



## ЕЛЕМЕНТИ НЕЙРОМЕРЕЖНОЇ ТЕХНОЛОГІЇ АНАЛІЗУ СТАВЛЕННЯ КОРИСТУВАЧІВ МЕРЕЖІ "ТВІТЕР" ДО БРЕНДІВ

**Вступ.** *Висвітлено проблему ефективної організації збору й аналізу інформації про ставлення користувачів мережі "Твітер" до брендів у формі програмного застосунку. Розглянуто проблеми дослідження сучасних засобів збору й аналізу інформації; визначення функціоналу, який має реалізовувати застосунок; аналізу архітектурних рішень і вибору програмних засобів, необхідних для його реалізації.*

**Методи.** *У процесі досліджень застосовано теорію маркетингу у сфері збору інформації про висновки споживачів, досліджено методи аналізу інформації з метою класифікації настрою споживачів, емпіричний аналіз і синтез архітектур, що застосовувалися у створенні й порівнянні моделей нейронних мереж для класифікації тексту, розроблення та побудови власної моделі для класифікації.*

**Результати.** *У межах задачі програмної реалізації аналізу тексту твітів досліджено архітектуру згорткових і рекурентних нейронних мереж, здійснено порівняння різних значень гіперпараметрів нейронних мереж, зокрема і функцій активації, функцій втрат, кількості епох навчання, кількості шарів мережі, виконано порівняння різних Python-бібліотек для оброблення природної мови в контексті оцінювання твітів.*

**Висновки.** *Практичне значення дослідження полягає у створенні програмного засобу для ефективного аналізу ставлення користувачів мережі "Твітер" до брендів, який може слугувати для підвищення ефективності маркетингової діяльності брендів.*

**Ключові слова:** *твітер, аналіз інформації, класифікація тексту, сентиментальний аналіз, нейронні мережі, програмний застосунок, бібліотеки Python.*

### Вступ

Важливим аспектом у виборі маркетингової стратегії для брендів виступають настрої їхніх споживачів. Сфера збору інформації про сприйняття бренду охоплює паперові чи онлайн-огляди продукту, спілкування із службами підтримки компаній, спілкування з іншими потенційними споживачами, зацікавленість щодо придбання пробних версій продуктів і реакція на рекламу, вислови у соціальних мережах. Кожне із цих джерел допомагає проаналізувати сприйняття бренду в суспільстві, що в подальшому може впливати на успіх компаній у разі використання відповідних маркетингових стратегій.

Сприйняття бренду й досвід формують його репутацію. Компанія з гарною репутацією бренду приваблює лояльних клієнтів, які обирають продукт або послугу серед конкурентів.

### Методи

Для оцінювання сприйняття бренду використовують такі методи (Kirsch, 2023):

- проведення опитувань щодо сприйняття бренду, думки про нього та його протистояння конкурентам. Постановка запитань, які стосуються емоційних, когнітивних і орієнтованих на дії факторів;

- відстеження згадувань бренду в інтернеті за допомогою соціальних інструментів прослуховування. Такі згадування бувають представлені у коментарях у соціальних мережах, онлайн-оглядах, відповідних хештегах, форумах і згадках у новинах. Мірою зростання бренду ростиме і цей набір даних;

- аудит бренду для оцінювання сприйняття бренду порівняно із сприйняттям конкурентів. Він має включати дослідження первинної та вторинної цільової аудиторії, конкурентний аналіз продуктів або послуг, ретельну оцінку комунікаційної стратегії та глибоке занурення в позиціонування бренду;

- збір даних від клієнтів у кожній точці шляху покупця. Може містити дані про те, як покупці досліджують інформацію, оцінюють продукти, порівнюють між конкурентами, взаємодіють зі службою підтримки клієнтів, приймають рішення про покупку, підключаються та взаємодіють із брендом після покупки.

Із зростанням популярності бренду зростає і обсяг інформації, що підлягає обробленню в задачі аналізу споживачів до нього. Тому збільшенню ефективності аналізу сприятиме його автоматизація, яка широко представлена у вигляді онлайн-сервісів. Їхнє використання дозволяє повністю охоплювати споживачів продукту або послуги, знаходити потенціальних споживачів та ефективно оцінювати їхні настрої. Такі програмні продукти спрямовані на економію часу працівників і збільшення капіталу брендів.

На ринку наявні продукти для виконання зазначених вище завдань. Згідно з відгуками користувачів представлені продукти не завжди задовольняють їхні потреби. Тож постає проблема створення продукту, який би враховував необхідні метрики для аналізу ставлення користувачів до брендів, забезпечував високий відсоток покриття інформаційних джерел і відображав реальну ситуацію, щоб давати змогу здійснювати ефективний маркетинг для бізнесів. Метою цього дослідження є реалізація ефективної організації збору й аналізу інформації про ставлення користувачів соціальної мережі "Твітер" до брендів у формі програмного застосунку.

© Жуланова Ольга, Ващіліна Олена, 2023



**Особливості предметної області й робота з ними в задачі класифікації тексту.** В задачі аналізу сприйняття бренду через збір інформації у соціальних мережах необхідно враховувати особливості цільових медіа. До них належать:

- характеристики цільової аудиторії соціальної мережі;
- тематика ресурсу;
- технічні та функціональні особливості ресурсу;
- представлені засоби висловлення думки споживачів (текст, аудіо, відео, зображення);
- переважальний пристрій для користування ресурсом.

Інструменти для автоматизації збору й аналізу інформації з інтернет-ресурсів повинні враховувати вказані особливості й опрацьовувати їх для об'єктивної роботи й оцінювання.

Соціальна мережа "Твітер" (англ. Twitter) є однією з найпопулярніших у світі, оскільки охоплює 319 мільйонів активних користувачів. Мережу "Твітер" як платформу для контент-маркетингу використовують більше 80 % маркетологів, які працюють у сегменті B2B (Nistor et al., 2021). Мережа "Твітер" має такі значення згаданих критеріїв (табл. 1).

Таблиця 1

Характеристики соціальної мережі	
Критерії	Показники
Характеристики цільової аудиторії соціальної мережі	43,6 % – жінки, 56,4 % – чоловіки (Statista, 2021) 38,5 % – 25–34 роки, 20,7 % – 35–49 років (Островський, 2021)
Тематики ресурсу	Відсутні
Технічні та функціональні особливості ресурсу	Максимальна кількість символів у твіті – 140; необмежена кількість твітів для одного користувача
Представлені засоби висловлення думки споживачів	Текст, посилання, зображення
Переважальні пристрої для користування ресурсом	Смартфони

Для створення сервісу щодо аналізу текстів у мережі "Твітер" важливим є факт відсутності тематики ресурсу та свобода вираження у ньому, що треба враховувати у підготовці вхідних даних для аналізу. Одним із факторів є наявність граматичних помилок і сленгу.

