УДК 004.8 DOI https://doi.org/10.35546/kntu2078-4481.2025.4.2.8

В. І. КЛІМЕНКО
асистент кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0001-5869-4269

І. О. БОЯРЧУК
студент кафедри комп'ютерних наук
Хмельницький національний університет

М. О. МОЛЧАНОВА
Ph.D., старший викладач кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0001-9810-936X

О. В. СОБКО
Ph.D., старший викладач кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0001-5371-5788

О. В. МАЗУРЕЦЬ
кандидат технічних наук,
доцент кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0002-8900-0650

ПРОЄКТУВАННЯ АРХІТЕКТУРИ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ НА ОСНОВІ ОБ'ЄКТНО-ОРІЄНТОВАНОГО ПІДХОДУ ДЛЯ РЕАЛІЗАЦІЇ АЛГОРИТМУ НЕЙРОМЕРЕЖЕВОГО ВИЯВЛЕННЯ ЕТНІЧНОЇ ВОРОЖНЕЧІ У ПОВІДОМЛЕННЯХ

У статті розглядається сучасний підхід до нейромережевого виявлення етнічної ворожнечі в цифровому текстовому контенті в умовах великих обсягів даних, що постійно зростають. Запропоновано використовувати технології машинного навчання та обробки природної мови для автоматичного виявлення етнічної ворожнечі у цифрових текстах.

Алгоритм, розроблений у роботі, автоматизує процес аналізу публікацій у мережі для виявлення етнічної ворожнечі та конфліктних висловлювань між етнічними групами. Він складається з двох етапів: попередньої обробки тексту (видалення стоп-символів і смайлів) та передбачення за допомогою моделі FastForest, яка перетворює текст в числову послідовність для подальшого аналізу. Алгоритм генерує відсоткову оцінку ймовірності наявності етнічної ворожнечі в тексті, що полегшує процес виявлення і експертизи таких проявів.

Запропоновано конвеєр обробки та класифікації текстових даних, який включає перетворення текстових повідомлень у числові ознаки та їх об'єднання в вектор характеристик для подальшого аналізу. Цільова змінна перетворюється в числові ключі, після чого застосовується метод OneVersusAll і алгоритм FastForest для навчання моделі на кожному класі окремо, з використанням ансамблю з чотирьох дерев рішень, що підвищує точність і знижує ризик перенавчання. Після навчання модель класифікує новий контент на наявність або відсутність етнічної ворожнечі.

Програмне забезпечення складається з кількох підсистем: обробки даних, виявлення етнічної ворожнечі, попередньої обробки даних, а також допоміжних підсистем для формування навчальної вибірки та навчання моделі. Підсистема формує набір даних для навчання класифікатора, а підсистема навчання моделі відповідає за її налаштування і збереження. Підсистеми були інтегровані в єдину систему для автоматизації виявлення етнічної ворожнечі в цифровому контенті.

Оцінка ефективності алгоритму проводилась за допомогою метрик, таких як MicroAccuracy, MacroAccuracy, LogLoss, ConfusionMatrix, f1-міра та Recall. Результати показали високий рівень точності: MicroAccuracy 0.9890, MacroAccuracy 0.9889, LogLoss 0.0463, з високими значеннями Precision та Recall для обох класів. Модель

успішно класифікувала більшість випадків, передбачивши 4 854 з 4 918 випадків з етнічною ворожнечею.

Отримані результати відкривають нові можливості для створення більш безпечних онлайн-середовищ, що є важливим кроком у боротьбі з дискримінацією та ненавистю в цифровому просторі.

Ключові слова: *етнічна ворожнеча, FastForest, алгоритм нейромережевого виявлення етнічної ворожнечі.*

V. I. KLIMENKO
Assistant at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0001-5869-4269

I. O. BOIARCHUK
Student at the Computer Science Department
Khmelnytskyi National University

M. O. MOLCHANOVA
Ph.D., Senior Lecturer at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0001-9810-936X

O. V. SOBKO
Ph.D., Senior Lecturer at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0001-5371-5788

O. V. MAZURETS
Ph.D. in Engineering Science,
Associate Professor at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0002-8900-0650

DESIGNING OF SOFTWARE ARCHITECTURE BASED ON OBJECT-ORIENTED APPROACH FOR IMPLEMENTATION OF NEURAL NETWORK ALGORITHM FOR DETECTING ETHNIC HATE IN MESSAGES

The article considers a modern approach to neural network detection of ethnic hatred in digital text content in conditions of large, constantly growing amounts of data. It is proposed to use machine learning and natural language processing technologies for automatic detection of ethnic hatred in digital texts.

The algorithm developed in the work automates the process of analyzing publications on the network to detect ethnic hatred and conflicting statements between ethnic groups. It consists of two stages: pre-processing of the text (removing stop characters and emoticons) and prediction using the FastForest model, which converts the text into a numerical sequence for further analysis. The algorithm generates a percentage estimate of the probability of the presence of ethnic hatred in the text, which facilitates the process of detecting and examining such manifestations.

A pipeline for processing and classifying text data is proposed, which includes converting text messages into numerical features and combining them into a vector of characteristics for further analysis. The target variable is converted into numeric keys, after which the OneVersusAll method and the FastForest algorithm are used to train the model on each class separately, using an ensemble of four decision trees, which increases accuracy and reduces the risk of overtraining. After training, the model classifies new content for the presence or absence of ethnic hatred.

The software consists of several subsystems: data processing, ethnic hatred detection, data preprocessing, as well as auxiliary subsystems for forming the training sample and training the model. The subsystem forms a data set for training the classifier, and the model training subsystem is responsible for its configuration and storage. The subsystems were integrated into a single system to automate the detection of ethnic hatred in digital content.

