Lithuanian Textual Corpus Development for Propaganda and Propaganda Techniques Detection
## Ieva Rizgelienė

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

CROSS-ANNOTATION PROCESS

Cross-annotation is a method that ensures high-quality data by having multiple annotators independently annotate each document. This process involves dividing the data batch among annotators, with two parts annotated by two individuals. The resulting annotations are then combined, highlighting any conflicts. All annotators collectively review and resolve these conflicts, fostering a collaborative learning environment that improves the consistency and accuracy of future annotations.
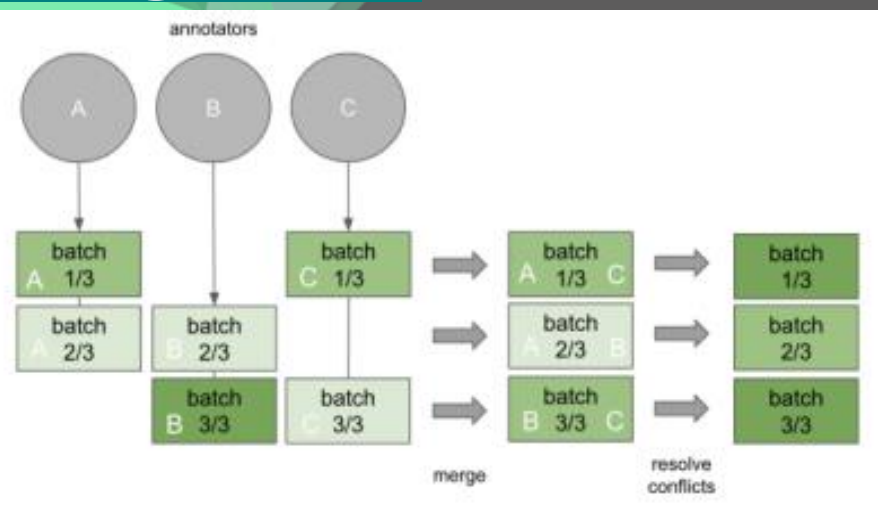
LITHUANIAN TEXTUAL PROPAGANDA CORPORA
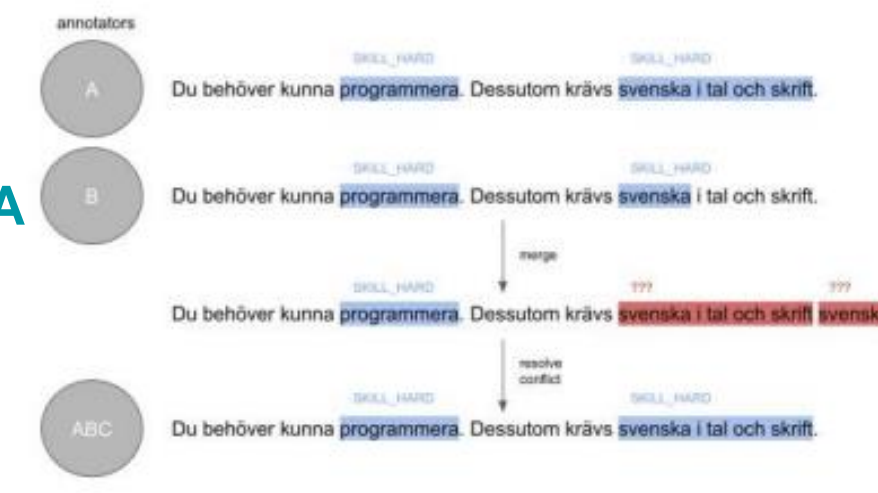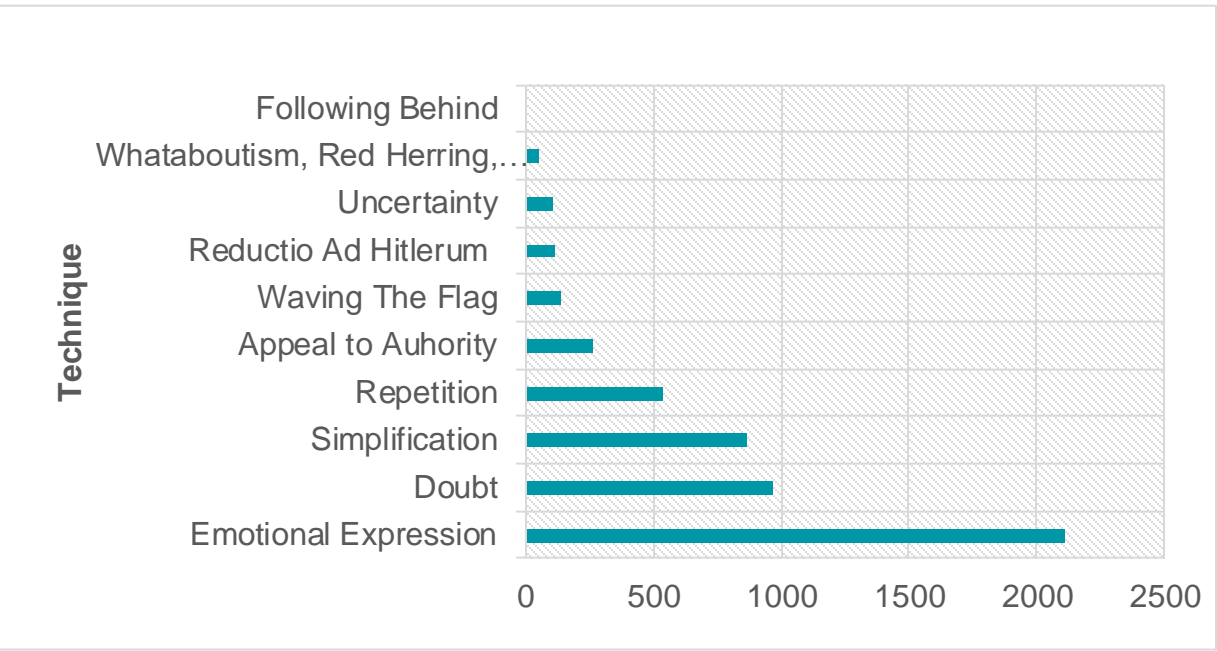
Three experts are involved in a two-stage labeling process: first, they determine if an article is propagandistic; then, they identify specific propaganda techniques within those articles. To date, 1,295 articles have been labeled for binary propaganda detection, and 352 articles for specific propaganda techniques.

MACHINE LEARNING APPLICATION RESULTS

Lithuanian propaganda textual corpora have been used for binary potential propaganda detection tasks. The analysis focused on how balancing techniques influence the performance of traditional machine learning models in detecting propaganda. The models' performance, generalization capabilities, and ability to handle new, unseen data were evaluated in all instances.

Evaluation type	In-domain		Cross-domain	
MODEL	ACC	F1	ACC	F1
LR + Adasyn	**0.89**	**0.89**	**0.72**	**0.71**
SVM + Adasyn	0.89	0.89	0.59	0.51
RF + Adasyn	0.86	0.86	0.61	0.55
XGB + Adasyn	0.85	0.85	0.7	0.69

• Rizgelienė, I., Korvel, G. (2024). Comparative Analysis of Various Data Balancing Techniques for Propaganda Detection in Lithuanian News Articles. In: Lupeikienė, A., Ralyté, J., Dzemyda, G. (eds) Digital Business and Intelligent Systems. DB&IS 2024. Communications in Computer and Information Science, vol 2157. Springer, Cham.
• *Project: 'Propaganda and Disinformation Research: Machine Learning Based Automatic Detection, Impact and Societal Resilience'. Grant number S-VIS-23-8.*