sciendo

DOI 10.2478/sbe-2024-0039
SBE no. 19(2) 2024

NATURAL LANGUAGE PROCESSING METHODS APPLICATION IN DEFENSE BUDGET ANALYSIS

Tetiana ZATONATSKA
Taras Shevchenko National University of Kyiv, Ukraine

Ganna KHARLAMOVA
Taras Shevchenko National University of Kyiv, Ukraine
Lucian Blaga University of Sibiu, Romania

Vadym PAKHOLCHUK
Taras Shevchenko National University of Kyiv, Ukraine

Alim SYZOV
Taras Shevchenko National University of Kyiv, Ukraine

Abstract:

Transferring to economy 5.0 makes a great emphasis on Artificial Intelligence technologies implementation in civil and military areas. The aim of the article is to model relation between the Ukrainian Ministry of Defense budget programs and strategic goals and tasks. The classical budget analysis methodology is extended with NLP technics. The analysis is performed for defense budget program 2101020 - Ensuring the activities of the Armed Forces of Ukraine, training of personnel and troops, medical support of personnel, military service veterans and their family members, and war veterans. Either TF-IDF or more advanced NLP methods are used along with Python libraries and packages. It is found that some goals intersect with each other by semantic similarity. Despite the lack of data, we build proof of concept machine learning model and proved its effectiveness.

Key words: *Natural Language Processing, word embedding, defense budget, defense expenditures, transparency*

1. Introduction

Natural Language Processing (NLP) isn't exactly new and has been around for decades. As digitalization became more and more prominent in our everyday life including but not limited to social media, healthcare, finances, government, etc., as machine learning algorithms became more complex and as our computing capabilities grew, Natural

Language Processing turned into a serious multifaceted tool that can help analyze huge data that's being uploaded every day and use it our advantage.

One field, where NLP plays an important role today is financial and banking systems. In the study Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016) analyzed various NLP algorithms and their practical application. NLP algorithms can retrieve and organize data which can be helpful when working with financial statement content and automating data extraction. One important function that NLP makes possible is fraud detection and prediction, which is paramount in modern banking systems. Authors argue that NLP tools can be beneficial for the government as parsing data using text-based classifiers could be useful for taxonomy creation and alteration.

In another study Valle-Cruz, D., Fernandez-Cortez, V., & Gil-Garcia, J. R. (2022) concluded that with the development of e-government delivery of public services became much easier through the Internet. Not only that but data collected online could be utilized to improve governmental decision-making. Budgeting, being one of the main financial activities of government, has been significantly altered by AI, where tools such as NLP can make budgetary procedures not only more precise based on the data in use but also can make the whole budget system more transparent as well as help improve accountability of governmental institutions.

Further proving this point Davies, J., Arana-Catania, M., Procter, R., van Lier, F. A., & He, Y. (2021, October) in the article discussed a similar issue, where they evaluated NLP application in Scotland budgeting processes and came to a similar conclusion. The authors analyzed participatory budgeting – a practice where citizens can influence how to allocate municipal or public budgets to help empower local communities. Naturally, such a process is quite complex, and upscaling it would only make it harder. The use of NLP algorithms in the digital participatory budgeting process showed promising results which helped by making data for participants clearer and simplifying interaction.

All in all, NLP has proved to be useful in several fields that affect us every day. Finances, auditing, and budgeting on corporate and governmental levels are just a couple of practical applications of such technology, new approaches are being created and improved almost daily, making better data-driven decisions.

Today we see a significant impact of artificial intelligence on the public finance sector. This is due to its effectiveness and increased transparency in the administration of financial resources. Another driving factor is the significant increase in the number of transactions that need to be processed in both structured and unstructured formats. Therefore, understanding the economic nature of expenditures, and comparing them with the goals of organizations can help improve the efficiency of the functioning of public organizations as a whole. Therefore, the task of analyzing costs and comparing them with strategic goals, objectives, and activities is one of the most important budgeting activities, which increases the importance of current research and the necessity of finding effective methods for its solving.

The purpose of this study is the application of machine learning models for defense budget expenditure, in particular, the construction of models for predicting the similarity among expenditures and Ministry of Defense strategic goals and tasks.