The algorithm's effectiveness was assessed using metrics such as MicroAccuracy, MacroAccuracy, LogLoss, ConfusionMatrix, f1-measure, and Recall. The results showed a high level of accuracy: MicroAccuracy 0.9890, MacroAccuracy 0.9889, LogLoss 0.0463, with high Precision and Recall values for both classes. The model successfully classified the majority of cases, predicting 4,854 out of 4,918 cases with ethnic hatred.

The results open up new opportunities for creating safer online environments, which is an important step in combating discrimination and hatred in the digital space.

Key words: *ethnic hostility, FastForest, neural network algorithm for detecting ethnic hostility.*

Постановка проблеми

Виявлення етнічної ворожнечі у цифровому текстовому контенті є критично важливим завданням в умовах зростання використання соціальних мереж, онлайн-дискусійних платформ та інших цифрових середовищ для комунікації [1]. З поширенням дезінформації, мови ворожнечі та маніпуляційного контенту, особливо у конфліктних і поляризованих суспільствах, зростає потреба у створенні ефективних методів автоматизованого моніторингу та модерації [2, 3].

Етнічна ворожнеча в онлайн-просторі може спричиняти серйозні соціальні наслідки, зокрема радикалізацію, дискримінацію, а також провокувати конфлікти як на національному, так і на міжнародному рівнях [4]. Традиційні методи аналізу та ручної модерації контенту є недостатньо масштабованими через величезні обсяги цифрових даних, що постійно зростають.

Сучасні методи на основі глибинного навчання, обробки природної мови та штучного інтелекту дозволяють автоматизовано ідентифікувати ознаки етнічної ворожнечі у текстах, враховуючи контекст, стилістику та приховані патерни дискримінаційних висловлювань [5, 6]. Використання таких технологій може значно підвищити ефективність виявлення та протидії шкідливому контенту, що сприятиме зменшенню соціальної напруги та захисту прав людини в цифровому просторі.

Ефективне виявлення етнічної ворожнечі у цифровому текстовому контенті потребує не лише якісних алгоритмів нейромережевого аналізу, а й відповідного програмного забезпечення, здатного забезпечити їхню продуктивність, масштабованість та інтеграцію у реальні системи модерації контенту [7].

Таким чином, проєктування архітектури програмного забезпечення є основою для ефективної реалізації алгоритму нейромережевого виявлення етнічної ворожнечі.

Аналіз останніх досліджень і публікацій

Попри активний розвиток моделей глибокого навчання для класифікації текстового контенту, сучасні методи мають низку обмежень, зокрема проблему узагальнення на нові дані та мовну упередженість. Багато існуючих рішень демонструють високу точність на тестових наборах, проте стикаються зі значним зниженням продуктивності у реальних умовах. Це зумовлює необхідність подальшого вдосконалення архітектур нейронних мереж, адаптації трансферного навчання та врахування контекстуальних особливостей мовлення, що є предметом активних досліджень.

У роботі [8] розглядається проблема онлайн-ворожнечі, зокрема hate speech, що виступає деструктивним контентом в Інтернеті та спрямований на пропаганду ненависті або напади на окремих осіб чи групи за їхніми реальними чи уявними характеристиками (етнічність, релігія тощо). Останнім часом завдання автоматичного виявлення такого контенту набуває все більшої актуальності в галузі обробки природної мови. Проте сучасні моделі демонструють недостатню здатність до узагальнення на нові дані. У дослідженні виконано огляд ефективності цих моделей та проаналізовано причини їх обмежень.

Автори [9] акцентують увагу на поширенні ворожнечі в соціальних мережах і пропонують фреймворк для автоматичного виявлення образливих висловлювань в арабськомовних текстах. Для цього було сформовано корпус, зібраний із Twitter, та визначено чотири ключові категорії ворожнечі: релігійну, етнічну, національну та гендерну. Експериментальна перевірка підтвердила високу ефективність розробленого підходу, що перевершує результати попередніх досліджень у цій сфері.

Дослідження [10] присвячене вивченню методів виявлення онлайн-ворожнечі та оцінює ефективність кросмовного навчання для цієї задачі. Запропонований метод ґрунтується на використанні попередньо навчених мовних моделей (BERT та XLM-R) та трансферного навчання для аналізу текстів у малоресурсних мовах. Проведені експерименти для англійської, італійської та португальської мов засвідчили перспективність підходу, продемонструвавши високу точність: для корпусу OffComBr-2 досягнуто F1-вимір 92%, що перевищує результати попередніх методів.

Формулювання мети дослідження

Метою роботи є проєктування архітектури програмного забезпечення, яке забезпечить реалізацію алгоритму нейромережевого виявлення етнічної ворожнечі у цифровому текстовому контенті.

Викладення основного матеріалу дослідження

Алгоритму нейромережевого виявлення етнічної ворожнечі у цифровому текстовому контенті

Алгоритм нейромережевого виявлення етнічної ворожнечі в цифровому текстовому контенті призначений для автоматизованого аналізу текстів, що публікуються в інтернет-просторі, з метою виявлення тнічної ворожнечі або конфліктних висловлювань між різними етнічними групами. Використовуються методи обробки природної мови, зокрема нейронні мережі для класифікації, що здійснюють перетворення вхідних текстових повідомлень на результат у вигляді відсоткової оцінки прояву етнічної ворожнечі в тестовому цифровому контенті [11]. Кроки алгоритму наведені на рис. 1.

