ВИЗНАЧЕННЯ АВТОРСТВА УКРАЇНОМОВНИХ ПИСЬМЕННИКІВ ВИКОРИСТОВУЮЧИ БІБЛІОТЕКИ МОВИ ПРОГРАМУВАННЯ PYTHON

Лупей М.І.

ДВНЗ “Ужгородський національний університет”, м.Ужгород

В роботі проведено аналіз поєднань різних методів векторизації та архітектур штучних нейронних мереж для ефективного визначення авторства україномовних письменників. Для дослідження було створено датасет з більш, ніж 1000 творів 15 україномовних письменників. Для початкової обробки було використано стемінг Ukrainian stemmer. Далі було розглянуто поєднання різних методів векторизації з різними структурами архітектур штучних нейронних мереж [1], які базувались на багатошаровому персептроні (MLP) [2] і використовували алгоритм навчання з вчителем [3]. В якості алгоритму навчання використовувався алгоритм зворотного поширення похибок.

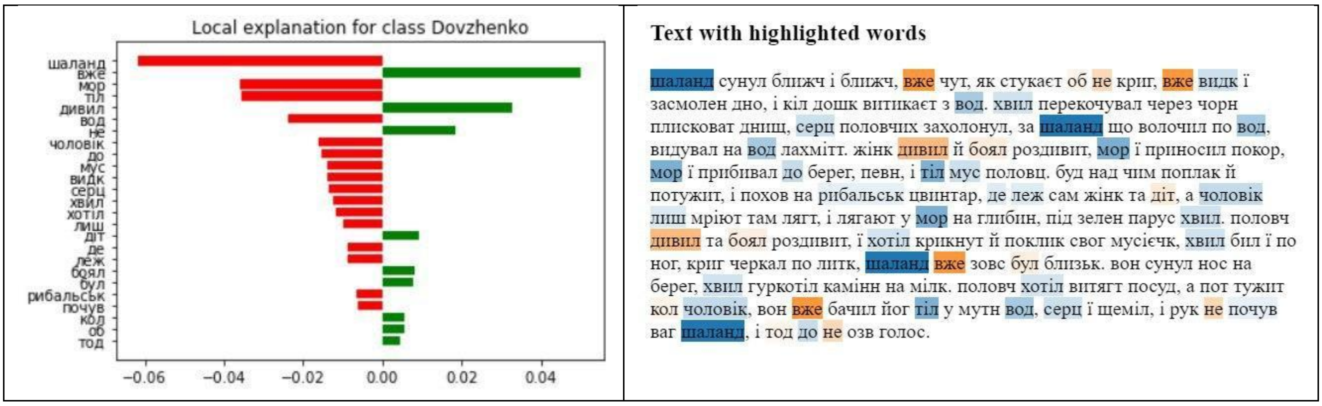

Рис. 1. Використання бібліотек мови програмування Python для знаходження ваг-коефіцієнтів слів при визначенні авторства Довженка

Найбільш ефективні результати були отримані поєднанням векторизатору TfidfVectorizer з архітектурою штучної нейронної мережі C-Support Vector Classification. Вони реалізовані у вигляді бібліотек на мові програмування Python і можуть бути використанні при написанні відповідного програмного продукту. Їх використання дозволило досягнути точність визначення авторства Олександра Довженка 98,54%. Допоміжні бібліотеки, які дозволяють візуалізувати (рис. 1) певні статистичні дані є цікавим емпіричним матеріалом для подальших лінгвістичних студій.

1. Lupei, M., Mitsa, A., Repariuk, V., Sharkan, V. Identification of authorship of Ukrainian-language texts of journalistic style using neural networks. *Eastern-European Journal of Enterprise Technologies*, 2020, 1 (2 (103)). pp. 30–36.

2. Bodyanskiy, Yevgeniy. Computational Intelligence Techniques for Data Analysis. *Leipziger Informatik-Tage*. 2005. pp. 15-36.

3. Лупей М.І., Міца О.В., Шаркань В.В. Використання штучних нейронних мереж в задачі рекрутингу. *V Міжнародна науково-практична конференція "Обчислювальний інтелект (результати, проблеми, перспективи) - 2019"*. Ужгород: ПП "Інвазор". С.253-254.