

УДК 004.89, 004.912

С. Мостовий; А. Северін, д-р філософії

(Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна)

ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ АНАЛІЗУ ТЕКСТУ НА ВИЯВЛЕННЯ ПРОПАГАНДИ

UDC 004.89, 004.912

S. Mostovyi; A. Severin, PhD.

INTELLIGENT TEXT ANALYSIS SOFTWARE FOR PROPAGANDA DETECTION

Україномовні месенджери та соціальні мережі є одним із провідних каналів поширення як оперативної інформації, так і дезінформації та пропаганди. Значні обсяги текстового контенту ускладнюють ручну модерацію, а наявні програмні рішення не завжди враховують особливості української мови. Це зумовлює потребу у спеціалізованому програмному забезпеченні для автоматизованого виявлення пропаганди.

Метою роботи є розроблення та аналіз програмної системи інтелектуального аналізу тексту на виявлення пропаганди, яка поєднує статистичний підхід TF-IDF з тонально-емоційними ознаками та застосовує алгоритм пасивно-агресивної класифікації. Система орієнтована на обробку потоків україномовних повідомлень та подальшу інтеграцію з сервісами моніторингу інформаційних ресурсів.

Програмне забезпечення реалізовано мовою Python та має модульну архітектуру. Модуль попередньої обробки здійснює очищення тексту від HTML-тегів, посилань, згадок користувачів, зайвої пунктуації, нормалізацію регістру, а також токенізацію і лематизацію за допомогою бібліотеки spaCy, що забезпечує коректну обробку українських словоформ. Модуль формування ознак поєднує TF-IDF-представлення тексту (засоби scikit-learn) з тонально-емоційними характеристиками, обчисленими на основі «Українського тонального словника»: полярність повідомлення, частка емоційно забарвлених слів, довжина тексту.

Модуль класифікації реалізує пасивно-агресивний класифікатор, який добре працює з розрідженими векторами ознак, підтримує онлайн-навчання та стійкий до шумних даних, характерних для коротких повідомлень у месенджерах. Допоміжний модуль відповідає за серіалізацію моделей та векторизаторів, логування результатів, розрахунків основних метрик. Логіка роботи організована у вигляді конвеєра: вхідний текст послідовно проходить етапи попередньої обробки, лематизації, формування TF-IDF та тональних ознак і подається на вхід класифікатору, який повертає бінарне рішення.

Система підтримує два основні режими використання. У режимі командного рядка аналіз окремих повідомлень або файлів з колекцією текстів здійснюється через виклик відповідних скриптів, що дозволяє інтегрувати рішення з чат-ботами або системами моніторингу. В експериментальному режимі передбачені засоби для завантаження корпусу, розбиття даних на тренувальну та тестову вибірки, навчання базової моделі (TF-IDF) та розширеного варіанту (TF-IDF + тональні ознаки), а також порівняння їх показників.

Отже, розроблено програмне забезпечення для виявлення пропаганди в коротких україномовних текстах, яке реалізує гібридний підхід на основі TF-IDF та тональних ознак із використанням пасивно-агресивного класифікатора. Модульна архітектура, застосування бібліотек spaCy, scikit-learn і tone-dict-uk забезпечують гнучкість, розширюваність та придатність системи до інтеграції з зовнішніми сервісами моніторингу інформаційних потоків.