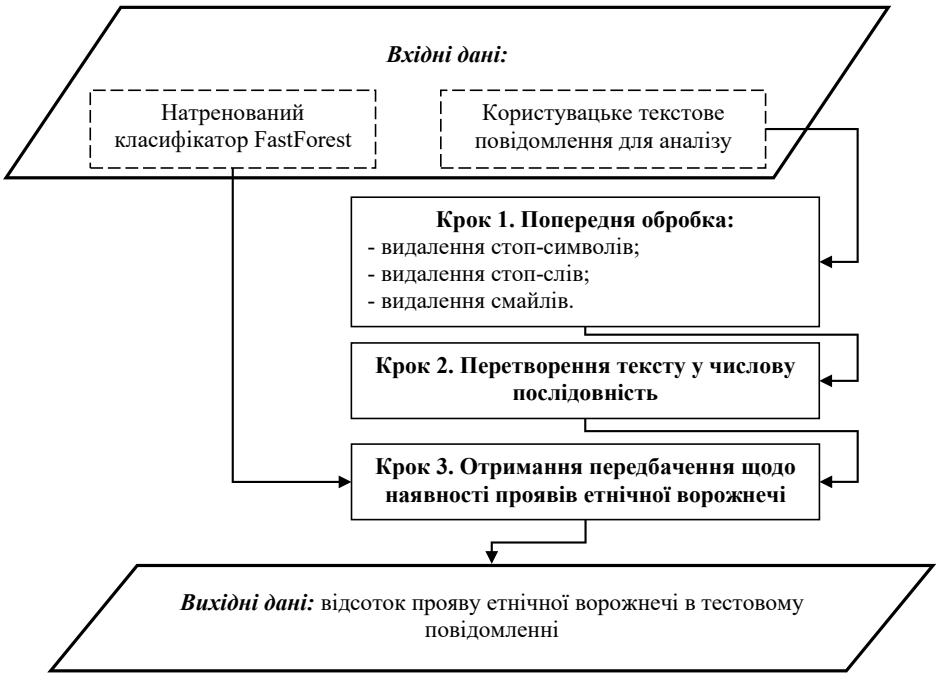

Рис. 1. Схематичне подання алгоритму нейромережевого виявлення етнічної ворожнечі у цифровому текстовому контенті

Вхідними даними алгоритму нейромережевого виявлення етнічної ворожнечі в цифровому текстовому контенті є натренований класифікатор FastForest та цифровий текстовий контент для аналізу. Процес роботи алгоритму складається з двох основних етапів: попередня обробка вхідного тексту (Крок 1) та передбачення за допомогою класифікатора FastForest (Крок 2 та Крок 3).

На першому етапі здійснюється попередня обробка тексту, яка включає видалення стоп-символів, стоп-слів та смайлів [12].

На другому етапі текст перетворюється у числову послідовність, яка є вхідними даними для натренованої моделі машинного навчання FastForest. Далі алгоритм отримує передбачення щодо ймовірності наявності етнічної ворожнечі на основі обробленої числової послідовності, що була подана до моделі. Вихідними даними алгоритму є відсоткова оцінка прояву етнічної ворожнечі в тестовому повідомленні.

Отже, описаний алгоритм забезпечує автоматизований аналіз цифрового текстового контенту з метою виявлення етнічної ворожнечі у текстах, що публікуються в інтернет-просторі, значно спрощуючи процес експертизи таких проявів.

Конвеєр обробки та класифікації текстових даних для виявлення етнічної ворожнечі

Конвеєр обробки текстових даних для виявлення етнічної ворожнечі забезпечує послідовне перетворення текстових повідомлень у формат, придатний для аналізу та класифікації (рис. 2). Спочатку текстові дані конвертуються у числові ознаки, які використовуються для подальшого навчання моделі. Потім відбувається об'єднання отриманих ознак у вектор характеристик, що дозволяє ефективно представляти вхідні дані.

Для коректної роботи алгоритму цільова змінна, що містить мітки класів, перетворюється у числові ключі. Після цього застосовується метод класифікації OneVersusAll, що використовує алгоритм FastForest для навчання моделі на кожному класі окремо. У даному підході використовується ансамбль з чотирьох дерев рішень, що підвищує точність класифікації, мінімізуючи ризик перенавчання, а обмеження максимальної кількості листків у кожному дереві забезпечує прозорість моделі.

Після навчання прогнозовані числові значення повертаються у вихідні мітки. Таким чином, отримана модель здатна аналізувати цифровий текстовий контент, яких не було в навчальному наборі, і здійснювати їх автоматичну класифікацію щодо наявності ознак етнічної ворожнечі.

Проектна архітектура програмного забезпечення для виявлення етнічної ворожнечі та взаємозв'язок компонентів

Архітектура програмного забезпечення для реалізації алгоритму нейромережевого виявлення етнічної ворожнечі в цифровому текстовому контенті складається з трьох основних підсистем: «Підсистема обробки експериментальних даних», «Підсистема виявлення етнічної ворожнечі», «Підсистема попередньої обробки даних»; двох допоміжних підсистем: «Підсистема формування навчальної вибірки з датасету» та «Підсистема навчання моделі машинного навчання», а також загального датасету та робочого набору даних (рис. 3).

