VILNIUS TECH
Faculty of Fundamental Sciences

Peculiarities of GenAI Usage in Higher Education: Possibilities and Additional Challenges

Urtė Radvilaitė, Birutė Pliuskuvienė, Pavel Stefanovič, Simona Ramanauskaitė

ABSTRACT. The fast-growing popularity of generative artificial intelligence (GenAI) has raised new challenges for educational institutions. Higher education students have wide access to those tools and utilise their opportunities for the study process and individual task implementation. Accordingly, the study and student evaluation process requires changes to adapt to the changes and ensure up-to-date technology usage as well as student skill growth. In this research, we overview the main GenAI-raised challenges at the higher institution and teacher level. The biggest attention is dedicated to student cheating prevention and unfair work identification in text-based tasks. A decision support system for Lithuanian written text generation fact automated identification is proposed. Also, the guidelines for the GenAI-generated text detection model development are provided by reflecting feature engineering and data augmentation's effect on the model accuracy and performance.

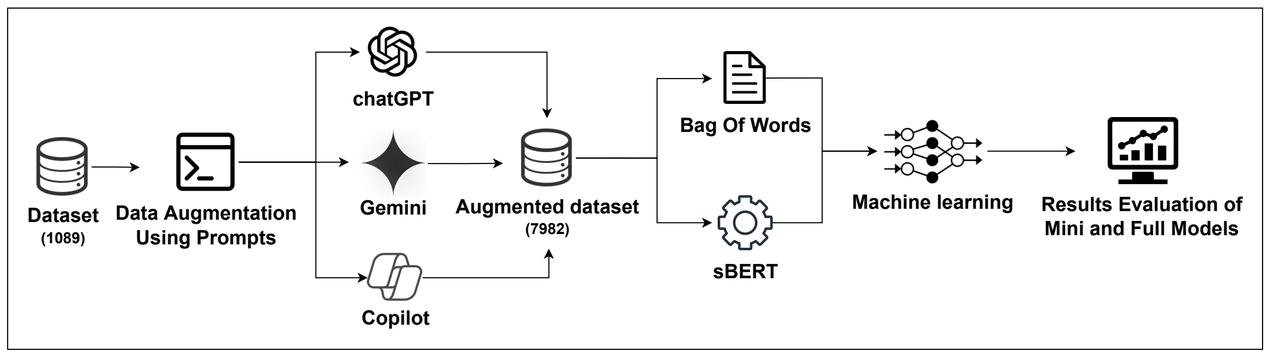


Figure 1. The research methodology.

