ІНФОКОМУНІКАЦІЙНІ ТЕХНОЛОГІЇ ТА ЕЛЕКТРОННА ІНЖЕНЕРІЯ
INFORMATION AND COMMUNICATION TECHNOLOGIES, ELECTRONIC ENGINEERING

№ 5 (1), 2025

ЗАСТОСУВАННЯ МАШИННОГО НАВЧАННЯ ДЛЯ АНАЛІЗУ НАСТРОЇВ КОРИСТУВАЧІВ В ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ СИСТЕМАХ

М. Брич [ORCID: 0009-0004-9767-5991], **О. Коваліско** [ORCID:0009-0007-7035-8429], **І. Балас** [ORCID: 0009-0002-3762-1006]

Національний університет «Львівська політехніка», вул. С. Бандери, 12, 79013, Львів, Україна

Відповідальний за рукопис: Микола Брич (e-mail: mykola.v.brych@lpnu.ua)

(Подано 23 березня 2025)

У статті розглядаються сучасні методи застосування машинного навчання та рекомендаційних систем для аналізу настроїв користувачів в інформаційно-комунікаційних середовищах. Соціальні мережі та цифрові платформи стали важливими джерелами громадської думки, щодня генеруючи великі обсяги текстових даних. Традиційні методи аналізу, такі як словникові методи або класичні алгоритми машинного навчання, мають обмеження щодо визначення контексту, сарказму, сленгу та емоційних відтінків тексту. Це ускладнює точне визначення емоцій користувачів і соціально значущих тем. У зв'язку з цим у дослідженні запропоновано ефективну модель, яка поєднує метод BERT (Bidirectional Encoded Representation from Transformers) і рекомендаційний алгоритм для поглибленого аналізу текстових даних. Розроблена модель не лише класифікує емоційне забарвлення тексту, а й визначає ключові теми, які набувають соціального резонансу, тому може швидко адаптуватися до динамічних змін інформаційного середовища. Запропонований підхід спрощує автоматичний моніторинг громадської думки, персоналізацію інформаційних потоків та ефективне управління неструктурованими даними. Результати дослідження підтвердили ефективність розробленої системи: точність моделі поступово підвищувалася в процесі навчання - від 60% на початковому етапі до понад 98% на кінцевому для навчальних даних. Для тестових даних точність класифікації досягла 100%, що свідчить про те, що модель має високу здатність до узагальнення інформації та низьку ймовірність помилки. Мінімізація функції втрат доводить ефективність процесу навчання та надійність запропонованого алгоритму. Інтеграція моделей BERT в комунікаційні та інформаційні системи пропонує широкі можливості для автоматизованого аналізу текстових даних. Такий підхід не тільки покращує якість аналізу контенту, але й забезпечує швидке виявлення соціально цікавих тем, що особливо важливо для соціальних платформ, аналізу медіа та цифрових комунікацій. Використання запропонованої моделі може значно підвищити ефективність управління інформаційними потоками, особливо у сферах штучного інтелекту, автоматичного аналізу громадської думки та моніторингу соціальних подій. Додаткові дослідження можуть бути зосереджені на розширенні можливостей моделі для аналізу багатомовного контенту, покращенні її адаптації до нових стилів письма та покращенні обробки коротких, неповних або неформальних текстів. У майбутньому запропонований підхід може бути застосований для автоматизованого управління великими обсягами даних, що сприятиме розвитку інтелектуальних інформаційних сервісів та кращому розумінню особливостей соціальних взаємодій у цифровому середовищі.

Ключові слова: *машинне навчання, рекомендаційні системи, модель BERT, аналіз контенту*
УДК: 004.7

Вступ

У сучасному цифровому суспільстві соціальні мережі відіграють ключову роль у формуванні громадської думки, поширенні інформації та визначенні актуальних суспільних тем. Щодня мільйони користувачів взаємодіють онлайн, залишаючи текстові відгуки, емоції та реакції на події. Такі дані можуть стати цінним ресурсом для аналізу настроїв і загальних тенденцій, що потребує застосування сучасних методів обробки природної мови.

Проте традиційні методи аналізу тексту – наприклад, засновані на словниках або класичних алгоритмах машинного навчання – мають свої обмеження. Вони часто не враховують складні мовні конструкції, контекст, сленг і сарказм. Також вони не завжди здатні вчасно визначати суспільно важливі теми в умовах швидкого потоку інформації. З огляду на ці виклики, дане дослідження пропонує ефективну модель машинного навчання, яка поєднує метод BERT та систему рекомендацій для аналізу тону тексту в соціальних мережах і виявлення ключових суспільних тем. Такий підхід не лише допомагає класифікувати емоційне забарвлення текстів, а й визначати теми, що викликають значний суспільний відгук. Це особливо важливо для інформаційних і комунікаційних систем, які працюють із великими масивами неструктурованих даних у режимі реального часу. Запропонована модель здатна швидко адаптуватися до нових даних, що критично важливо для аналізу текстових потоків. Її інтеграція з інформаційно-комунікаційними технологіями дозволяє автоматизувати моніторинг громадської думки, покращити персоналізацію інформаційних послуг та ефективно керувати потоками даних. Подальший розвиток моделі відкриває перспективи її застосування у сферах медіа-аналізу, соціальних досліджень, штучного інтелекту та автоматизованого управління інформаційними системами [1-4].