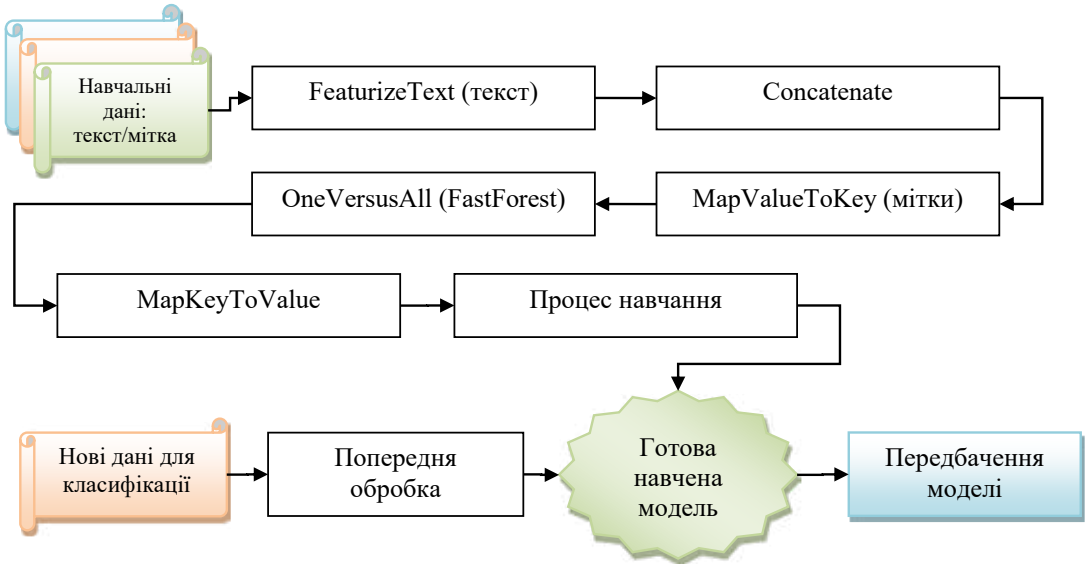

Рис. 2. Конвеєр обробки та класифікації текстових даних для виявлення етнічної ворожнечі

Рис. 3. Проектна архітектура програмного забезпечення для виявлення етнічної ворожнечі

Для забезпечення функціональних можливостей інформаційної системи, яка призначена для виявлення етнічної ворожнечі, була розроблена відповідна діаграма класів [13] (рис. 4).

Підсистема навчання моделі машинного навчання реалізована через клас «BoyarchukModel1», який містить методи для побудови пайплайну, навчання нейромережі та збереження результату. Кінцевим продуктом цієї підсистеми є навчена модель машинного навчання FastForest.

Клас «FormatedDataSet» реалізує функціональність допоміжної підсистеми формування навчальної вибірки з датасету та надає методи для виведення статистики по категоріях в датасеті. Продукт цієї підсистеми – це сформований робочий набір даних, який використовується як вхід для інших підсистем.

Підсистема роботи з експериментальними даними реалізована через класи «Work_with_DataSet» та «DataSetOperation»

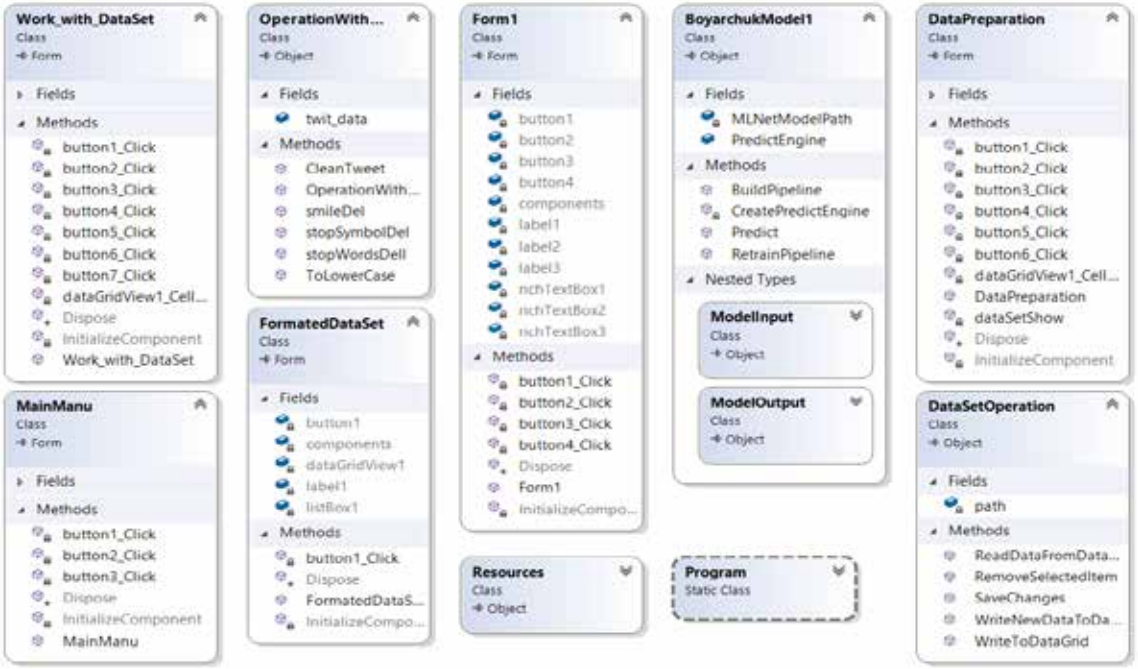

Рис. 4. Діаграма класів програмного забезпечення для виявлення етнічної ворожнечі

Таким чином, була представлена архітектура програмного забезпечення для реалізації алгоритму нейромережевого виявлення етнічної ворожнечі в цифровому текстовому контенті, яка складається з трьох основних підсистем: «Підсистема обробки експериментальних даних», «Підсистема виявлення етнічної ворожнечі», «Підсистема попередньої обробки даних»; двох допоміжних підсистем: «Підсистема формування навчальної вибірки з датасету» та «Підсистема навчання моделі машинного навчання», а також загального датасету та робочого набору даних, який є результатом роботи підсистеми формування навчальної вибірки з датасету.

Дослідження ефективності алгоритму нейромережевого виявлення етнічної ворожнечі у цифровому текстовому контенті спроектованим розробленим програмним забезпеченням

Для всіх експериментів було розроблено програмне забезпечення у вигляді віконного застосунку, для розробки якого було використано платформу.NET 7.0, середовище програмування Visual Studio, та мову програмування C# [14]. Приклад роботи алгоритму нейромережевого виявлення етнічної ворожнечі у цифровому текстовому контенті спроектованим розробленим програмним забезпеченням наведено на рис. 5.

