Fake News detection in Moldova's Information Space

Maria Smarcevscaia, Victoria Bobicev

Technical University of Moldova, maria.smarcevscaia@iis.utm.md,
victoria.bobicev@ia.utm.md, ORCID: 0000-0003-4450-3964, http://lilu.fcim.utm.md/

Abstract. The described research is dedicated to the development of an
automatic fake news detection system. Fake news, also known as
disinformation or misinformation, poses a significant threat to society. Fake
news can erode trust in traditional media outlets, leading people to question
credible sources, they can also undermine trust in government institutions and
democratic processes. Emotionally written fake news can exacerbate social
divisions by spreading false or misleading information to reinforce existing
prejudices [1]. Disinformation can damage the reputation of individuals,
organizations, and even entire countries; they can influence elections by
spreading false or misleading information about candidates or policies
disrupting democratic processes by sowing confusion and distrust.

To combat the problem of fake news, it is essential to promote critical
thinking, media literacy, and fact-checking. Additionally, platforms and
governments need to work together to address the spread of misinformation
and disinformation.

Despite the existence of similar research, there is no adequate system for
detecting fake news in Russian for the Republic of Moldova and neighboring
regions. The complexity of developing such a system lies in the need to create
a representative dataset, which involves identifying reliable sources that do
not publish fake news and finding the fake news itself. The creation of such a
dataset is included in this work [2].

The sites specializing in fact-checking in Russian were selected as sources
of fake news: stopfals.md, stopfake.org, factcheck.kz, veridica.ro,

provereno.media, factcheck.kg, stopfake.kz. As a result, a dataset was formed containing 2108 news articles, of which 614 are fake and 1494 are reliable.

A balanced collection of 614 fake and 614 reliable news articles has been used in the machine learning experiments. The texts were cleaned, tokenized, vectorized and split in train-test parts in proportion 70%-30%. Several algoritms have been tested including Naive Bayes classifier, logistic regression, the k-nearest neighbors method, the support vector machine, and the random forest.

The technologies used in the experiments included various machine learning algorithms such as the naive Bayes classifier, logistic regression, the k-nearest neighbors method, the support vector machine, and the random forest, ensuring high accuracy in fake news detection [3].

The support vector machine and random forest algorithms showed the best results, achieving an accuracy of 90% and 91%, respectively. The error analysis allowed us to study and correct common classification improving the quality and robustness of the models.

The developed system helps protect society from disinformation, which is especially important in the modern world, where information warfare and political manipulation have become commonplace.

References

[1] E. Rahmanian, *Fake news: a classification proposal and a future research agenda*. Spanish Journal of Marketing – ESIC, 2023, №. 1, pp. 60-78.
[2] R. Oshikawa,J. Qitan, W. Wang, A Survey on Natural Language Processing for Fake News Detection, 2018, https://ar5iv.labs.arxiv.org/html/1811.00770.
[3] M.C. Buzea, S. TRAUSAN-MATU, T. Rebedea, Automatic Fake News Detection for Romanian Online News. B: Information. 2022; 13(3):151.