2. Аналіз проблем та постановка задачі досліджень

Соціальні мережі перетворилися на потужний інструмент комунікації, що не лише об’єднує людей, а й надає великий масив даних про громадську думку, соціальні настрої та актуальні суспільні проблеми. Однак аналіз цієї інформації за допомогою традиційних методів залишається складним і ресурсозатратним процесом. Саме тому зростає потреба в автоматизованих системах, здатних ефективно оцінювати емоційний тон публікацій, коментарів і взаємодій користувачів у соцмережах. Це дозволить здійснювати оперативний моніторинг громадської думки та швидко реагувати на важливі події.

Аналіз настроїв у соціальних мережах є одним із ключових завдань у сфері обробки природної мови (Natural Language Processing, NLP) та машинного навчання (Machine Learning, ML). Визначення тональності повідомлень (позитивної, негативної чи нейтральної) допомагає виявляти теми, що викликають суспільний резонанс. Проте, незважаючи на значний прогрес у цій сфері, дослідники все ще стикаються з низкою викликів, зокрема необхідністю підвищення точності аналізу під час обробки великих обсягів неструктурованих даних, багатомовним характером контенту, а також складністю мовних конструкцій і використанням неформальних виразів у соцмережах.

Традиційно для аналізу настроїв застосовували алгоритми машинного навчання, такі як наївний класифікатор Байєса або метод опорних векторів (Support Vector Machine, SVM). Однак ці методи демонструють обмежену ефективність при роботі з великими та неоднорідними даними. Останні досягнення в галузі глибокого навчання, зокрема використання трансформерних моделей, таких як BERT, відкривають нові можливості для покращення точності та адаптивності алгоритмів. Саме тому впровадження моделей класифікації настроїв на основі BERT є важливим кроком до вдосконалення аналізу емоційного забарвлення текстів.

Іншим перспективним напрямом є інтеграція аналізу настроїв із системами рекомендацій, що дозволяє персоналізувати контент для користувачів залежно від їхніх емоційних реакцій.

Дослідження показують, що такі системи підвищують залученість аудиторії та покращують взаємодію з контентом. Водночас ефективність подібних систем залежить від низки технічних і методичних факторів. Одна з основних проблем – недостатня кількість якісних даних для навчання моделей. Обмежений доступ до великих обсягів розмічених даних ускладнює навчання глибоких нейронних мереж і може впливати на точність результатів, особливо при аналізі текстів у реальному часі. Важливим завданням є розробка методів, що дозволять досягати високої точності навіть за умов обмежених навчальних даних.

Використання сленгу, емодзі, хештегів, змішаних мов та неформальних виразів ускладнює традиційні підходи до аналізу. З урахуванням цих викликів основною метою дослідження є створення ефективної моделі машинного навчання для автоматичного аналізу настроїв у соціальних мережах. Така модель повинна забезпечувати високу точність класифікації навіть при обмеженій кількості навчальних даних, інтегруватися з рекомендаційними системами для персоналізації контенту, покращувати ефективність функціонування інформаційно-комунікаційних технологій.

Запропонований підхід сприятиме подальшому розвитку методів аналізу текстових даних у цифровому середовищі та розширенню можливостей автоматизованого моніторингу громадської думки [5,6].

3. Основні поняття аналізу настроїв у соціальних мережах

NLP відіграють ключову роль в аналізі текстових даних, особливо у визначенні емоційного забарвлення повідомлень у соціальних мережах. Одне з основних завдань NLP — перетворення необробленого тексту у формат, який можна аналізувати за допомогою комп’ютерних алгоритмів. Для цього застосовують різні методи, що дозволяють структурувати текст і виділяти з нього змістовні одиниці. Одним із важливих підходів є кодування тексту, яке перетворює слова та фрази в числові вектори, придатні для аналізу машинними моделями. Ще одним ключовим процесом є лематизація — зведення слів до їхніх базових форм (лем). Це допомагає зменшити варіативність мовних конструкцій і підвищити точність аналізу, особливо під час класифікації настроїв або виявлення емоційного забарвлення тексту.

Щоб зробити аналіз ще ефективнішим, важливим етапом є видалення стоп-слів — термінів, які не мають значного емоційного чи змістовного навантаження. Це дозволяє зосередитися на ключових словах, що визначають тональність тексту. У соціальних мережах, де контент змінюється динамічно, наявність зайвих слів може спотворити результати аналізу, тому їхнє вилучення сприяє точнішій інтерпретації даних. Одним із головних завдань NLP у цій сфері є виявлення емоційних відтінків у тексті. Для цього використовують семантичний аналіз, який дозволяє визначати, чи містить текст позитивні, негативні або нейтральні емоційні стани. Цей метод базується на ідентифікації ключових слів і фраз, що мають певний емоційний заряд. Важливим аспектом є також врахування контексту: окремі слова можуть змінювати свій зміст залежно від оточення, у якому вони використовуються. Сучасні контекстні моделі, такі як BERT, значно покращують аналіз настроїв, оскільки здатні враховувати значення слова в конкретному контексті. Це дозволяє точніше визначати тональність повідомлень навіть у випадках, коли окремі слова мають нейтральне значення, але в поєднанні з іншими змінюють емоційний зміст висловлювання. Такий підхід є особливо важливим для аналізу текстів у соціальних мережах, де повідомлення часто містять іронію, сарказм або специфічний сленг. Таким чином, методи обробки природної мови є основою для автоматизованого аналізу настроїв у цифровому середовищі. Вони дозволяють перетворювати неструктуровані дані у формат, придатний для машинного аналізу, що дає змогу швидко й точно виявляти емоційні відтінки у великих обсягах інформації.