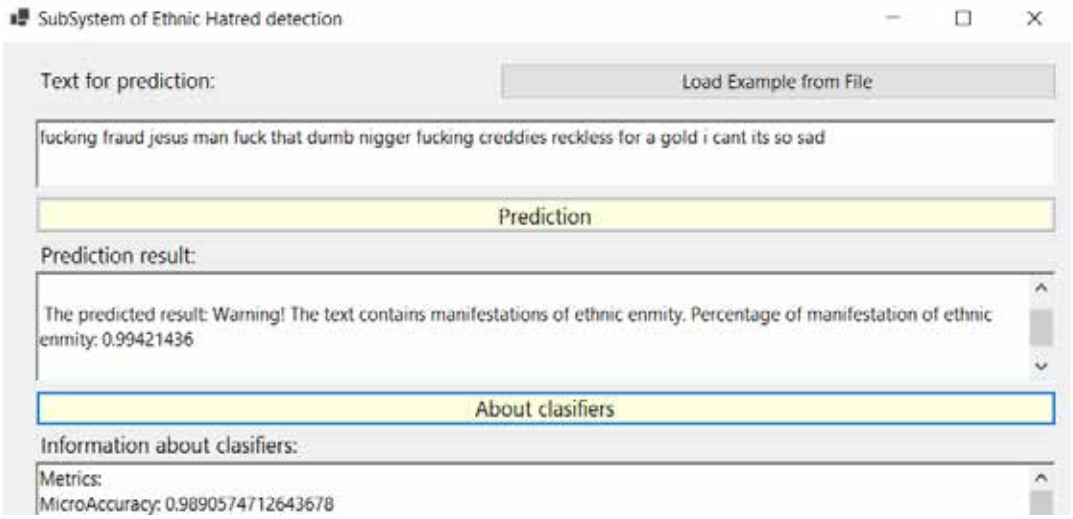

Рис. 5. Приклад виявлення етнічної ворожнечі у цифровому текстовому контенті спроектованим розробленим програмним забезпеченням

Оцінка ефективності алгоритму нейромережевого виявлення етнічної ворожнечі у цифровому текстовому контенті здійснювалась шляхом порівняння результатів роботи алгоритму з валідаційним набором даних, а також виконанням детальної оцінки навченої моделі машинного навчання FastForest. Для цього використовувалися такі метрики, як MicroAccuracy, MacroAccuracy, LogLoss, ConfusionMatrix, f1-міра та Recall, які дозволяють всебічно оцінити ефективність алгоритму з точки зору точності та здатності до виявлення етнічної ворожнечі в цифровому текстовому контенті [15].

За результатами оцінки алгоритму, виконаної за базовим навчальним набором даних, було отримано значення метрик, що свідчать про високий рівень точності та надійності роботи алгоритму. Зокрема, значення MicroAccuracy склало 0.9890, що вказує на високу загальну точність моделі, а MacroAccuracy досягло 0.9889, що свідчить про хорошу здатність алгоритму працювати з усіма категоріями даних. Логарифмічний збиток (LogLoss) становить 0.0463, що також підтверджує ефективність алгоритму в мінімізації помилок під час класифікації.

Зокрема, матриця сплутувань демонструє, що алгоритм правильно класифікує більшість випадків, не містячих ознак етнічної ворожнечі, з кількістю 5 902 правильно передбачених випадків, що складає високий рівень точності для цього класу. Для випадків, що містять ознаки етнічної ворожнечі, модель успішно передбачила 4 854 випадки, з кількістю помилок у 64 випадки. Це дозволяє стверджувати, що алгоритм ефективно справляється з завданням виявлення ознак етнічної ворожнечі в цифрових текстових повідомленнях. Метрика Precision для класу «Не містить ознак етнічної ворожнечі» складає 0.9893, а для класу «Містить ознаки етнічної ворожнечі» – 0.9888. Ця метрика відображає точність класифікації позитивних випадків серед усіх випадків, які модель класифікує як позитивні. Високе значення Precision свідчить про те, що модель здатна правильно відокремлювати класи та мінімізувати кількість помилок при класифікації випадків, що належать до певного класу.

Метрика Recall для класу «Не містить ознак етнічної ворожнечі» дорівнює 0.9908, а для класу «Містить ознаки етнічної ворожнечі» – 0.9870. Recall вимірює, наскільки ефективно модель виявляє всі істинно позитивні випадки в даному наборі даних. Високе значення Recall означає, що модель здатна виявити більшість випадків етнічної ворожнечі в текстах, не пропускаючи їх. Порівняння отриманих метрик наведено на рис. 6.

Загалом, ці метрики вказують на високу точність та здатність моделі до ефективного виявлення як випадків з ознаками етнічної ворожнечі, так і випадків без таких ознак.

