

МЕТОД НЕЙРОМЕРЕЖЕВОГО ВИЗНАЧЕННЯ АВТОРСТВА СТУДЕНТСЬКИХ РОБІТ ЗА СТИЛІСТИЧНИМИ ХАРАКТЕРИСТИКАМИ ТЕКСТІВ

В умовах стрімкої цифровізації освіти та зростання обсягу дистанційного навчання питання забезпечення академічної доброчесності набуває особливої важливості. Студентські роботи дедалі частіше перевіряються на плагіат за допомогою автоматизованих систем, однак вони не завжди здатні ефективно виявляти приховані запозичення, написані сторонніми особами. У таких умовах методи нейромережевого аналізу, засновані на вивченні стилістичних характеристик текстів, відкривають нові можливості для точної ідентифікації авторства.

Дослідження стилю письма студента дозволяє враховувати унікальні мовні та стилістичні патерни, які важко підробити або повторити. Це особливо актуально в контексті боротьби з використанням сторонніх сервісів для написання академічних робіт, а також зростання популярності текстових генераторів, які можуть впливати на якість та достовірність навчальних досягнень.

Крім того, розвиток методів визначення авторства на основі нейронних мереж має значення для підвищення довіри до освітнього процесу, зокрема, у разі дистанційного навчання, де прямий контроль за написанням робіт є обмеженим. Такий підхід дозволяє зменшити ризики шахрайства та сприяти формуванню культури доброчесності серед студентів.

Запропонований метод полягає у використанні нейромережевого аналізу для автоматичного визначення авторства студентських робіт на основі стилістичних характеристик текстів. Спочатку здійснюється попередня обробка тексту, яка включає визначення мови, очищення від стоп-слів і зайвих символів [1]. Цей етап спрямований на підготовку даних до аналізу нейромережею.

Далі проводиться завантаження попередньо навченої моделі машинного навчання, яка спеціалізується на інтелектуальному аналізі тексту [2]. На основі стилістичних характеристик моделі здійснюється процес ідентифікації автора. Нейромережа порівнює

вхідний текст із зразками текстів, що містяться в базі даних, визначаючи найімовірнішого автора.

На завершальному етапі формується висновок щодо приналежності тексту певному автору. Аналіз проводиться з урахуванням не лише основного результату, але й оцінки можливості належності тексту іншим авторам. Цей підхід забезпечує високу точність і автоматизацію процесу визначення авторства.

Оскільки вхідними даними підходу є попередньо натренована неймережа для визначення авторства студентських робіт за стилістичними характеристиками текстів, необхідно спочатку провести навчання цієї неймережевої моделі та оцінити її здатність до інтелектуального визначення авторства на основі стилістичних ознак тексту.

Запропонований метод неймережевого визначення авторства студентських робіт за стилістичними характеристиками текстів призначений для моніторингу зміни поведінки акаунтів, які були зламані, а також для виявлення випадків несанкціонованого використання текстів. Метод полягає у перетворенні текстових даних для визначення авторства за допомогою попередньо натренованої неймережі, яка оцінює приналежність тексту до конкретного автора, а також здійснює порівняння з іншими авторами в базі даних для визначення ймовірності належності тексту до інших учасників.

Для оцінки ефективності методу неймережевого визначення авторства студентських робіт за стилістичними характеристиками текстів було розроблено та протестовано програмне забезпечення. У дослідженні використовувалися дві альтернативні моделі машинного навчання: FastForestOva та sdcaLogisticRegressionOva, які були порівняні між собою. Для порівняння авторських робіт застосовувався попередньо розмічений набір даних, на якому обчислювались основні метрики, зокрема точність, влучність та повнота.

У ході експерименту використовувались тексти 8 авторів, роботи яких були попередньо розмічені, але не брали участі в процесі навчання та тестування моделей. Результати на тестових даних показали, що за макрометрикою точності модель FastForestOva досягла значення 78,12 %, тоді як sdcaLogisticRegressionOva – 74,56 %.

Отримані результати свідчать про високу ефективність методу неймережевого визначення авторства на основі стилістичних характеристик текстів. Модель FastForestOva продемонструвала кращі результати в порівнянні з sdcaLogisticRegressionOva, що вказує на її більшу точність у визначенні авторства серед студентських

робіт. Таким чином, цей метод може бути використаний для ефективного моніторингу змін у стилі написання та виявлення авторських патернів у текстах.

Література:

1. Yurchenko D., Mazurets O., Didur V., Molchanova M. Approach to Using Cloud Services for Visual Analytics of Neural Network Analysis of Texts Emotional Tonality. *The Future of Scientific Discoveries: New Trends and Technologies. Proceedings of the XLVII International scientific and practical conference*. 2024. Pp. 108–113.
2. Shevchuk P., Molchanova M., Mazurets O. Software for Text Messages Reliability Analysis Based on the Machine Learning Models Ensemble. *Proceedings of IV International Scientific and Practical Conference «Innovative research and perspectives of the development of science and technology»*. 2024. Pp. 347–354.