Методи класифікації тексту є основою для аналізу настроїв у великих обсягах текстових даних, особливо в соціальних мережах, де користувачі активно висловлюють свої думки та емоції. Класифікація тексту дозволяє визначити тон повідомлення, що є важливим етапом у процесі

аналізу настроїв. Це дає змогу оцінювати емоційний контекст обговорень суспільно значущих тем. Однак для ефективної роботи з великими масивами даних необхідні потужні алгоритми, здатні враховувати мовні особливості, контекст і динаміку публікацій. Традиційні підходи до класифікації тексту базуються на статистичних моделях. Наприклад, наївні байєсівські класифікатори працюють на припущенні, що всі слова в документі незалежні одне від одного, і на цій основі обчислюється ймовірність приналежності тексту до певної категорії. Цей метод добре підходить для простих завдань, однак його ефективність значно знижується при обробці текстів із глибоким контекстом. Іншим поширеним методом є підтримувальні векторні машини (Support Vector Machines, SVM), які працюють шляхом побудови гіперплощин для розділення текстових даних у багатовимірному просторі. SVM здатний ефективно обробляти великі обсяги інформації та враховувати складні текстові патерни. Однак його головним недоліком є потреба у ретельному налаштуванні параметрів моделі та значних обчислювальних ресурсах при роботі з великими наборами даних. Це ускладнює використання SVM у реальному часі для аналізу настроїв у соціальних мережах.

Останніми роками значного поширення набули методи глибокого навчання, які демонструють високу ефективність у класифікації тексту. Одним із найпотужніших інструментів є рекурентні нейронні мережі (Recurrent Neural Networks, RNN), зокрема їхня вдосконалена версія – довготривала короткочасна пам’ять (Long Short-Term Memory, LSTM). Ці моделі здатні аналізувати послідовність слів, враховуючи їхній контекст, що особливо важливо для точного визначення емоційного тону тексту.

Проте найбільш ефективними на сьогодні є трансформерні моделі, зокрема. На відміну від традиційних підходів, BERT аналізує контекст слова з обох боків, що дозволяє точніше розуміти значення фраз у тексті. Це значно покращує результати аналізу настроїв у соціальних мережах, оскільки модель може правильно інтерпретувати навіть складні емоційні нюанси та сарказм. Окрім окремого використання класичних та глибоких методів, усе більшої популярності набувають гібридні моделі, які поєднують традиційні алгоритми (SVM, Naive Bayes) із можливостями нейронних мереж. Такий підхід дозволяє використовувати переваги обох методів: з одного боку – ефективність нейромереж у виявленні складних закономірностей, а з іншого – стабільність і швидкість роботи статистичних алгоритмів. Це особливо корисно в ситуаціях, коли кількість мічених даних обмежена, але висока точність класифікації залишається критично важливою. Методи класифікації тексту відіграють ключову роль у визначенні настроїв користувачів соціальних мереж. Завдяки сучасним алгоритмам глибокого навчання та трансформерним моделям, таким як BERT, точність аналізу значно зросла. Поєднання різних підходів дозволяє створювати гнучкі системи, здатні ефективно працювати з великими обсягами неструктурованих текстових даних. Це відкриває нові можливості для моніторингу громадської думки та глибшого розуміння соціальних процесів у цифровому середовищі.

Незважаючи на значні досягнення в галузі класифікації тексту, існують певні труднощі, зумовлені мовними особливостями, зокрема використанням сленгу, смайликів і нестандартних виразів, характерних для соціальних мереж. Без додаткового налаштування класичні методи, а подекуди навіть сучасні моделі, можуть не впоратися з такими текстами. Це ускладнює процес класифікації, оскільки алгоритми повинні враховувати не лише граматичні правила, а й специфіку мови, яку використовують різні онлайн-спільноти.

З розвитком глибоких нейронних мереж, зокрема моделей Transformer, таких як BERT, точність і ефективність аналізу тексту значно покращилися. Вони дозволяють точніше визначати емоційні відтінки в текстах користувачів, що відкриває нові можливості для автоматизованого аналізу громадської думки. Це особливо важливо в умовах швидкої зміни інформації та контексту в соціальних мережах.