The rest of the paper is organized as follows: Section 2 discusses the existing application of the data science in budgeting linking natural language processing and finance. Section 3 reflects the used methodology. Section 4 establish the general theoreticsl concepts of the experiment. Section 5 presents the empirical analysis and tests conducted for the existence of any relation between budget programs and strategic tasks. Finally, Section 6 presents the conclusions and proposals.

2. Literature review

Many works of modern scientists are devoted to the relevance of using big data for the analysis of various processes.

In their study, Loughran and McDonald (2011) investigate the use of textual analysis to measure liabilities in corporate disclosures. They argue that existing word lists used to identify negative words in text are not appropriate for financial contexts and propose an alternative word list based on financial dictionaries and accounting standards. They demonstrate that their word list provides a more accurate and consistent measure of liabilities than existing word lists and can capture variations in reporting behavior across firms and industries. The authors conclude that textual analysis can enhance financial reporting quality and transparency by revealing hidden or understated liabilities.

Li (2010) provides a comprehensive survey of the literature on textual analysis of corporate disclosures. He discusses the motivations and methodology issues for this research area, such as data sources, text processing techniques, textual attributes, and empirical models. Li reviews recent papers that use textual analysis to test economic hypotheses related to disclosure quality, information asymmetry, market efficiency, corporate governance, and litigation risk. He also identifies challenges and opportunities for future research, such as improving text processing methods, incorporating contextual information, exploring new data sources and textual attributes, and addressing endogeneity and causality issues.

The paper by Bochkay (2014) explores how textual information can enhance empirical accounting models. The paper consists of three essays that apply different methods of textual analysis to various accounting topics. The first essay examines how sentiment extracted from financial news articles can improve stock return predictability. The second essay investigates how textual complexity of annual reports affects audit fees and audit delays. The third essay proposes a new approach to measure earnings management using textual disclosures. The paper contributes to the literature by demonstrating the value of textual information for accounting research and practice, and by providing insights into the challenges and opportunities of using textual analysis in empirical accounting models.

In research of Elaine Henry and J. Andrew Leone (2016) conducted a study to evaluate different measures of the tone of financial narrative. The study compared three wordlists commonly used in capital markets research to determine the positivity or negativity of qualitative information in financial disclosure. These included a wordlist developed for financial disclosure, a wordlist from Diction software for general discourse analysis, and a wordlist based on commonly used financial document words. The study found that domain-specific wordlists better predicted market reaction to earnings announcements and had

higher correlations with analyst forecast revisions and dispersion compared to general wordlists. Additionally, the study demonstrated that different methods of aggregating tone scores within a document could affect the performance of tone measures. The paper provides guidance on measuring qualitative information in capital markets research and suggests future research directions.

El-Haj, Rayson, Walker, Young, and Simaki (2019) presented an overview of the current methods and resources for computational analysis of financial discourse. Financial discourse encompasses any text or speech related to financial markets, such as annual reports, earnings calls, and analyst reports. The paper reviewed the steps involved in computational analysis of financial discourse, including corpus creation, preprocessing, annotation, analysis, and evaluation. It discussed the challenges and opportunities associated with each step and provided examples of existing tools and datasets that could be adapted for financial discourse analysis. The paper also identified gaps and limitations in the current research and suggested future research directions.

Luo and Zhou (2020) conducted a literature review on the topic of textual tone in corporate financial disclosures. Textual tone refers to the use of positive or negative language in financial texts that conveys the writer's sentiment. The paper covers various research areas related to textual tone, including its determinants, usefulness for investors and stakeholders, measurement methods, and implications for accounting and finance practice and regulation. The paper also identifies the limitations and challenges of current studies and proposes future research directions. Overall, the paper provides a comprehensive and critical overview of the current state of knowledge on textual tone in corporate financial disclosures and offers guidance and inspiration for interested parties.

Chakraborty and Bhattacharjee (2020) conducted a review and synthesis of automated textual analysis of corporate disclosure, which refers to the information firms communicate to their stakeholders, such as financial statements, annual reports, and press releases. The paper covers various techniques and approaches for automated textual analysis of corporate disclosure, such as dictionary-based, machine learning, deep learning, and hybrid methods. The paper compares and contrasts the advantages and disadvantages of each method and discusses how they have improved the accuracy and efficiency of measuring disclosure tone, which refers to the use of positive or negative language in corporate texts that reflects the firm's sentiment or attitude. The paper offers a comprehensive and updated overview of automated textual analysis of corporate disclosure and identifies gaps and challenges in the current research, proposing future research directions.