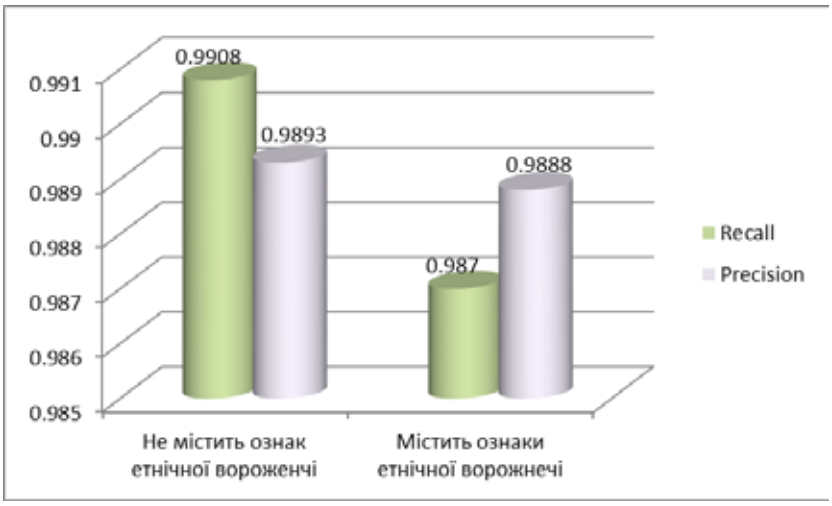

Рис. 6. Оцінка навчених нейромережевих моделей за метриками

Напрями потенційного застосування алгоритму виявлення етнічної ворожнечі та відповідної програмної системи можуть включати платформи соціальних мереж та вебсайти, де такі системи можуть бути використані для автоматичного виявлення та фільтрації образливих, расистських або ненависницьких висловлювань. Це сприятиме створенню більш безпечних онлайн-середовищ, де користувачі зможуть взаємодіяти без загрози для психічного благополуччя. Крім того, отримані результати можуть бути використані дослідниками, соціальними активістами та організаціями, які займаються аналізом та вивченням ознак етнічної ворожнечі. Це дозволить краще зрозуміти мотиви та наслідки таких висловлювань і сприятиме розробці більш ефективних стратегій протидії цим явищам.

Висновки

У статті розглянуто поточний стан нейромережевого виявлення етнічної ворожнечі у цифровому текстовому контенті. Зазначено, що традиційні підходи до аналізу та модерації не здатні справитися з величезними обсягами даних, що постійно зростають. Використання технологій глибинного навчання, обробки природної мови та

штучного інтелекту дозволяє автоматично виявляти ознаки етнічної ворожнечі у текстах, враховуючи контекст та приховані патерни дискримінаційних висловлювань. Такий підхід може значно підвищити ефективність виявлення шкідливого контенту і зменшити соціальну напругу в цифровому просторі. Для забезпечення ефективної реалізації цих алгоритмів необхідно створити відповідне програмне забезпечення, яке забезпечить масштабованість та інтеграцію в реальні системи модерації контенту.

В рамках роботи запропоновано алгоритм нейромережевого виявлення етнічної ворожнечі в цифровому текстовому контенті, який автоматизує процес аналізу текстів, що публікуються в інтернет-просторі, для виявлення проявів етнічної ворожнечі або конфліктних висловлювань між етнічними групами. Він складається з двох основних етапів: попередньої обробки тексту, що включає видалення стоп-символів та смайлів, та передбачення за допомогою натренованої моделі FastForest, яка перетворює текст в числову послідовність для подальшого аналізу. Результатом роботи алгоритму є відсоткова оцінка ймовірності наявності етнічної ворожнечі в аналізованому тексті, що дозволяє автоматизувати виявлення таких проявів у цифровому контенті та спрощує процес їх експертизи.

Запропоновано конвеєр обробки та класифікації текстових даних для виявлення етнічної ворожнечі, який включає послідовне перетворення текстових повідомлень у числові ознаки, які потім об'єднуються в вектор характеристик для подальшого аналізу та класифікації. Спочатку цільова змінна перетворюється в числові ключі, після чого застосовується метод OneVersusAll з алгоритмом FastForest для навчання моделі на кожному класі окремо за допомогою ансамблю з чотирьох дерев рішень, що підвищує точність і знижує ризик перенавчання. Після навчання модель може автоматично класифікувати новий цифровий текстовий контент на наявність або відсутність ознак етнічної ворожнечі, використовуючи прогнозовані числові значення, які повертаються до вихідних міток.

Спроектовано архітектуру програмного забезпечення для реалізації алгоритму нейромережевого виявлення етнічної ворожнечі, що складається з трьох основних підсистем: «Підсистема обробки експериментальних даних», «Підсистема виявлення етнічної ворожнечі» та «Підсистема попередньої обробки даних», а також двох допоміжних підсистем: «Підсистема формування навчальної вибірки з датасету» та «Підсистема навчання моделі машинного навчання». Підсистема формування навчальної вибірки створює робочий набір даних для подальшого навчання класифікатора, тоді як підсистема навчання моделі відповідає за налаштування, навчання і збереження натренованої моделі. В результаті цього процесу утворюється модель FastForest, яка здатна виявляти ознаки етнічної ворожнечі у текстах. Спроектовані підсистеми реалізовано у вигляді застосунку, у якому всі розроблені компоненти працюють у тандемі для досягнення цілей автоматизованого виявлення етнічної ворожнечі в цифровому контенті.

Також було виконано оцінку ефективності розробленого алгоритму нейромережевого виявлення етнічної ворожнечі спроєктованим та розробленим програмним забезпеченням, що проводилась шляхом порівняння результатів роботи алгоритму з валідаційним набором даних за допомогою метрик, таких як MicroAccuracy, MacroAccuracy, LogLoss, ConfusionMatrix, f1-міра та Recall. За результатами оцінки алгоритм продемонстрував високий рівень точності та надійності, з MicroAccuracy 0.9890, MacroAccuracy 0.9889, LogLoss 0.0463, а також з високими значеннями Precision та Recall для класів з та без ознак етнічної ворожнечі. Зокрема, модель успішно класифікувала більшість випадків, правильно передбачивши 4 854 з 4 918 випадків з етнічною ворожнечею, що підтверджує її ефективність. Потенційні сфери застосування алгоритму включають платформи соціальних мереж та вебсайти для автоматичного виявлення образливих або расистських висловлювань, а також його використання дослідниками та соціальними активістами для глибшого аналізу етнічної ворожнечі.