Оцінка емоційних реакцій користувачів є ключовим інструментом для розуміння їхньої взаємодії з контентом у соціальних мережах. Ці платформи пропонують різні способи вираження емоцій – від вподобань і коментарів до пересилань і реакцій у вигляді смайлів. Аналіз таких даних

дозволяє не лише визначати загальний настрій аудиторії, а й оцінювати, наскільки певний контент емоційно резонує з користувачами. Вподобання та пересилання контенту часто свідчать про позитивну реакцію. Користувачі зазвичай реагують на контент, який викликає у них приємні емоції або відповідає їхнім переконанням. У той час як вподобання вказують на загальний позитивний настрій, пересилання можуть свідчити про важливість або актуальність інформації для ширшої аудиторії. Коментарі дають змогу виразити детальніші емоційні реакції. Вони можуть містити як позитивні, так і негативні відгуки, а також передавати більш складні емоційні відтінки – подяку, обурення, захоплення або критику. Для точного аналізу важливо враховувати не лише кількість [6-9]. Змішування різних мов та діалектів у текстах є серйозним викликом для алгоритмів аналізу повідомлень. У соціальних мережах користувачі часто перемикаються між мовами або комбінують кілька мов у межах одного речення, що ускладнює процес розпізнавання змісту. Це явище вимагає розробки адаптивних моделей, здатних правильно обробляти подібні змішані структури. При цьому необхідно враховувати не лише граматичні аспекти, а й культурний контекст, притаманний різним мовним спільнотам.

Одним із можливих підходів до вирішення цієї проблеми є оптимізація існуючих алгоритмів під специфічні умови спілкування в мережі. Важливим кроком є створення технологій, що зможуть розпізнавати та коректно обробляти такі змішані тексти. Це може включати застосування спеціалізованих словників або навчання моделей на базі реальних даних, які містять характерні риси таких текстів. Додатково варто впроваджувати алгоритми, що оцінюють значення слів залежно від контексту, щоб точніше визначати сенс повідомлення та наміри користувачів. Ефективна обробка неформальних текстів є ключовою складовою точного аналізу комунікацій. Розвиток технологій, що зможуть впоратися з такими труднощами, сприятиме створенню більш гнучких та точних систем визначення емоційних відтінків у повідомленнях. Це особливо важливо, оскільки платформи для спілкування відіграють значну роль у формуванні суспільних настроїв. Детальне дослідження таких текстів дає змогу краще розуміти загальні тенденції, відстежувати зміни у громадських дискусіях і покращувати взаємодію між людьми.

Ще одним важливим етапом підготовки тексту до аналізу є вилучення так званих "нейтральних" слів, які не впливають на змістовне навантаження. Ці слова, такі як “та”, “або”, “в”, виконують важливу роль у побудові речень, проте можуть ускладнювати аналіз настроїв, оскільки не несуть суттєвого емоційного забарвлення. Виключення таких елементів допомагає зосередитися на ключових словах, що містять емоційні або смислові акценти, тим самим підвищуючи точність моделей для аналізу текстів. Оптимізація процесу підготовки текстових даних не лише покращує якість аналізу, а й сприяє ефективнішому використанню обчислювальних ресурсів. Враховуючи, що обсяг даних у соціальних мережах постійно зростає, скорочення зайвого тексту без втрати смислового навантаження може значно прискорити процес обробки. Це особливо важливо в ситуаціях, коли необхідно швидко реагувати на зміни у спілкуванні та оперативно аналізувати актуальні теми.

Зрештою, удосконалення підходів до підготовки текстів з урахуванням мовних і культурних особливостей є важливим фактором у підвищенні точності аналізу настроїв. Такий підхід не лише покращує якість прогнозів, а й дозволяє адаптувати результати до особливостей різних груп користувачів. Удосконалення цих методів є невід’ємною складовою розвитку сучасних технологій аналізу комунікацій у режимі реального часу.

4. Дослідження та моделювання алгоритму рекомендацій на основі BERT

Традиційні методи аналізу настрою, такі як Naive Bayes і SVM, мають обмежену точність через відсутність врахування контексту слів і складних мовних конструкцій, таких як сленг або сарказм. Вони також не є ефективними для обробки великих обсягів даних в реальному часі, що є важливим для аналізу соціальних медіа та оптимальної роботи рекомендаційних систем. Зазвичай

ці моделі погано адаптуються до змін у мовних трендах та нових форматах текстів, що знижує їх точність на різних типах даних. Крім того, більшість з них фокусується тільки на виявленні індивідуальних емоційних реакцій, без здатності визначати загальні теми, які викликають широкий інтерес. Більшість традиційних методів також орієнтовані на англомовне середовище, що ускладнює їх використання для аналізу текстів іншими мовами. Інтеграція BERT з рекомендаційними системами вирішує ці проблеми, надаючи можливість виконувати більш точний контекстуальний аналіз, адаптуватися до нових даних та ефективно виявляти важливі соціальні теми, при цьому зберігаючи високу продуктивність навіть за обмеженими обсягами даних.

Наукова новизна запропонованого підходу полягає у поєднанні BERT з рекомендаційними системами, які не лише здійснюють аналіз настроїв, але й дозволяють ідентифікувати соціально значущі теми. Хоча BERT широко використовується для аналізу тональності тексту, цей підхід є інноваційним завдяки інтеграції з рекомендаційними системами для виявлення актуальних соціальних тем. Запропонована модель є високоефективною та адаптивною, що дозволяє їй працювати навіть з обмеженими даними в динамічному середовищі соціальних мереж. Застосовуючи глибший контекстуальний аналіз, система краще враховує культурні та мовні особливості, зокрема сленг і сарказм. На відміну від традиційних методів, які орієнтовані на емоційну оцінку текстів, цей підхід фокусується на виявленні соціально значущих тем та їх поширенні серед користувачів. Таким чином, новизна підходу полягає у його гнучкості, здатності працювати в реальному часі та визначати важливі соціальні тренди, що спрощує обробку даних в інформаційних системах.