**Засоби аналізу та класифікації текстів та їхнє порівняння.** Для аналізу текстів часто застосовують програмні інструменти сентиментального аналізу, який є задачею класифікації тексту (Minaee et al., 2021), аналізу намірів і контекстуального семантичного пошуку.

Програмні продукти, що використовують алгоритми для визначення значень необхідних метрик, покликани надавати користувачам згруповану та зрозумілу інформацію, яка може бути додатково ілюстрована числами, графіками, діаграмами, посиланнями на джерела.

Для сентиментального аналізу зазвичай поєднують використання статистики, оброблення природної мови та машинного навчання для визначення та вилучення суб'єктивної оцінки автора про предмет висловлювання.

Для різних цілей використовують різні рівні деталізації сентиментального аналізу, що означає дослідження різного обсягу інформації та виокремлення її значущих частин. Результатом сентиментального аналізу є висновок про тональність (позиція автора відносно згаданої теми), виражену словами або реченнями, яка може бути нейтральною, позитивною або негативною.

Існують такі підходи до здійснення сентимент-аналізу тексту:

- застосування правил із використанням раніше виявлених шаблонів. Текст поділяють на окремі слова та послідовності слів, у яких оцінюється тональність. Такі виокремлені й оцінені елементи стають шаблонами для правил, які застосовують для визначення тональності текстів;
- машинне навчання без учителя. Підхід базується на тому, що найбільше значення мають слова або послідовності слів, що зустрічаються у досліджуваному тексті або уривку найчастіше. Отже, визначивши ці терміни та їхню тональність, можна дійти висновку про тональність повного тексту;
- машинне навчання з учителем. Метод потребує навчального набору даних, що мають визначений емоційний характер. На його базі будують і використовують статистичний класифікатор;
- застосування графових моделей тексту. Підхід враховує те, що слова у тексті нерівнозначні, і ті, які зустрічаються частіше, сильніше впливають на його тональність.

Останній метод включає кілька етапів, серед яких:

- побудова графа на основі тексту, що підлягає дослідженню;
- визначення рангів вершин графа;
- оцінювання та класифікація знайдених слів;
- підсумок аналізу й визначення результату.

Також існують гібридні методи, які поєднують у собі всі або кілька з наведених вище. Оскільки це дає можливість дослідити тональність тексту з різних боків, можна стверджувати, що поєднання кількох підходів є найефективнішим, оскільки дозволяє покрити всі значущі для тональності елементи.

Однак, враховуючи те, що будь-які методи сентиментального аналізу працюють над природною мовою, існують поширені проблеми, пов'язані із цим, що не дають можливості досягти максимальної високої точності оцінки тональності.

Ці проблеми часто спричинені:

- орфографічними помилками у висловлюваннях користувачів електронними ресурсами;
- тональність оцінюється для повного тексту, а не для конкретного об'єкта висловлювання;
- оцінка автора не завжди визначається наявністю створених ним позитивних, нейтральних або негативних висловлювань про продукт;
- покращити результати автоматичного визначення тональності тексту можна за допомогою використання методів автоматичного виправлення орфографічних помилок, удосконалення словників (для методів, заснованих на



словниках) і навчальної вибірки (для методів машинного навчання). Також можна підвищити точність роботи алгоритмів, застосовуючи розробки з інших проблем природної мови, як-от:

- автоматичне реферування (automatic summarization);
- виявлення кореферентності, референціональної множини (coreference resolution);
- аналіз порівнянь;
- вилучення об'єктів із текстів (Named-entity recognition, NER) та виявлення відносин між ними (Relationship extraction).

Частота орфографічних помилок і сленгу в повідомленнях твітів набагато вища, ніж в інших доменах, оскільки користувачі зазвичай публікують повідомлення з найрізноманітніших електронних пристроїв, включаючи смартфони та планшети. Крім того, в такому середовищі користувачі розвивають свою власну культуру та специфічний словниковий запас, який, хоч і є обмеженим із погляду довжини (напр., кількості символів), може передавати насичене значення.

**Теоретичне порівняння методів аналізу тексту.** Для використання аналізу на базі лексикону або словників створюють набори слів із наперед визначеною тональністю, які надалі порівнюють зі словами з досліджуваного тексту. Також словники можуть містити і словосполучення, послідовності слів, фразеологізми, їхнє використання підвищує точність алгоритму.

Після розгляду методу з використанням словників, з погляду реалізації у ньому можна виокремити такі етапи:

1. Завантаження сентимент-словників, що містять уже марковані слова або послідовності слів.
2. Завантаження тренувального набору даних із мережі "Твіттер", на якій проводитиметься навчання.
3. Очищення твітів від стоп-слів.
4. Токенізація слів із набору.
5. Порівняння слів із словниковим варіантом, присвоєння їм відповідної тональності й урахування їхньої тональності для кінцевого результату.
6. Обчислення співвідношення кількості слів різної тональності та визначення результату.

Указаний підхід не можна назвати ефективним, оскільки у ньому не враховано способи поєднання слів, що часто призводить до зміни їхнього значення та настрою автора.

Цих недоліків можна позбутися з використанням машинного навчання, оскільки цей метод, подібно розуму людини, використовує власний досвід, а не "сухі" дані і найкраще підходить для аналізу текстового контенту із соціальних мереж. Серед недоліків традиційних методів машинного навчання є необхідність ретельного відбору ознак для класифікації і, відповідно, трудомістке попереднє оброблення даних. До того ж, в аналізі тексту із соціальних мереж не враховується наявність граматичних помилок і сленгу, який може використовуватися доволі часто.

Найпоширенішим методом класифікації тексту нині є використання методів глибокого машинного навчання із застосуванням нейронних мереж, оскільки вони показують найкращі результати. Серед переваг також є відсутність необхідності "ручного" відбору ознак у даних і гнучкість – за зміни постановки задачі не потрібно масштабних змін у коді нейронної мережі.

Порівнюючи результати роботи, бачимо, що нейронні мережі показують кращі результати ніж методи логічної регресії, однак час роботи перших набагато більший (Sheng, Socher, & Manning, 2015).