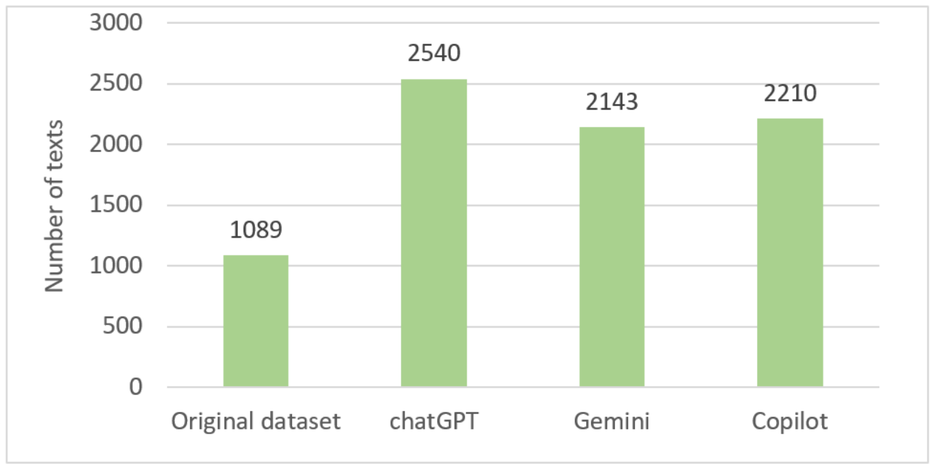


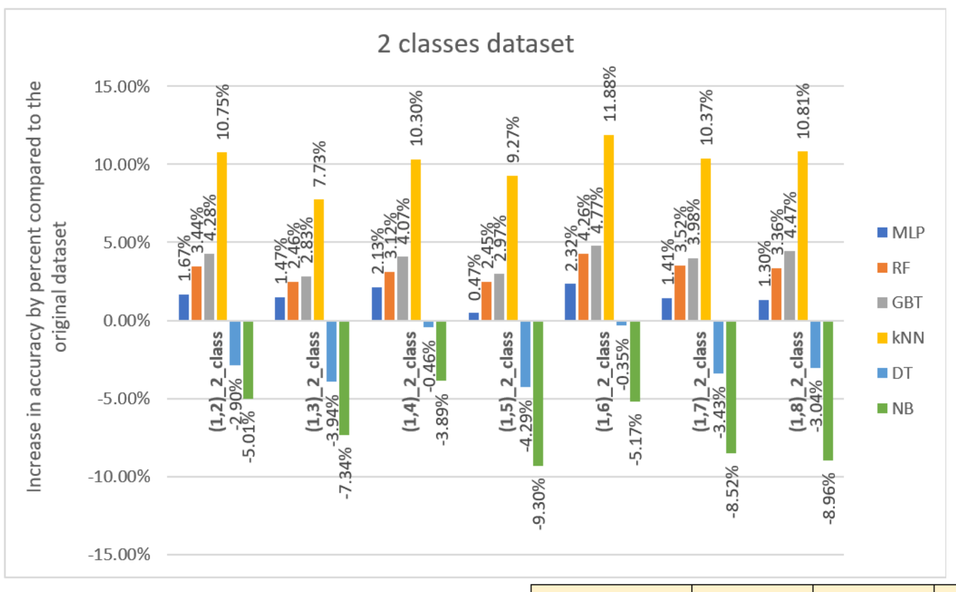
Figure 2. The number of additionally generated texts.

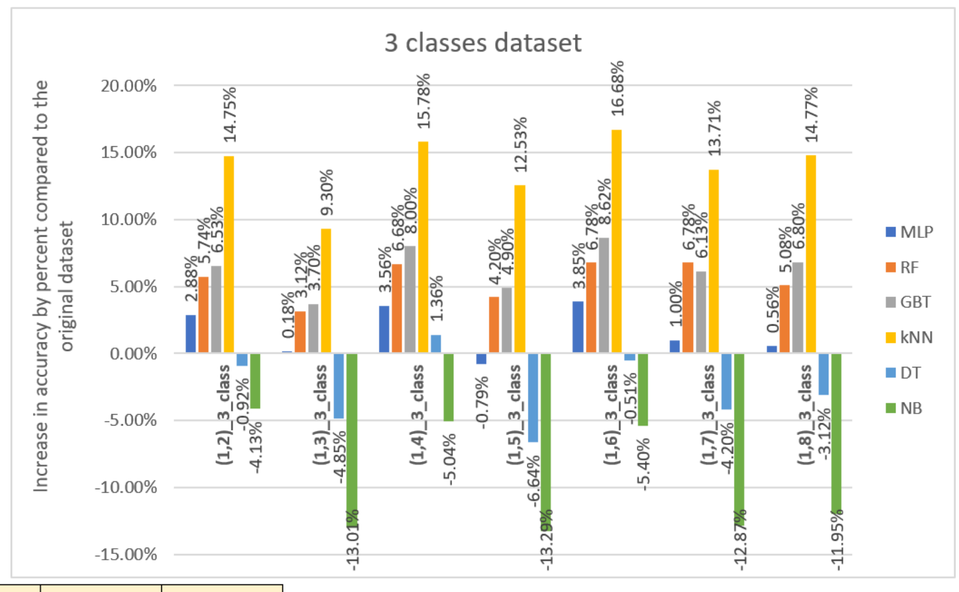
DATASET. The original dataset was collected during the midterm exam of the subject Fundamentals of Data Mining (IV course, information systems study program) at VILNIUS TECH and fully described in the research by Stefanovič et al. [1]. The dataset consists of 1089 texts, 566 of which are original students' answers to given questions (15 different level open-ended questions), 261 - answers paraphrased by various generators, and 262 - direct answers provided by generators. Manually dataset has been augmented using the original dataset as a prompt, where all generators analyzed have been asked to paraphrase the answers several times. Each class has been augmented using one of the text generators: original students' answers, paraphrased answers, and generated answers. The total overall number of additional texts obtained by text generators is presented in Figure 2.