У дослідженні BERT, модель, заснована на архітектурі Transformer, є основою для аналізу настроїв у соціальних мережах. Основні переваги BERT у цій галузі:

- Глибоке розуміння контексту: на відміну від класичних методів, таких як LSTM або CNN, BERT враховує контекст кожного слова з обох напрямків (зліва направо та справа наліво), що дозволяє краще інтерпретувати значення тексту.
- Трансферне навчання: BERT можна попередньо тренувати на великих корпусах текстів і потім налаштувати для специфічних завдань, таких як аналіз настроїв. Це дозволяє зменшити потребу у великих обсягах даних і досягти високих результатів швидше.
- Висока точність: моделі, побудовані на архітектурі Transformer, демонструють відмінну ефективність на стандартних наборах даних для обробки природної мови (NLP), зокрема для визначення настроїв у текстах.

Після вибору моделі наступним важливим етапом є налаштування параметрів і тренування на підготовлених даних. Основні етапи цього процесу:

1. Підготовка навчального набору: на цьому етапі здійснюється попередня обробка даних, включаючи токенізацію, нормалізацію та видалення стоп-слів. Потім кожен текстовий фрагмент перетворюється в векторний формат за допомогою спеціальних бібліотек, таких як Hugging Face або TensorFlow.
2. Навчання моделі: на цьому етапі навчальні дані використовуються для тренування BERT, де кожен текстовий зразок (публікація або коментар) класифікується за тоном (позитивний, негативний або нейтральний). Для оптимізації ваг моделі застосовується метод градієнтного спуску.
3. Оцінка ефективності: після завершення навчання проводиться перевірка точності моделі за допомогою стандартних показників (точність, функція втрат тощо).

Блок-схема запропонованого алгоритму рекомендацій на основі BERT представлена на рис.1.

Традиційні методи аналізу настроїв ґрунтуються на двох основних підходах: словникові та машинного навчання. Словникові підходи використовують попередньо створені словники, де кожному слову або фразі надається певне емоційне забарвлення (позитивне, негативне, нейтральне).

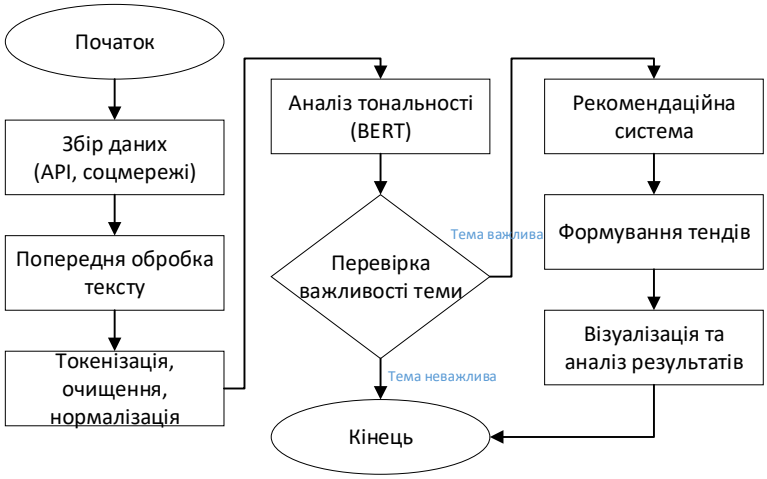

Рис.1. Блок-схема запропонованого алгоритму рекомендацій на основі BERT

Приклади популярних словників: AFINN, SentiWordNet, VADER (спеціально для соціальних мереж). Обмеження: ці методи мають проблеми з обробкою сленгу, сарказму та складних контекстів. Підходи машинного навчання використовують класичні алгоритми машинного навчання (наївний байєсівський класифікатор, SVM, методи логістичної регресії — ефективні, але обмежені у розумінні багатозначності слів).

Основні недоліки традиційних підходів полягають в тому, що вони не враховують контекст слів у реченні. Зі складними мовними явищами, такими як багатозначність, сарказм та сленг, вони працюють погано, а також потребують ручного створення словників або відбору ознак.

Сучасні підходи глибокого навчання значно покращують точність аналізу настроїв, використовуючи нейронні мережі, зокрема трансформерні моделі. Архітектура BERT використовує механізм самоуваги (self-attention), що дозволяє враховувати значення кожного слова в контексті всього речення. Навчена на великих обсягах текстових даних, з можливістю трансферного навчання та забезпечує двосторонній аналіз тексту (зліва направо і справа наліво), що дозволяє глибше розуміти контекст. Перевагами BERT для аналізу настроїв є вища точність порівняно з традиційними методами, ефективність роботи з сарказмом, сленгом і складними емоційними відтінками. Модель автоматично адаптується до нових стилів письма без потреби в ручному створенні словників. Порівняння BERT з класичними методами наведено в таблиці 1.