За порівняння найпоширеніших нейронних мереж для класифікації тексту майже однаково себе показують рекурентні і згорткові мережі, причому результати можуть не набагато відрізнитись у разі зміни теми вхідних даних. Середній результат для обох мереж в усіх випадках дорівнює приблизно 87 %.

Дослідження свідчать про те, що на точність результатів роботи рекурентних нейронних мереж впливає кількість прихованих шарів у мережі, архітектура, кількість комірок пам'яті і їхніх типів (GRU/LSTM), а також кількість шарів, яким належать ці комірки, механізм попереднього оброблення даних, кількість кроків оброблення (Minaee et al., 2021).

**Архітектура нейронних мереж для задач класифікації тексту.** Класифікація тексту в основному зосереджена на трьох аспектах:

- пошук ознак: найчастіше для цього використовують "мішок слів", проте існує ще кілька мультиплексних функцій, наприклад теги частини мови, ядра дерева та фрази-іменники;
- вибір ознак: призначений для видалення "зашумлених" ознак і збільшення ефективності класифікації; найпоширеніший метод вибору – видалення стоп-слів;
- алгоритми машинного навчання: використання класифікаторів, наприклад, наївний баєсів класифікатор (NB), логістична регресія (LR) і метод опорних векторів (SVM).

Ефективність нейронної мережі залежить від результатів кожного з аспектів, її підвищення є головною метою в задачі побудови моделі для класифікації тексту. До метрик, які визначають ефективність мережі, зазвичай належать значення точності (1) і функції втрат моделі:

$$Acc = \frac{N_{yPred}}{N_{yTrue}} \cdot 100 \%, \quad (1)$$

де  $Acc$  – точність моделі;  $N_{yTrue}$  – кількість коректних класифікацій;  $N_{yPred}$  – кількість значень, що класифікуються.

Функції втрат, які поділяють на втрати класифікації і втрати регресії, мають різне застосування і вибираються залежно від задачі, тому можуть показувати різні результати у різних контекстах використання.

Для проблем класифікації найчастіше використовують функцію cross-entropy (2), значення якої зростає із збільшенням різниці між прогнозованою оцінкою та фактичною:

$$CrossEntropyLoss = -(y_i \hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i)), \quad (2)$$

де  $\hat{y}_i$  – значення функції втрат;  $y_i$  – значення фактичної оцінки.

З рівняння (2) випливає, що коли фактична оцінка дорівнює 1 ( $y_i = 1$ ), то друга половина функції стає рівною 0; у разі, якщо фактична оцінка дорівнює 0 ( $y_i = 0$ ), перша половина стає рівною 0. У такому випадку обчислюється добуток фактичної прогнозованої імовірності для основного класу істинності. Важливим аспектом цього є те, що функція робить значущими прогнози, які є високоімовірними, але помилковими.

Відповідно значення функції втрат у процесі побудови моделі для класифікації підлягає мінімізації.



Порівнюючи методи машинного навчання, відразу звертаємо увагу на недолік методу машинного навчання з учителем, що полягає у потребі в анотованих наборах даних. Тому може бути необхідною ручна анотація, але це тривалий, вартісний процес, який часто призводить до помилкових висновків. Для методів напівконтрольованого навчання (SSL) використовують у навчальному процесі немарковані дані. Немарковані дані можуть покращити класифікацію у програмах, де кількість маркованих даних дуже мала. Через це методи на основі SSL є перспективною альтернативою для аналізу тональності твітів, оскільки доступна переважна кількість некоментованих твітів, на відміну від обмеженої кількості анотованих. З огляду на це прості напівкервані моделі класифікації, такі як Self-training, можуть мати велике практичне значення.

Кращі результати (Appel et al., 2016) показують гібридні методи, що поєднують у собі методи машинного навчання з використанням словників. Згортова нейронна мережа, розширена за допомогою методу опорних векторів і рекурентного шару й аналіз на основі словників із машинним навчанням показують покращений результат. Комбінація CNN, LSTM (Long short-term memory) і SVM спрямована на використання переваг двох моделей архітектури глибокої мережі й алгоритмів SVM під час виконання аналізу настроїв у різних доменах і типах наборів даних. Крім того, існують різні типи вхідних даних, отриманих із соціальних мереж, наприклад твіти та огляди, різноманітність тем у кожному наборі даних, розмір вибірки та більшу чи меншу присутність явних настроїв і невідповідних інформації. Деякі підходи можуть бути неефективними в різних областях, з недостатньою точністю та неефективністю в аналізі настроїв. Як наслідок, певні підходи можуть бути погано застосовними і не знаходити використання для певних типів вхідних даних.

Згортова нейронна мережа, що може бути елементом архітектури нейронної мережі для класифікації текстів, включає серію фільтрів різних розмірів і форми, які згортають вхідну матрицю даних, щоб зменшити її до матриці розірності ще нижчого рівня. Для розв'язання задачі класифікації текстів згортові нейронні мережі застосовують до векторних представлень речень. Зниження дискретизації, що використовується у згортовій нейронній мережі, має назву  $L_2$ -регуляризація. Згортова мережа застосовує функцію активації, що допомагає їй працювати в ядрі, тобто у просторі високої розмірності для нейронного оброблення.

Дослідження (Conneau et al., 2017) показують, що збільшення глибини згортової мережі підвищує її точність. На результати також впливає метод масштабування, що застосовується до згорнутих даних. Зазначимо, що алгоритм MaxPooling зарекомендував себе краще, ніж KmaxPooling і Convolution. MaxPooling у цілому досягає кращих результатів, які, проте, близькі до результатів з використанням Convolution із кроком рівним 2. Обидва алгоритми значно ефективніші, ніж KmaxPooling. Обидва механізми масштабування виконують локальну операцію max, яка обмежена трьома послідовними значеннями, тоді як алгоритм  $k$ -max враховує відразу ціле речення. Згідно з експериментами, він погано впливає на ефективність класифікатора під час використання у проміжних шарах (за винятком менших наборів даних) (Conneau et al., 2017).

Рекурентна нейронна мережа (RNN – від англ. Recurrent neural network) – це один із типів нейронних мереж, який використовує рекурсивний алгоритм. Такі нейронні мережі найчастіше застосовують у разі проблем з оброблення природної мови, оскільки вони "вміють" працювати з контекстом у тексті. Рекурсивні нейронні мережі відрізняються своїм механізмом "запам'ятовування", який використовує інформацію з попередніх ітерацій для кожної наступної, що дозволяє впливати на результат кожного згортового шару мережі на їхній основі. Цим RNN відрізняються від інших нейронних мереж, які не враховують впливу попередніх виходів на поточний.

