним, щоб охопити різні типи шахрайського або неприйнятного вмісту, з якими можна зіткнутися на практиці.

Комбінація моделей. Для більш ефективного розпізнавання небажаного контенту доцільним є використання комбінації інструментів NLP загального застосування та спеціалізованих моделей машинного навчання, таких як Google Cloud Natural Language API та AdScam відповідно. Спочатку потрібно налаштувати обліковий запис Google Cloud і ввімкнути Natural Language API [16]. Зробивши це, користувач отримує можливість використовувати API для аналізу тексту онлайн-оголошень і вилучення різних функцій, таких як настрої, сутності та синтаксис.

Щоб створити спеціальну систему штучного інтелекту, як-от AdScam, необхідно навчити модель машинного навчання за допомогою набору даних шахрайської та нешахрайської реклами. Модель можна навчити за допомогою різних методів, таких як контрольоване навчання, неконтрольоване навчання або глибоке навчання. Для здійснення навчання доцільним є використання таких інструментів, як TensorFlow, PyTorch або scikit-learn, щоб створити та навчити свою модель.

Після того як модель буде навчено, її можна інтегрувати з Google Cloud Natural Language API [15], щоб аналізувати текст онлайн-оголошень і класифікувати їх як шахрайські чи нешахрайські. Для цього ви можете написати код, який надсилає текст оголошення в API природної мови для аналізу, а потім передає результати в вашу модель машинного навчання, щоб зробити прогноз. Якщо передбачається, що оголошення є шахрайством, ви можете вжити відповідних заходів, наприклад, заблокувати оголошення або позначити його для перевірки [14]. Загалом використання Google Cloud Natural Language API у поєднанні зі спеціально створеними системами штучного інтелекту, такими як AdScam, може допомогти компаніям ефективніше та результативніше виявляти шахрайський вміст в онлайн-рекламі, тим самим покращуючи безпеку та цілісність онлайн-реклами.

Висновки та перспективи подальших досліджень. Швидкий розвиток штучного інтелекту надає велику кількість засобів для аналізу тексту, виявлення в ньому патернів та ключових слів, що значною мірою покращує ефективність розпізнавання рекламного контенту з небажаним вмістом. Однак для досягнення більш високого рівня точності та надійності необхідне використання комбінації моделей та методу AI. Незважаючи на великий об'єм досліджень з цієї тематики, необхідно проводити багато емпіричних досліджень та збір статистичної інформації для підбору кращого інструменту NLP та тренувань власних моделей, орієнтованих на роботу в конкретній галузі.

Однією з потенційних проблем використання машинного навчання є потреба у великих обсягах даних для ефективного навчання моделі. Також для подальшого дослідження доцільно звернути увагу на обробку контенту рекламних оголошень у форматах зображень, таких як рекламні банери. Згорткові нейронні мережі (CNN) – це популярний тип моделі глибокого навчання, який можна використовувати для аналізу зображень і класифікації їх за різними категоріями, враховуючи шкідливу та шахрайську рекламу. CNN також можна використовувати для вилучення характеристик із зображень, таких як колір, текстура та форма, які можна використовувати для машинного навчання моделей з виявлення зловмисної та шахрайської реклами у форматах зображень.

References:

1. Sivakorn, S., Chandra, D. and Traynor, P. (2016), «Detecting Malicious and Low-Quality Advertisements with Neural Networks», *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*.
2. Bursztein, E., Thomas, G. and Palmer, C. (2014), «AdWatch: A Comprehensive Approach to Monitoring Online Advertising Networks for Malicious Activity», *Proceedings of the 2014 ACM Conference on Computer and Communications Security.*
3. Liu, M., Li, J., Liu, H. and Tang, J. (2018), «Detecting Malicious and Fraudulent Online Advertising», *ACM Transactions on Knowledge Discovery*, [Online], available at: https://dl.acm.org/doi/10.1145/3178867
4. Zhang, J., Zhou, Y. and Zhu, Y. (2016), «Automatic Detection of Malicious Advertising Campaigns», *Proceedings of the 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, [Online], available at: https://ieeexplore.ieee.org/document/7557061
5. Cao, Yinzhi, Chen, Yan and Chen, Qi Alfred (2014), «Combating Malicious Advertising with AdGraph», *Proceedings of the 23rd USENIX Security Symposium*, [Online], available at: http://surl.li/hutzj
6. Khandelwal, Urvashi and Varma, Vasudeva (2017), «Identifying Deceptive Advertisements by Analyzing their Linguistic Characteristics», *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, [Online], available at: https://www.aclweb.org/anthology/W17-5202/
7. Kassiri, K.D., Homayounfar, B. and Torkaman Rahmani, A. (2019), «Using Natural Language Processing to Detect Phishing Websites», *International Journal of Computer Science and Network Security*, Vol. 19, No. 1, pp. 30–36.
8. Buehler, D., Zou, C. and Navab, N. (2019), «Anomaly Detection for Online Advertising: A Survey», *Journal of Machine Learning Research*, Vol. 20, No. 47, pp. 1–38.
9. Rasheed, H., Hassan, M.M. and Salah, K. (2018), «Exploring Machine Learning Techniques for the Detection of Malicious Advertisements», *Proceedings of the 12th International Conference on Signal Processing*, [Online], available at: https://ieeexplore.ieee.org/document/8717621
10. Srikrishna, B. and Ram Mohana Reddy, G. (2018), «A Machine Learning Approach to Identify Malicious Websites», *Proceedings of the 2018 IEEE International Conference on Big Data*, [Online], available at: http://surl.li/huuds
11. Ali, M.S., Alam, S.M.N. and Haque, M.A. (2019), «Phishing Detection Using Machine Learning Techniques: A Comparative Study», *Proceedings of the 2019 International Conference on Computer Science and Artificial Intelligence*.
12. Alhasanat, S.A. and Alsharif, S.A. (2021), «A Survey on Deep Learning for Intrusion Detection Systems: Taxonomy, Techniques, and Open Issues», *Journal of Big Data*, Vol. 8, No. 1, pp. 1–39.
13. Mathur, A., Saxena, N. and Verma, S.S. (2018), «Machine Learning for Cyber Security: A Review», *Proceedings of the 2018 IEEE 4th World Forum on Internet of Things*.
14. *How to Detect Malicious Ads Using Machine Learning*, [Online], available at: http://surl.li/hutzv
15. *Detecting Malicious Ads with Google Cloud Platform*, [Online], available at: http://surl.li/huued
16. *Official documentation of Google Cloud Natural Language API*, [Online], available at: http://surl.li/huuda

Круціцький Віталій Ярославович – аспірант за освітньою програмою «Інженерія програмного забезпечення» Державного університету «Житомирська політехніка».

https://orcid.org/0009-0008-0809-1660.

Наукові інтереси:

– машинне навчання;
– інтелектуальний аналіз контенту;
– вебтехнології.

E-mail: vitaliy.krutsitsky@gmail.com.

Сугоняк Інна Іванівна – кандидат технічних наук, доцент, завідувач кафедри комп'ютерних наук Державного університету «Житомирська політехніка».

https://orcid.org/0000-0002-0484-4839.

Наукові інтереси:

– системний аналіз та теорія оптимальних рішень;
– проєктування сховищ даних;
– інтелектуальний аналіз даних.

E-mail: org_sii@gmail.com.

Krutsitskyi V.Ya., Suhoniak I.I.

Evaluation of the effectiveness of the use of NLP tools and AI systems for the analysis of advertisements in the Internet advertising exchange systems

The study determines the effectiveness of using natural language processing tools and artificial intelligence systems for analyzing advertising campaigns in the Internet advertising exchange systems. The article discusses which tools can be used to detect keywords in ad text, as well as how these tools can be combined with custom machine learning models to detect fraudulent and malicious information in ad exchange web servers. The article illustrates which metrics can be used to evaluate ad content for unwanted content using modern artificial intelligence systems. An analysis of existing tools and their results is conducted using a real high-risk advertisement. A detailed report is provided according to different evaluation metrics. The expediency of integrating the above technologies into the business logic of advertising networks is determined.

Keywords: NLP; AI; advertising campaigns; the Internet advertising exchange systems; fraudulent content; malicious content; online advertisements; detection.

Стаття надійшла до редакції 07.04.2023.