Таблиця 1

Переваги BERT у порівнянні з класичними методами

Методи	Врахування контексту	Робота з сарказмом та сленгом	Точність
Словникові методи	Ні	Погано	Низька (50-70%)
SVM, Naïve Bayes	Частково	Не враховують	Середня (70-85%)
BERT	Так	Добре	Висока (98-100%)

Для проведеного дослідження були використані відкриті дані з соціальних мереж, зокрема Instagram, з метою оцінки ефективності розробленої моделі. Дані включають пости, коментарі, вподобання та хештеги, що дозволяє аналізувати реакції користувачів на різні теми. Використано Instagram Graph API для отримання загальнодоступних публікацій та коментарів. Інформація була зібрана з публічних профілів за допомогою спеціальних аналітичних інструментів. Дані були відібрані за ключовими словами, хештегами та популярними обліковими записами, що відображають актуальні соціальні теми.

Текст перед аналізом був очищений і нормалізований: видалено стоп-слова, проведена токенізація з урахуванням специфіки мови у соціальних мережах (сленг, абревіатури, емодзі). Аналізувався розвиток настроїв користувачів з часом на основі їхніх реакцій на різні події.

Виконано анонімізацію даних для забезпечення конфіденційності користувачів. Використовувалися тільки відкриті дані, без збору особистої інформації.

Зібрані дані дозволяють протестувати модель у реальних умовах і оцінити її ефективність у виявленні важливих суспільних тем і аналізі настроїв користувачів.

Умови навчання: Дані були розподілені на дві вибірки: навчальна вибірка (70% даних, використовується для тренування та оптимізації моделі) та тестова вибірка (30% даних — служить для перевірки точності моделі та порівняння з іншими підходами (без застосування BERT).

В експериментах використовувалася мова програмування Python і деякі спеціалізовані бібліотеки (Transformer, TensorFlow/Keras, Sklearn, Matplotlib/Seaborn).

Для оцінки продуктивності моделі в дослідженні були використані такі метрики, як точність (accuracy) та функція втрат (loss), що продемонстровано на рис. 2 та рис. 3. На рис. 2 зображено зміни точності в процесі роботи моделі, аналізуючи навчальні та тестові дані. Графік відображає динаміку точності на етапах тренування та тестування. Підвищення точності свідчить про успішне навчання моделі. Коли крива тестування наближається до кривої навчання, це означає, що модель добре узагальнює дані та не страждає від перенавчання.

Точність визначає те, наскільки правильно модель передбачає класи:

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN},$$

де TP (true positive) - кількість правильних позитивних класифікацій; TN (true negative) - кількість правильних негативних класифікацій; FP (false positive) - кількість помилкових позитивних класифікацій; FN (false negative) - кількість помилкових негативних класифікацій.

Accuracy показує частку правильних передбачень серед усіх результатів у вибірці даних.

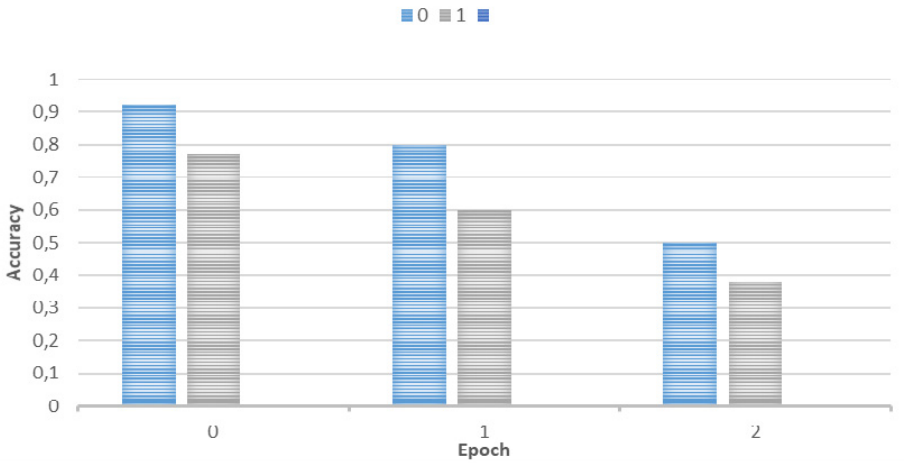

Рис 2. Точність рекомендаційнох моделі на основі BERT під час навчання (Train) та тестування (Val)

Згідно рис.2 точність моделі під час навчання поступово зростає, що свідчить про успішне навчання нейронної мережі. Тестова точність залишається близькою до навчальної. Це означає, що модель добре узагальнює знання і не перенавчається. Для оцінювання рекомендаційної моделі, заснованої на BERT, в дослідженні використовується крос-ентропійна функція втрат (Cross-Entropy Loss), яка визначається як:

$$Loss = -\sum_{i=1}^{N} y_i \log(\hat{y}_i), \quad (2)$$

де y_i – істинне значення відповідності екземпляру вибірки до певного класу (1 або 0), $\hat{y}_i$ – передбачена ймовірність моделі для цього класу, N – кількість зразків у вибірці.

Функція втрат визначає, наскільки передбачені ймовірності відрізняються від реальних значень. Чим нижче значення $Loss$, тим краще модель формує рекомендації на основі даних [9-12]. На рис.3 наведено графік втрат моделі в процесі навчання. Зниження втрат свідчить про ефективність процесу оптимізації.