Переваги рекурентної нейронної мережі:

- врахування вибірок попередніх шарів на наступних;
- рекурентні нейронні мережі використовують зі згортовими шарами для розширення ефективного сусідства елементів.
- Недоліки рекурентної нейронної мережі:
  - проблеми зі зникненням і вибухом градієнта;
  - навчання RNN – складне завдання (Pascanu, Mikolov, & Bengio, 2013);
  - через нелінійні алгоритми оброблення даних і налаштування ваг моделі часто важко встановити стабільність системи (Chiang, Chang, & Chang, 2004);
  - не може обробляти дуже довгі послідовності з використанням функцій активації relu або tanh.

Мережі довготермінової пам'яті (LSTM) є модифікованою версією рекурентних нейронних мереж, що представляють гнучкий механізм керування пам'яттю мережі. Використовуючи їх, можна розв'язати проблему зникаючого градієнта, що властивий звичайним рекурентним мережам. LSTM дозволяє враховувати вплив вибірок даних із попередніх виходів шарів, не обмежуючи час життя цих вибірок за допомогою логіки відбору таких даних для наступних ітерацій. Такі моделі навчаються алгоритмом зворотного поширення.

У мережі LSTM наявні три види шлюзів:

1. Вхідний шлюз (англ. "input gate") відбирає значення із вхідних даних, які мають вплинути на такі етапи:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (4)$$

де  $i$  – вихідне значення шлюзу "input gate";  $\sigma$  – сигмоїдна функція;  $t$  – такт;  $W$  – вага;  $h$  – вектор вихідних даних шару;  $x$  – вектор вхідних даних шару;  $b$  – значення нейрона зсуву;  $\tilde{C}$  – значення комірки пам'яті.

2. Шлюз забуття (англ. "forget gate") відбирає дані, які підлягають виключенню з блоку. Вони визначаються також сигмоїдною функцією, яка залежить від вихідних даних попередньої ітерації і вхідних значень. Далі шлюз визначає число від 0 (виключити) і 1 (зберегти) для кожного значення у стані комірки C:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (5)$$

де  $\sigma$  – сигмоїдна функція;  $t$  – такт;  $W$  – вага;  $h$  – вектор вихідних даних шару;  $x$  – вектор вхідних даних шару;  $b$  – значення нейрона зсуву.



3. Вихідний шлюз (англ. "output gate") використовує поточний вхід і вихід попереднього блоку для визначення поточного виходу. Сигмоїдна функція (6) визначає, які значення пропускати через шлюз. Функція  $\tanh$  надає вагу значенням, які передаються, вирішуючи рівень їхньої важливості числом у діапазоні від  $-1$  до  $1$ , яке далі множиться на вихід сигмоїда (7):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (6)$$

$$h_t = O_t \cdot \tanh(C_t), \quad (7)$$

де  $O$  – вихідне значення шлюзу "output gate";  $\sigma$  – сигмоїдна функція;  $t$  – такт;  $W$  – вага;  $h$  – вектор вихідних даних шару;  $x$  – вектор вхідних даних шару;  $b$  – значення нейрона зсуву;  $\tanh$  – функція "гіперболічний тангенс".

Існують роботи з удосконалення моделей RNN і LSTM для класифікації тексту шляхом отримання більшого обсягу інформації з урахуванням, наприклад, деревоподібних структур природної мови, відношення між словами в тексті, теми документів тощо. З такою метою була розроблена модель Tree-LSTM (Sheng, Socher, & Manning, 2015), яка є узагальненням LSTM до деревоподібної типології мережі, задля дослідження насичених семантичних елементів документів. Її автори стверджують, що Tree-LSTM є кращою моделлю, ніж ланцюгова структурована LSTM для завдань NLP, оскільки природна мова проявляє синтаксичні властивості, які природно поєднують слова у фрази. Вони підтверджують ефективність Tree-LSTM у розв'язанні двох завдань: класифікації настроїв і прогнозування смислової спорідненості двох речень. Архітектуру цих моделей показано на рис. 1. Елементи  $x_i$  та  $y_i$  позначають вхід і вихід кожної комірки.

У процесі розширення ланцюгової LSTM із застосуванням деревоподібної структури, яка використовує комірку пам'яті для зберігання історії кількох дочірніх елементів або кількох комірок-нащадків у рекурсивному процесі, нова модель забезпечує ефективний спосіб розгляду взаємодії елементів на великій відстані в ієрархіях, наприклад, структурі аналізу мови або зображень (Zhu, Sobhani, & Guo, 2015).

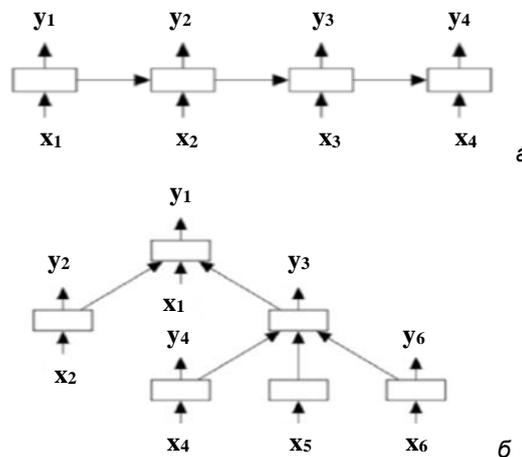


Рис. 1. LSTM-мережа: а – ланцюгова; б – деревоподібна

**Метою дослідження** є розроблення концепції функціонування проєктів електронної комерції, націлених на індивідуальні замовлення. В основу такої моделі покладено інформаційну аналітику інформаційного простору проєкту.

Ефективність нейронної мережі переважно визначають гіперпараметри, якими можуть бути представлені рівень навчання мережі, функції активації, кількість шарів і нейронів у них, кількість епох навчання тощо.

Вибір функції активації мережі має залежати від поставленої задачі, яку ця мережа розв'язує, і є важливим фактором в її ефективності, оскільки описує функцію перетворення суми ваг для шарів мережі. Ці функції покликані вирізати дані, які найбільше впливають на результат роботи мережі.

У створенні рекурентних і згорткових нейронних мереж можуть виникати проблеми зникаючого або вибухового градієнта. Явище зникаючого градієнта виникає за використання функцій активації з похідними, близькими до нуля. Під час навчання мережі через їхнє множення за декілька ітерацій градієнт стає настільки малим, що ваги майже не оновлюються і під час навчання не вдається досягти глобального мінімуму функції втрати мережі.