Список використаної літератури

1. Raza Ur Rehman H. M. Detecting hate in diversity: a survey of multilingual code-mixed image and video analysis / H. M. Raza Ur Rehman, M. Saleem, M. Z. Jhandir, E. S. Alvarado, H. Garay, I. Ashraf // Journal of Big Data. 2025. Vol. 12, No. 1. P. 109. https://doi.org/10.1186/s40537-025-01167-w.

2. Okpala E. Large Language Model Annotation Bias in Hate Speech Detection / E. Okpala, L. Cheng // Proceedings of the International AAAI Conference on Web and Social Media. 2025. Vol. 19. P. 1389–1418. https://doi.org/10.1609/icwsm.v19i1.35879.

3. Akram M. H. ISE-Hate: A benchmark corpus for inter-faith, sectarian, and ethnic hatred detection on social media in Urdu / M. H. Akram, K. Shahzad, M. Bashir // Information Processing & Management. 2023. Vol. 60, No. 3. P. 103270. https://doi.org/10.1016/j.ipm.2023.103270.

4. Guan T. Countering Ethnic Minority-Targeted Hate Speech in a Multicultural Society / T. Guan, Y. Yang, T. Liu // Race and Social Problems. 2025. Vol. 17, No. 3. P. 233–248. https://doi.org/10.1007/s12552-024-09426-w

5. Овчарук О.М. Нейромережеве діагностування проявів ПТСР у текстовому контенті з використанням помилко-орієнтованого навчального набору даних / О.М. Овчарук, О.В. Мазурець // Вісник Хмельницького національного університету. Серія: Технічні науки. 2024. № 6, Т. 1 (343). С. 195–200. http://doi.org/10.31891/2307-5732-2024-343-6-30.

6. Залуцька О.О. Метод інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів нейромережевими засобами / О.О. Залуцька, М.О. Молчанова, О.В. Мазурець, О.І. Мельник, Т.К. Скрипник // Вісник Хмельницького національного університету. Серія: Технічні науки. 2023. № 5 (325), Т. 1. С. 67–73. https://doi.org/10.31891/2307-5732-2023-325-5.

7. Молчанова М.О. Підхід до тестування об'єктно-орієнтованих систем керування в електронній комерції / М.О. Молчанова, О.В. Мазурець, П.О. Шевчук, В.І. Кліменко, О.О. Тищенко // Наука і техніка сьогодні. 2025. № 4 (45). С. 1273–1285. https://doi.org/10.52058/2786-6025-2025-4(45)-1273-1285.

8. Yin W. Towards generalisable hate speech detection: a review on obstacles and solutions / W. Yin, A. Zubiaga // PeerJ Computer Science. 2021. Vol. 7. P. e598. https://doi.org/10.7717/peerj-cs.598

9. Alsafari S. Hate and offensive speech detection on Arabic social media / S. Alsafari, S. Sadaoui, M. Mouhoub // Online Social Networks and Media. 2020. Vol. 19. P. 100096. https://doi.org/10.1016/j.osnem.2020.100096.

10. Firmino A. A. Improving hate speech detection using cross-lingual learning / A. A. Firmino, C. de Souza Baptista, A. C. de Paiva // Expert Systems with Applications. 2024. Vol. 235. P. 121115. https://doi.org/10.1016/j.eswa.2023.121115.

11. Мазурець О.В. Метод виявлення депресивного стану пов'язаного із навчанням у закладах освіти із використанням нейромережі дуальної архітектури / О.В. Мазурець, І.А. Тимофієв, В.І. Кліменко, О.О. Тищенко // Вісник Херсонського національного технічного університету. 2024. № 4 (91). С. 311–318. https://doi.org/10.35546/kntu2078-4481.2024.4.41.

12. Віт Р.В. Підхід до тематичної класифікації текстової інформації засобами обробки природної мови / Р.В. Віт, О.В. Мазурець // Наукові праці Донецького національного технічного університету. Серія: Проблеми моделювання та автоматизації проєктування. 2025. № 1 (21). С. 94–99. https://doi.org/10.31474/2074-7888.

13. Собко О.В. Особливості програмної інженерії та тестування програмного забезпечення для нейромережевого аналізу фотоданих залишків зруйнованих будівель із роботизованої техніки / О.В. Собко, В.І. Кліменко, О.В. Мазурець, О.О. Залуцька, О.В. Гладун // Наука і техніка сьогодні. 2025. № 4 (45). С. 1566–1581. https://doi.org/10.52058/2786-6025-2025-4(45)-1566-1581.

14. Молчанова М.О. Алгоритм виявлення аб’юзивного вмісту в україномовному аудіоконтенті для імплементації в об'єктно-орієнтовану інформаційну систему / М.О. Молчанова, О.В. Мазурець, О.В. Собко, Р.В. Віт, В.В. Назаров // Вісник Хмельницького національного університету. Серія: Технічні науки. 2024. № 1 (331). С. 101–106. https://doi.org/10.31891/2307-5732-2024-331-17.

15. Sathyanarayanan S. Confusion matrix-based performance evaluation metrics / S. Sathyanarayanan, B. R. Tantri // African Journal of Biomedical Research. 2024. Vol. 27, No. 4S. P. 4023–4031. https://doi.org/10.53555/AJBR.v27i4S.4345.

References

1. Raza Ur Rehman, H. M., Saleem, M., Jhandir, M. Z., Alvarado, E. S., Garay, H., & Ashraf, I. (2025). Detecting hate in diversity: a survey of multilingual code-mixed image and video analysis. Journal of Big Data, 12(1), 109. https://doi.org/10.1186/s40537-025-01167-w.