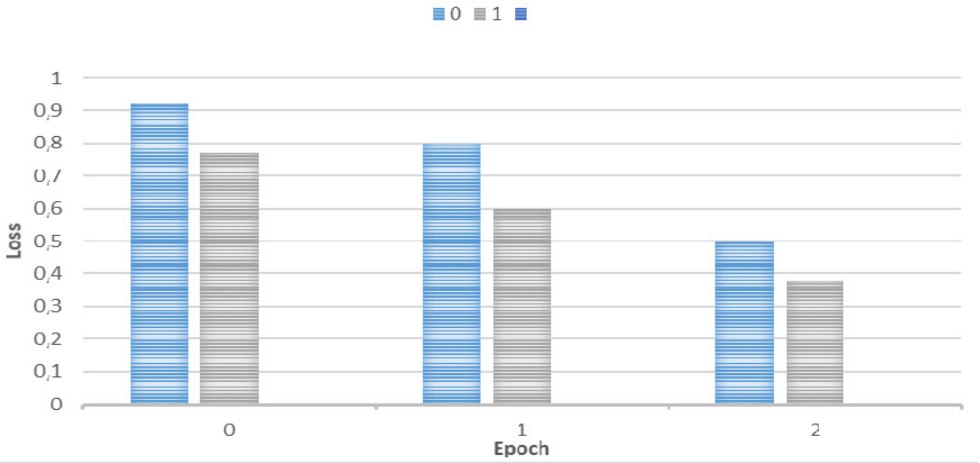

Рис.3. Втрати рекомендаційної моделі на основі BERT під час навчання (Train) та тестування (Val)

Аналіз отриманих результатів (див. рис. 2 та рис. 3) свідчить про високу ефективність розробленого підходу до рекомендацій на основі BERT при оцінці емоційного забарвлення текстів у соціальних мережах. Під час навчання модель демонструвала поступове зростання точності: починаючи з приблизно 60% на початковому етапі, вона перевищила 98% на третьому. Тестування показало максимальний рівень точності, що свідчить про здатність системи адаптуватися до нових даних із мінімальним рівнем помилок. Також простежується стабільне зниження значень функції втрат, що підтверджує якість навчання та надійність роботи алгоритму.

Поєднання високої точності, швидкої адаптації до змін і стабільності функціонування робить цю методику ефективним рішенням для аналізу великих обсягів текстових даних у режимі реального часу. Інтеграція рекомендаційного механізму на основі BERT у комунікаційні та інформаційні платформи сприяє більш якісній обробці контенту, покращенню розуміння інтересів користувачів та автоматичному виявленню актуальних суспільних тем. Це, своєю чергою, дозволяє оптимізувати управління інформаційними потоками та підвищити персоналізацію сервісів відповідно до індивідуальних уподобань аудиторії.

Висновки

Метою цього дослідження є створення ефективної моделі для аналізу текстів у соціальних мережах, яка поєднує сучасні підходи до обробки мовних даних із алгоритмами рекомендаційних систем. Оскільки кількість інформації в цифровому просторі стрімко зростає, особливо в середовищі онлайн-комунікацій, актуальним завданням є розробка алгоритмів, здатних оперативно оцінювати суспільні настрої та виокремлювати ключові теми.

Основні завдання, які вирішуються в цьому дослідженні, включають покращення точності аналізу емоційного забарвлення текстів, адаптацію моделі до змін у мовних конструкціях, забезпечення її ефективності на малих обсягах даних, а також можливість швидкого визначення актуальних дискусійних питань. Отримані результати демонструють високу ефективність запропонованого рішення: модель поступово вдосконалювала свої прогнози під час навчання, досягаючи понад 98% точності на тренувальних вибірках і 100% на тестових даних. Це свідчить про здатність алгоритму не лише точно аналізувати текстовий контент, а й адаптуватися до нових умов без значного зниження якості. Динаміка функції втрат підтверджує, що модель працює стабільно, не схильна до перенавчання та демонструє високу надійність у класифікації.

Впровадження цього підходу у сферу автоматизованого аналізу текстів відкриває широкі можливості для більш глибокої обробки інформації. Це дозволяє покращити розуміння користувацького контенту, виявляти важливі соціальні теми та вдосконалювати персоналізацію інформаційних потоків. Така технологія є особливо корисною для систем, що працюють із великими масивами неструктурованих текстових даних. Завдяки здатності функціонувати в режимі реального часу модель може бути використана для моніторингу змін у громадських дискусіях. Подальший розвиток цієї технології може розширити її застосування в таких сферах, як автоматизовані системи аналізу суспільних процесів, розумні інформаційні сервіси, а також удосконалення взаємодії користувачів із цифровими платформами.

Список використаних літературних джерел

[1]. S. Subramanian, B. Tseng, R. Barbieri and E. N. Brown, "Unsupervised Machine Learning Methods for Artifact Removal in Electrodermal Activity," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 399-402, doi: 10.1109/EMBC46164.2021.9630535.

[2]. T. R. N and R. Gupta, "A Survey on Machine Learning Approaches and Its Techniques:," 2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS), Bhopal, India, 2020, pp. 1-6, doi: 10.1109/SCEECS48394.2020.190.