За значень похідних функцій активації більше одиниці відбувається вибухання градієнта, тобто різке зростання ваг моделі з великою кількістю шарів при навчанні.

В обох випадках рівень навчання моделі наближається до нуля та досягається невисока її точність, що впливає з рівняння (8):

$$W_{new} = W_{old} - \eta \left( \frac{\partial L}{\partial W_{old}} \right), \quad (8)$$

де  $W_{new}$  – нова вага;  $W_{old}$  – стара вага;  $\eta$  – рівень навчання;  $L$  – значення функції втрат. Обидва явища спричиняють нестабільність моделі та великі значення результуючої функції втрат.

У продуктивності мережі важливу роль відіграють три основні аспекти нейронної мережі: архітектура мережі та шаблон зв'язків між блоками, алгоритм навчання та функції активації, що використовуються в мережі. Проте більшість досліджень з аналізу нейронних мереж зосереджують увагу на ролі алгоритму навчання, оскільки важливістю функцій активації, які використовуються нейронними мережами, переважно нехтують.

Властивості, яким загалом має відповідати функції активації, такі:

- функція активації має бути неперервною і обмеженою;



▪ функція повинна бути сигмоїдною або межі для нескінченності мають задовольняти таким рівнянням (Chandra, & Singh, 2004):

$$f(x) = \alpha, \quad (9)$$

$$f(x) = \beta, \quad (10)$$

$$\alpha < \beta, \quad (11)$$

де  $x$  – вектор вхідних значень функції;  $f(x)$  – функція активації;  $\alpha, \beta$  – довільні значення.

У дослідженні (Farzad, Mashayekhi, & Hassanpour, 2019) оцінено продуктивності мережі LSTM із 23 різними функціями активації, які можна використовувати замість сигмоїдної. Змінювали кількість прихованих шарів у мережі та використовували три різні набори даних для класифікації. Результати показали, що деякі менш відомі функції активації, такі як функція Елліота (LeRoy Elliott, 1998) та її модифікації можуть привести до меншого рівня помилок порівняно з найпопулярнішими функціями: *relu*, *tanh* і *sigmoid* (Farzad, Mashayekhi, & Hassanpour, 2019).

До проблем, пов'язаних із навчанням нейронних мереж, належить проблема перенавчання. Однією з її причин може бути тісний зв'язок між нейронами мережі, що веде до гарних результатів на тренувальних даних, але до протилежного результату на тестових. Для розв'язання проблеми застосовують метод стохастичної регуляризації Dropout – виключення нейронів разом із їхніми зв'язками із шару з певною ймовірністю під час навчання. За використання методу мінімізація функції втрат відбувається стохастично при розподілі шуму (Srivastava et al., 2014). Це можна розглядати як мінімізацію очікуваної функції втрат. Отже, під час тренування моделі, поступове оновлення ваг відбувається без урахування виключених нейронів.

За використання Dropout із шарів випадковим чином виключаються нейрони із заданою ймовірністю  $p$  з розподілу Бернуллі, а ваги нейронів, що залишились, підлягають масштабуванню за формулою (12):

$$W_{in} = W / \left(\frac{n_{ex}}{N}\right), \quad (12)$$

де  $W_{in}$  – вага включеного нейрона;  $W$  – оригінальна вага нейрона;  $n_{ex}$  – кількість виключених нейронів шару;  $N$  – кількість усіх нейронів шару.

Ефективність тієї чи іншої архітектури нейронної мережі можливо визначити лише емпіричним шляхом, і хоча результати досліджень точно показують найефективніші алгоритми і функції, що можуть бути використані у тренуванні моделі, різне їхнє поєднання може приводити до іншого результату. Тому для визначення оптимальної будови моделі для аналізу текстів необхідно проводити експерименти з використанням різних архітектур. Нейронну мережу з тією архітектурою, яка показує найвищу точність і найменше значення функції втрат на тестових даних, вважають найефективнішою.

**Мова програмування для реалізації застосунку.** Для реалізації сентиментального аналізу твітів найкраще використовувати мову Python, оскільки вона:

- є простою зі структурного погляду побудови програм;
- має відкритий вихідний код;
- містить багато зручних методів для роботи з даними;
- використовується у великій кількості бібліотек для машинного навчання й аналізу даних, написаних мовою Python.

Для вибору оптимального варіанта бібліотеки для класифікації виконано порівняння найпоширеніших нині бібліотек для прогнозування оцінки природної мови: NLTK, TextBlob і Flair. Найкращий результат показала бібліотека Flair із точністю 74 %. Результати використання кожної з бібліотек наведено у табл. 2.

Таблиця 2

Точність бібліотек NLTK, TextBlob і Flair у прогнозуванні сентименту тексту, %

NLTK	TextBlob	Flair
54	44	74

**Порівняння архітектур моделі нейронної мережі для класифікації тексту.** Модель для оцінювання емоційного забарвлення тексту, побудовану з використанням шарів SpatialDropout1D та LSTM, показала найвищу точність оцінки – 80 % і найменше значення функції втрат – 0,51.

```

model.add(Embedding(vocab_size, 100, input_length = X.shape[1]))
model.add(SpatialDropout1D(0.5))
model.add(LSTM(176, dropout=0.2, recurrent_dropout=0.3))
model.add(Dense(7, activation='softmax'))
model.compile(loss = 'categorical_crossentropy', optimizer='adam',
metrics = ['accuracy'])

```

Рис. 2. Код побудови моделі нейронної мережі з використанням LSTM-шару



Для порівняння створено моделі нейронної мережі з додаванням більшої кількості шарів. Припущення, що додавання шару згорткової мережі підвищить точність моделі, не справджується у разі побудови моделі, де такий шар передує рекурентному LSTM-шару:

```

model.add(Embedding(vocab_size, 100, input_length = X.shape[1]))
model.add(SpatialDropout1D(0.5))
model.add(Conv1D(filters=32, strides=3, activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(LSTM(50, dropout=0.2, recurrent_dropout=0.3))
model.add(Dense(7, activation='softmax'))
model.compile(loss = 'categorical_crossentropy', optimizer='adam',
metrics = ['accuracy'])
    
```

Рис. 3. Код побудови моделі нейронної мережі з використанням LSTM-шару і CNN-шару

Клас Conv1D, що реалізує функцію згортки, у конструкторі набуває параметрів:

- **filters:** ціле число, розмірність вихідного простору (тобто кількість вихідних фільтрів у згортці);
- **kernel\_size:** ціле число або кортеж/список одного цілого числа, що вказує довжину вікна одновимірної згортки;
- **activation:** функція активації шару (Keras, 2023).