2. Okpala, E., & Cheng, L. (2025). Large Language Model Annotation Bias in Hate Speech Detection. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 19, pp. 1389-1418). https://doi.org/10.1609/icwsm.v19i1.35879.

3. Akram, M. H., Shahzad, K., & Bashir, M. (2023). ISE-Hate: A benchmark corpus for inter-faith, sectarian, and ethnic hatred detection on social media in Urdu. Information Processing & Management, 60(3), 103270. https://doi.org/10.1016/j.ipm.2023.103270.

4. Guan, T., Yang, Y., & Liu, T. (2025). Countering Ethnic Minority-Targeted Hate Speech in a Multicultural Society. Race and Social Problems, 17(3), 233-248. https://doi.org/10.1007/s12552-024-09426-w.

5. Ovcharuk O.M., Mazurets O.V. (2024) Neiromerezheve diahnostuvannia proiaviv PTSD u tekstovomu kontenti z vykorystanniam pomylko-oriientovanoho navchalnoho naboru danykh [Neural network diagnostics of PTSD manifestations in textual content using an error-oriented dataset]. Visnyk Khmelnytskoho natsionalnoho universytetu. Seriia: Tekhnichni nauky, № 6, T.1 (343), pp. 195–200. http://doi.org/10.31891/2307-5732-2024-343-6-30.

6. Zalutska O.O., Molchanova M.O., Mazurets O.V., Melnyk O.I., Skrypnyk T.K. (2023) Metod intelektualnoho analizu emotsiinoi tonalnosti tekstovoi informatsii dlia vyznachennia povedinkovykh namiriv neiromerezhsevymy zasobamy [Method of intelligent analysis of emotional tone in textual information for determining behavioral intentions using neural network tools]. Visnyk Khmelnytskoho natsionalnoho universytetu. Seriia: Tekhnichni nauky, № 5 (325), T.1, pp. 67–73. https://doi.org/10.31891/2307-5732-2023-325-5.

7. Molchanova M.O., Mazurets O.V., Shevchuk P.O., Klimenko V.I., Tyshchenko O.O. (2025) Pidhid do testuvannia obiektno-oriientovanykh system keruvannia v elektronnii komertsi [Approach to testing object-oriented control systems

in e-commerce]. Nauka i tekhnika siohodni, № 4 (45), pp. 1273–1285. https://doi.org/10.52058/2786-6025-2025-4(45)-1273-1285.

8. Yin, W., & Zubiaga, A. (2021). Towards generalisable hate speech detection: a review on obstacles and solutions. PeerJ computer science, 7, e598. https://doi.org/10.7717/peerj-cs.598.

9. Alsafari, S., Sadaoui, S., & Mouhoub, M. (2020). Hate and offensive speech detection on Arabic social media. Online Social Networks and Media, 19, 100096. https://doi.org/10.1016/j.osnem.2020.100096.

10. Firmino, A. A., de Souza Baptista, C., & de Paiva, A. C. (2024). Improving hate speech detection using cross-lingual learning. Expert Systems with Applications, 235, 121115. https://doi.org/10.1016/j.eswa.2023.121115.

11. Mazurets O.V., Tymofiiev I.A., Klimenko V.I., Tyshchenko O.O. (2024) Metod vyiavlennia depresyvnoho stanu poviazanoho iz navchanniam u zakladakh osvity iz vykorystanniam neiromerezhi dualnoi arkhitektury [Method for detecting depression related to studying in educational institutions using a dual-architecture neural network]. Visnyk Khersonskoho natsionalnoho tekhnichnoho universytetu, № 4 (91), pp. 311–318. https://doi.org/10.35546/kntu2078-4481.2024.4.41.

12. Vit R.V., Mazurets O.V. (2025) Pidhid do tematychnoi klasyfikatsii tekstovoi informatsii zasobamy obrobky pryrodnoi movy [Approach to thematic classification of textual information using natural language processing]. Naukovi pratsi Donetskoho natsionalnoho tekhnichnoho universytetu. Seriia: Problemy modeliuvannia ta avtomatyzatsii proektuvannia, № 1 (21), pp. 94–99. https://doi.org/10.31474/2074-7888.

13. Sobko O.V., Klimenko V.I., Mazurets O.V., Zalutska O.O., Hladun O.V. (2025) Osoblyvosti prohramnoi inzhenerii ta testuvannia prohramnoho zabezpechennia dlia neiromerezhsevoho analizu fotodanykh zalyshkiv zruynovanykh budivel iz robotyzovanoi tekhniky [Features of software engineering and testing for neural network analysis of photographic data of destroyed buildings using robotic technology]. Nauka i tekhnika siohodni, № 4 (45), pp. 1566–1581. https://doi.org/10.52058/2786-6025-2025-4(45)-1566-1581.

14. Molchanova M.O., Mazurets O.V., Sobko O.V., Vit R.V., Nazarov V.V. (2024) Alhorytm vyiavlennia ab'iuzyvnoho vmistu v ukrayinomovnomu audiokontenti dlia implementatsii v obiektno-oriientovanu informatsiinu systemu [Algorithm for detecting abusive content in Ukrainian audio content for implementation in an object-oriented information system]. Visnyk Khmelnytskoho natsionalnoho universytetu. Seriia: Tekhnichni nauky, № 1 (331), pp. 101–106. https://doi.org/10.31891/2307-5732-2024-331-17.

15. Sathyanarayanan, S., & Tantri, B. R. (2024). Confusion matrix-based performance evaluation metrics. African Journal of Biomedical Research, 27(4S), 4023-4031. https://doi.org/10.53555/AJBR.v27i4S.4345.

Дата першого надходження рукопису до видання: 15.11.2025
Дата прийнятого до друку рукопису після рецензування: 12.12.2025
Дата публікації: 31.12.2025