[3]. B. Wang and W. Zhang, "Research on Edge Network Topology Optimization Based on Machine Learning," 2023 5th International Conference on Applied Machine Learning (ICAML), Dalian, China, 2023, pp. 41-46, doi: 10.1109/ICAML60083.2023.00018.

[4]. B. Walek and P. Fajmon, "A Recommender System for Recommending Suitable Products in E-shop Using Explanations," 2022 3rd International Conference on Artificial Intelligence, Robotics and Control (AIRC), Cairo, Egypt, 2022, pp. 16-20, doi: 10.1109/AIRC56195.2022.9836983.

[5]. A. K. John and B. Thomas, "Recent Trends and Growth in E-Learning Recommender System: A Bibliometric Analysis," 2024 International Conference on Data Science and Network Security (ICDSNS), Tiptur, India, 2024, pp. 1-5, doi: 10.1109/ICDSNS62112.2024.10691270.

[6]. N. Rani and S. L. Chu, "Does the Type of Recommender System Impact Users' Trust? Exploring Context-Aware Recommender Systems in Education," 2023 IEEE International Conference on Advanced Learning Technologies (ICALT), Orem, UT, USA, 2023, pp. 41-43, doi: 10.1109/ICALT58122.2023.00017.

[7]. K. S. Yogi, V. Dankan Gowda, D. Sindhu, H. Soni, S. Mukherjee and G. C. Madhu, "Enhancing Accuracy in Social Media Sentiment Analysis through Comparative Studies using Machine Learning Techniques," 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), Chikkaballapur, India, 2024, pp. 1-6, doi: 10.1109/ICKECS61492.2024.10616441.

[8]. R. Singh and P. Sharma, "An Overview of Social Media and Sentiment Analysis," 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-4, doi: 10.1109/ISCON52037.2021.9702359.

[9]. V. Joseph, C. P. Lora and N. T, "Exploring the Application of Natural Language Processing for Social Media Sentiment Analysis," 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/INOCON60754.2024.10511841.

[10]. A. R. Lubis, Y. Fatmi and D. Witarsyah, "Comparison of Transformer Based and Traditional Models on Sentiment Analysis on Social Media Datasets," 2023 6th International Conference of Computer and Informatics Engineering (IC2IE), Lombok, Indonesia, 2023, pp. 163-168, doi: 10.1109/IC2IE60547.2023.10331232.

[11]. B. Mridula, A. H. Juliet and N. Legapriyadharshini, "Deciphering Social Media Sentiment for Enhanced Analytical Accuracy: Leveraging Random Forest, KNN, and Naive Bayes," 2024 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2024, pp. 1410-1415, doi: 10.1109/ICCSP60870.2024.10543836.

[12]. W. C. Yap, K. K. J. Sing and C. P. Goh, "Restaurant Recommendations with Implicit User Behavior via Image Classification and Sentiment Analysis from Social Media," 2024 3rd International Conference on Digital Transformation and Applications (ICDXA), Kuala Lumpur, Malaysia, 2024, pp. 39-44, doi: 10.1109/ICDXA61007.2024.10470702.

APPLICATION OF MACHINE LEARNING FOR USER SENTIMENT ANALYSIS IN INFORMATION AND COMMUNICATION SYSTEMS

Mykola Brych, Oleksiy Kovalisko, Ihor Balias

Lviv Polytechnic National University, S. Bandery Str., 12, 79013, Lviv, Ukraine

The article examines modern methods of applying machine learning and recommendation systems for sentiment analysis of users in information and communication environments. Social networks and digital platforms have become important sources of public opinion, generating large volumes of textual data daily. Traditional analysis methods, such as lexical approaches or classical machine learning algorithms, have limitations in detecting context, sarcasm, slang, and emotional nuances in the text. This complicates the accurate identification of user emotions and socially significant topics. In this regard, the study proposes an effective model that combines the BERT (Bidirectional Encoded Representation from Transformers) method and a recommendation algorithm for an in-depth analysis of textual data. The developed model not only classifies the emotional tone of the text but also identifies key topics that gain social resonance, allowing it to quickly adapt to dynamic changes in the information environment. The proposed approach simplifies automated public opinion monitoring, personalization of information flows, and efficient management of unstructured data. The research results confirmed the effectiveness of the developed system: the model's accuracy progressively increased during training—from 60% at the initial stage to over 98% at the final stage for training data. For test data, the classification accuracy reached 100%, indicating a high capacity for information generalization and a low probability of error. The minimization of the loss function confirms the efficiency of the training process and the reliability of the proposed algorithm. Integrating BERT models into communication and information systems offers vast opportunities for automated text data analysis. This approach not only improves content analysis quality but also enables the rapid detection of socially relevant topics, which is particularly important for social platforms, media analysis, and digital communications. The proposed model can significantly enhance the efficiency of information flow management, especially in artificial intelligence, automated public opinion analysis, and social event monitoring. Further research may focus on expanding the model's capabilities for multilingual content analysis, improving its adaptation to new writing styles, and enhancing the processing of short, incomplete, or informal texts. In the future, the proposed approach may be applied to the automated management of large volumes of data, contributing to the development of intelligent information services and a better understanding of social interactions in the digital environment.

Keywords: *machine learning, recommendation systems, BERT model, content analysis*