**Результати**

Для узагальнення результатів згортки додано шар масштабування – MaxPooling1D, якого набуває параметр pool\_size: кратність масштабування.

Створена модель описується значеннями з рис. 4.

Layer (type)	Output Shape	Param #
embedding_13 (Embedding)	(None, 50, 100)	70000
spatial_dropout1d_11 (SpatialDropout1D)	(None, 50, 100)	0
conv1d (Conv1D)	(None, 48, 32)	9632
max_pooling1d (MaxPooling1D)	(None, 16, 32)	0
lstm_11 (LSTM)	(None, 50)	16600
dense_11 (Dense)	(None, 7)	357
Total params: 96,589		
Trainable params: 96,589		
Non-trainable params: 0		

Рис. 4. Вихідні розмірності і кількість вихідних параметрів шарів моделі з використанням шару згортки

Отримані значення майже не відрізняються від результатів мережі без згортки, але в першій мережі швидкість роботи довша за рахунок більшої кількості обчислень.

Обидві архітектури протестовано на кількості епох 5, 7, 10, 100. Модель без використання згорткового шару показує результати, наведені в табл. 3 та 4.

Таблиця 3

Метрики моделі без згорткового шару для кількості епох 5, 7, 10 і 100

Кількість епох	5	7	10	100
Точність моделі, %	80	79	79	76
Значення функції втрат моделі	0,51	0,51	0,51	0,69

Таблиця 4

Метрики моделі зі згортковим шаром для кількості епох 5, 7, 10 і 100

Кількість епох	5	7	10	100
Точність моделі, %	79	78	78	78
Значення функції втрат моделі	0,56	0,6	0,65	0,72

Згідно з результатами, можна сказати, що збільшення кількості ітерацій тренування моделі хоч і не сильно, але призводить до перенавчання й гірших результатів на тестових даних.



За значного збільшення кількості епох (до 100) вдалося досягти лише меншої точності.

У цьому експерименті варіювання розміру словника для тренування мережі впливає на точність моделі.

Використання розміру словника у 700 слів дало кращі результати, ніж використання словника обсягом 500 слів. Це може означати, що зменшення лексики призводить до меншої "обізнаності" моделі та підвищення вірогідності прогнозування на основі невідомих токенів.

Серед оптимізаторів значно кращі результати показує оптимізатор Adam порівняно з RMSprop та Adamax, хоча в ході дослідження (Choi et al., 2019) їхні метрики досить близькі. Самі ці оптимізатори обрано для порівняння, оскільки інші функції, такі як SGD та Momentum у експериментах з різним набором даних, з різною архітектурою та тривалістю навчання показують себе гірше.

Оптимізатор Adam – алгоритм для градієнтної оптимізації стохастичних цільових функцій першого порядку, заснований на адаптивних оцінках моментів нижчого порядку. Метод простий у реалізації, ефективний в обчислюванні, має невеликі вимоги до пам'яті, незалежний від діагонального масштабування градієнтів і добре підходить для задач, які є великими з погляду даних або параметрів. Гіперпараметри методу мають інтуїтивно зрозумілі інтерпретації і зазвичай вимагають невеликого налаштування. Емпіричні результати демонструють, що Адам добре працює на практиці і вигідно порівнюється з іншими методами стохастичної оптимізації (Kingma, & Ba, 2014).

**Переваги створення застосунку у формі вебсервісу.** Програмну реалізацію застосунку для аналізу ставлення користувачів мережі "Твітер" до брендів доцільно здійснити у формі вебсервісу. Створення стандартизованого інтерфейсу забезпечує гнучкість такого сервісу, програмний код якого може розвиватися, не залежачи від клієнтського застосунку. Клієнт, так само, може бути реалізований із застосуванням будь-яких технологій, у вигляді вебдодатка, настільного або мобільного додатка.

Нині компанії використовують веб-API для трансформації свого бізнесу. Підприємства використовують веб-API для збільшення доходу, покращення інновацій і скорочення їхнього часу виходу на ринок за допомогою двох стратегій:

- використання чужих API;
- надання власних API для внутрішніх і зовнішніх розробників (Higginbotham, 2015).

API дозволяють використовувати роботу інших розробників. Замість того, щоб створювати кожну функцію самостійно, команди можуть використовувати сторонні API, щоб додати нові функції та зосередитися на аспектах рішень, які є унікальними. Для деяких рішень це може привести до створення програми повністю зі сторонніх API, з невеликою кількістю або зовсім без спеціальних розробок.

**Алгоритм застосунку.** Після отримання запиту програма, яка здійснює аналіз, виконує послідовність операцій, показаних на рис. 5.

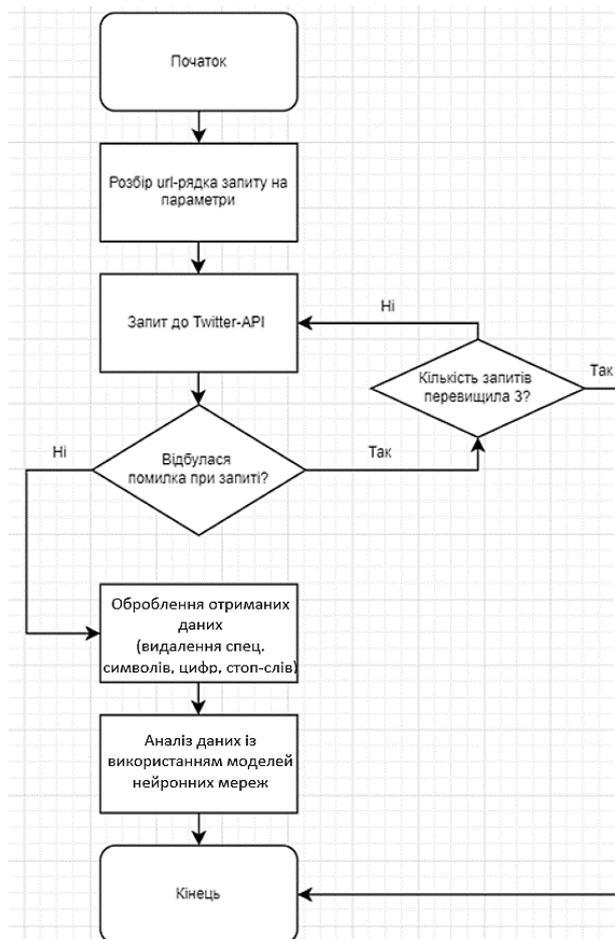


Рис. 5. Робота програми після отримання запиту



Алгоритм забезпечує оброблення ситуації, за якої на запит до Twitter-API відбувається помилка. В цьому випадку програма повторно надсилає запит, якщо загальна кількість таких запитів не перевищує 3. Причиною помилки може стати неробочий стан інтерфейсу або його зміна, коли надсилаються запити на неіснуючі методи.