Dataset ID	Original data	Data augmented using:		
		chatGPT	Gemini	Copilot
(1,1)_2_class	+			
(1,1)_3_class	+			
(1,2)_2_class	+	+		
(1,2)_3_class	+	+		
(1,3)_2_class	+		+	
(1,3)_3_class	+		+	
(1,4)_2_class	+			+
(1,4)_3_class	+			+
(1,5)_2_class	+	+	+	
(1,5)_3_class	+	+	+	
(1,6)_2_class	+	+		+
(1,6)_3_class	+	+		+
(1,7)_2_class	+		+	+
(1,7)_3_class	+		+	+
(1,8)_2_class	+	+	+	+
(1,8)_3_class	+	+	+	+

Figure 3. The data subsets used in the experiment.

EXPERIMENTS. The experiments have been performed using two techniques: Bag of Words (BoW) and sBERT embeddings to transform text into vectors. In the case of BoW, the Latent Dirichlet Allocation (LDA) method has been used to reduce the dimension of the vectors. The primary research has shown that the accuracy is not decreased when the size of the dimension is equal to 40. The six machine learning algorithms have been used in the experiments: Multi-layer Perceptron (MLP), Random Forest (RF), Gradient-oosted trees (GBT), k-nearest neighbors (kNN), Decision Trees (DT), and Naive Bayes (NB). Experiments were performed in various ways, where not only one generator was used for text augmentation, but also, different combinations were analyzed. The title of the dataset subsets used in experiments are presented in Figure 3. The title consists of a subset ID and the number how many classes the dataset has: 2 classes - original students' answers and generated answers (the paraphrased answers included); 3 classes - original students' answers, paraphrased answers, and generated answers.During the research, the hyperparameters have been changed using grid search. Total a of 15,296 models have been trained and tested using cross-validation (stratified k-folds are equal to 5).





	MLP	RF	GBT	kNN	DT	NB
(1,1)_2_class	86.87%	87.60%	86.69%	82.28%	85.12%	85.31%
(1,1)_3_class	82.74%	82.83%	81.45%	76.12%	78.05%	78.70%

Figure 4. Results of experimental investigation.

CONCLUSIONS. The experimental investigation has shown that data augmentation using any text generator tool decreases accuracy compared to the original dataset results when the Decision Tree and Naive Bayes algorithms are used and the two classes of datasets are analyzed. A similar tendency can be observed when the three classes of datasets are analyzed. Only once, the accuracy has increase when the dataset has been augmented using the Copilot generator. The highest increase in accuracy has been obtained when the kNN algorithm is used. Comparing the results of dataset augmentation using different text generators, the highest influence had two generators: chatGPT (10.75 % - 2 classes; 14.75 % - 3 classes) and Copilot (10.30 % - 2 classes; 15.78 % - 3 classes). A slightly smaller increase in accuracy has been obtained using the Gemini text generator (7.73 % - 3 classes; 9.30 % - 3 classes). Overall, the highest increase in accuracy was obtained in the case, when both generators Copilot and chatGPT were used for text data augmentation (11.88 % - 3 classes; 16.68 % - 3 classes).

REFERENCES.

[1] Stefanovič, P., Pliuskuvienė, B., Radvilaitė, U., & Ramanauskaitė, S. (2024). Machine learning model for ChatGPT usage detection in students' answers to open-ended questions: Case of Lithuanian language. Education and Information Technologies, 29 (14), p. 18403-18425.