**Організація виведення результатів.** Щоб забезпечити краще сприйняття інформації, виділення взаємозв'язків і відношень, що містяться в інформації, результати роботи застосунку з аналізу ставлення користувачів мережі "Твітер" до брендів доцільно оформити у вигляді інфографіки (рис. 6):

1. Хмара слів.
2. Загальна кількість опрацьованих твітів.
3. Графік залежності кількості твітів від дати.
4. Графік залежності кількості негативних, позитивних і нейтральних твітів від дати.
5. Кругова діаграма співвідношення кількості негативних, позитивних і нейтральних твітів.
6. Графік залежності кількості сумних, злих, щасливих, здивованих, люблячих і наляканих твітів від дати.
7. Кругова діаграма співвідношення кількості сумних, злих, щасливих, здивованих, люблячих і наляканих твітів.
8. Список проаналізованих твітів з індикатором визначеного сентименту у вигляді стрілки.

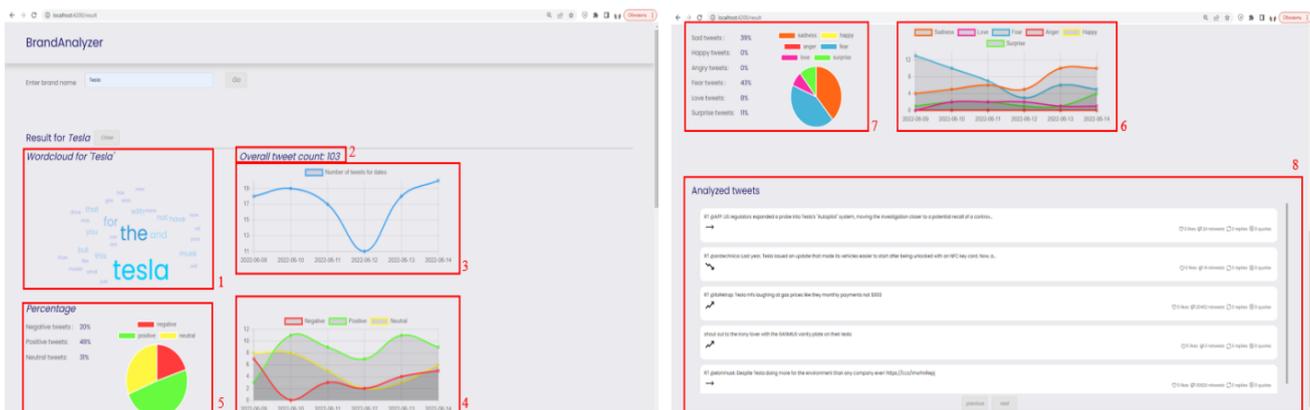


Рис. 6. Результат роботи застосунку за запитом "Tesla"

## Висновки

Створення програмного засобу для аналізу ставлення користувачів соціальної мережі "Твітер" до брендів слугуватиме підвищенню ефективності маркетингової діяльності брендів.

За підготовки вхідних даних для аналізу текстів із мережі "Твітер" слід враховувати такі особливості: відсутність тематики ресурсу, наявність граматичних помилок і сленгу;

Одним із кращих засобів аналізу текстової інформації є нейронні мережі, побудовані із застосуванням підходів, які дозволяють враховувати контекст інформації;

Оптимальною є архітектура нейронної мережі з використанням шару довготермінової пам'яті без згорткового шару, оскільки швидкість роботи першої є меншою при майже однаковій точності;

Програмна реалізація застосунку для аналізу ставлення користувачів мережі "Твітер" до брендів у формі вебсервісу дозволить використовувати функціонал аналізу тексту в різних застосунках без додаткових затрат.

**Внесок авторів:** Ольга Жуланова – збір емпіричних даних, написання програмного коду й оброблення результатів програмування; Олена Ваціліна – огляд літературних джерел, розроблення методів і методології дослідження, опис результатів і написання висновків.

## Список використаних джерел

- Островський, С. (2021). *Twitter – стисло про соціальну мережу короткого формату*. ApiX-Drive. <https://apix-drive.com/ru/blog/useful/twitter-korotko-o-socialnoj-seti-korotkogo-formata>
- Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, (108). <https://doi.org/10.1016/j.knosys.2016.05.040>
- Chandra, P., & Singh, Y. (2004). Feedforward Sigmoidal Networks—Equicontinuity and Fault-Tolerance Properties. *IEEE Transactions on Neural Networks*, 15(6), 1350–1366. <https://doi.org/10.1109/tnn.2004.831198>
- Chiang, Y. M., Chang, L. C., & Chang, F. J. (2004). *Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling*. *Hydrol.* 297–311.
- Choi, D., Shallue, C. J., Nado, Z., Lee, J., Maddison, C., J., & Dahl, G. E. (2019). *On Empirical Comparisons of Optimizers for Deep Learning*. Addis Ababa.
- Conneau, A., Schwenk, H., Barrault, L., & Lecun, Y. (2017). Very Deep Convolutional Networks for Text Classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, (1). Long Papers, Valencia, Spain. <https://doi.org/10.18653/v1/e17-1104>
- Farzad, A., Mashayekhi, H., & Hassanpour, H. (2019). A comparative performance analysis of different activation functions in LSTM networks for classification. *Springer*, 31(5), 2507–2521. doi:10.1007/s00521-017-3210-6
- Higginbotham, J. (2015). *Designing Great Web APIs*. O'Reilly Media, Inc.
- Keras (2023). *Conv1D layer*. [https://keras.io/api/layers/convolution\\_layers/convolution1d/](https://keras.io/api/layers/convolution_layers/convolution1d/)
- Kingma, D. P., & Ba, J. A. (2014). A Method for Stochastic Optimization. *ICLR*. <https://www.kaggle.com/datasets/ishantjuyal/emotions-in-text> <https://doi.org/10.48550/arXiv.1412.6980>
- Kirsch, K. (2023). *What Is Brand Perception? How to Measure It and 4 Examples*. Hubspot. <https://blog.hubspot.com/marketing/what-is-brand-perception>. July 05, 2023.
- LeRoy Elliott D. (1998). A better Activation Function for Artificial Neural Networks. *Technical Research Report* (3).
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning--based Text Classification. *A Comprehensive Review*. *ACM Computing Surveys*. 54(3), 1–40. <https://doi.org/10.1145/3439726>
- Nistor, S. C., Moca, M., Moldovan, D., Oprean, D. B. & Nistor, R. L. (2021). *Building a Twitter Sentiment Analysis System with Recurrent Neural Networks*. *Sensors*. <https://doi.org/10.3390/s21072266>



- Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training Recurrent Neural Networks. *Proceedings of the 30th International Conference on Machine Learning*. Atlanta GA.
- Sheng, T. K., Socher, R. D., & Manning, C. (2015). *Improved Semantic Representations from Tree-Structured Long Short-Term Memory Networks*. Cornell University. <https://doi.org/10.48550/arXiv.1503.00075>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958.
- Statista (2021). *Distribution of Twitter users worldwide as of April 2021, by age group*. <https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/>
- Zhu, X., Sobhani, P., & Guo, H. (2015). Long Short-Term Memory Over Recursive Structures. *Proceedings of Machine Learning Research*, (37), 1602–1612.

#### References

- Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems*, (108). <https://doi.org/10.1016/j.knosys.2016.05.040>
- Chandra, P., & Singh, Y. (2004). Feedforward Sigmoidal Networks—Equicontinuity and Fault-Tolerance Properties. *IEEE Transactions on Neural Networks*, 15(6), 1350–1366. <https://doi.org/10.1109/tnn.2004.831198>
- Chiang, Y. M., Chang, L. C., & Chang, F. J. (2004). *Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling*. *Hydrolog.* 297–311.
- Choi, D., Shallue, C. J., Nado, Z., Lee, J., Maddison, C., J., & Dahl, G. E. (2019). *On Empirical Comparisons of Optimizers for Deep Learning*. Addis Ababa.
- Conneau, A., Schwenk, H., Barrault, L., & Lecun, Y. (2017). Very Deep Convolutional Networks for Text Classification. *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, (1). Long Papers, Valencia, Spain. <https://doi.org/10.18653/v1/e17-1104>
- Farzad, A., Mashayekhi, H., & Hassanpour, H. (2019). A comparative performance analysis of different activation functions in LSTM networks for classification. *Springer*, 31(5), 2507–2521. doi:10.1007/s00521-017-3210-6
- Higginbotham, J. (2015). *Designing Great Web APIs*. O'Reilly Media, Inc.
- Keras (2023). *Conv1D layer*. [https://keras.io/api/layers/convolution\\_layers/convolution1d/](https://keras.io/api/layers/convolution_layers/convolution1d/)
- Kingma, D. P., & Ba, J. A. (2014). A Method for Stochastic Optimization. *ICLR*. <https://www.kaggle.com/datasets/ishantjuyal/emotions-in-text> <https://doi.org/10.48550/arXiv.1412.6980>
- Kirsch, K. (2023). *What Is Brand Perception? How to Measure It and 4 Examples*. Hubspot. <https://blog.hubspot.com/marketing/what-is-brand-perception>. July 05, 2023.
- LeRoy Elliott D. (1998). A better Activation Function for Artificial Neural Networks. *Technical Research Report* (3).
- Miniae, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep Learning--based Text Classification. *A Comprehensive Review*. *ACM Computing Surveys*. 54(3), 1–40. <https://doi.org/10.1145/3439726>
- Nistor, S. C., Moca, M., Moldovan, D., Oprean, D. B. & Nistor, R. L. (2021). *Building a Twitter Sentiment Analysis System with Recurrent Neural Networks*. *Sensors*. (1–24). doi.org:10.3390/s21072266doi.org/10.3390/s21072266
- Ostrovskiy, S. (2021). Twitter – Briefly about the social network of the short format. *ApiX-Drive*. <https://apix-drive.com/ru/blog/useful/twitter-korotko-o-socialnoj-seti-korotkogo-formata> [in Ukrainian].
- Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training Recurrent Neural Networks. *Proceedings of the 30th International Conference on Machine Learning*. Atlanta GA.
- Sheng, T. K., Socher, R. D., & Manning, C. (2015). *Improved Semantic Representations from Tree-Structured Long Short-Term Memory Networks*. Cornell University. <https://doi.org/10.48550/arXiv.1503.00075>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958.
- Statista (2021). *Distribution of Twitter users worldwide as of April 2021, by age group*. <https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/>
- Zhu, X., Sobhani, P., & Guo, H. (2015). Long Short-Term Memory Over Recursive Structures. *Proceedings of Machine Learning Research*, (37), 1602–1612.

Отримано редакцію журналу / Received: 21.11.22

Прорецензовано / Revised: 05.01.23

Схвалено до друку / Accepted: 15.01.23

Olga ZHULANOVA, Bachelor  
ORCID ID: 0000-0001-6231-0600  
e-mail: olga.zhulanova@knu.ua  
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Olena VASHCHILINA, PhD (Phys. & Math.), Assoc. Prof.  
ORCID ID: 0000-0001-6867-6216  
e-mail: olenavashchilina@knu.ua  
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

## ELEMENTS OF NEURAL NETWORKS TECHNOLOGY FOR ANALYZING THE ATTITUDE OF TWITTER USERS TOWARDS BRANDS

**Background.** The article is devoted to the issues of effective organization of collection and information analysis about the attitude of Twitter users to brands in the software application form. Issues such as research into modern means of collecting and analyzing information are considered; definition of the functionality that the application should implement; analysis of architectural solutions and selection of software necessary for its implementation.

**Methods.** When conducting research, marketing theory is used in the field of collecting information about consumer opinions, research on methods of information analysis for the purpose of classifying consumer mood, empirical analysis and synthesis of architectures used in the creation and comparison of neural network models for text classification, development and construction of own model for classification.

**Results.** As part of the task of software implementation of tweet text analysis, the architecture of convolutional and recurrent neural networks was investigated, a comparison of various hyper parameter values of neural networks was made, in particular, activation functions, loss functions, the number of learning epochs, the number of network layers, a comparison of different Python libraries for processing natural languages in the context of tweet evaluation.

**Conclusions.** The practical significance of the study is the creation of a software tool for effective analysis of Twitter users' attitudes towards brands, which can serve to improve the effectiveness of marketing activities of brands.

**Keywords:** Twitter, information analysis, text classification, sentiment analysis, neural networks, software application, Python libraries.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.