The paper by Wujec (2021) introduces a deep learning approach for sentiment analysis of financial texts in current reports. Current reports are documents that firms publish to disclose important events or information that may affect their performance or value. The paper applies deep neural networks (DNNs) to analyze the sentiment of financial texts, and uses the stocks' cumulative abnormal return (CAR) as a proxy for the market reaction to the current reports. The paper compares the DNNs with other methods such as support vector machines (SVMs) and naive Bayes (NB), and evaluates their performance on a dataset of 10,000 current reports from Polish companies. By avoiding manual labeling of data or the creation of dictionaries and rules, the paper contributes to the literature by introducing a new

approach for financial text analysis. Moreover, the paper demonstrates that DNNs outperform other methods in terms of accuracy and correlation with CAR.

The paper presents a method to predict abnormal stock returns by analyzing the text in annual reports of US firms. The method combines financial indicators, readability, sentiment categories, and bag-of-words (BoW) features. The paper shows that sentiment and BoW features significantly impact abnormal stock returns, and their combination improves prediction accuracy compared to using only financial indicators or readability. Additionally, the paper compares different machine learning models and finds that support vector machines perform the best for this task.

In recent research, the use of machine transfer learning methods to analyze accounting disclosures has been introduced and demonstrated (Siano & Wysocki, 2021). Transfer learning is a technique that involves fine-tuning pre-trained models on large generic datasets to analyze small domain-specific datasets. The study specifically utilized BERT, a state-of-the-art language model, and sentiment analysis of quarterly earnings disclosures to demonstrate how transfer learning can enhance the accuracy and efficiency of textual analysis. Additionally, the paper explores the potential benefits and challenges of transfer learning in accounting research and practice.

Loughran and McDonald (2016) conducted a survey of textual analysis methods and applications in accounting and finance. Textual analysis is a technique that utilizes natural language processing to extract information from textual data. The paper discusses the advantages and limitations of textual analysis, as well as the challenges and best practices for implementing it. Moreover, the paper reviews the existing literature on textual analysis in various accounting and finance topics, including earnings announcements, financial disclosures, analyst reports, corporate governance, fraud detection, market efficiency, and investor behavior.

A recent review discusses the use of computer-aided text analysis (CATA) in archival accounting research and offers insights and strategies for accounting systems' scholars (Zhang et al, 2019). CATA is a technique that employs computational tools to analyze textual data from different sources. The paper introduces a model of the corpus linguistic research production process, which comprises four stages: data collection, data preparation, data analysis, and data interpretation. Additionally, the paper examines the primary text archival data sources used in top accounting journals from 2010 to 2016, such as financial reports, press releases, conference calls, social media posts, and news articles. The paper discusses the advantages and challenges of using various text data sources and provides recommendations for future research.

Fernandez-Cortez, Valle-Cruz, and Gil-Garcia (2020) propose a methodology based on artificial intelligence (AI) to optimize the public budgeting process in the Mexican federal government. The paper utilizes a multi-objective genetic algorithm to generate alternative budget scenarios that maximize social and economic development outcomes while minimizing government spending and non-programmed budget items. Furthermore, the paper evaluates the smartness and public value of each scenario using a framework that considers efficiency, effectiveness, equity, transparency, accountability, and citizen participation. The paper demonstrates how AI can assist in optimizing the public budgeting process and offer useful insights for decision-makers.

In a recent study, Zuiderwijk, Chen, and Salem (2021) conduct a systematic literature review and research agenda on the implications of using artificial intelligence (AI) in public governance. The paper defines public governance as the process of steering and coordinating public affairs by various actors, including governments, citizens, businesses, and civil society organizations. The review identifies four primary themes in the literature: AI applications and challenges in government, AI governance frameworks and principles, AI governance mechanisms and instruments, and AI governance outcomes and impacts. Additionally, the paper proposes a research agenda that addresses gaps and limitations in the current literature, such as the lack of empirical studies, comparative analyses, power-related considerations, risk management strategies, performance and impact measurement methods, and impact evaluation approaches.

A literature review found that there are several potential applications of natural language processing methods to the analysis of budgets and reportings, including automatic text summarization and keyword extraction. Additionally, these methods could be used to identify trends in defense spending, detect anomalies and discrepancies in defense spending, and monitor changes in public opinion on military expenditure. Finally, it was suggested that the use of machine learning algorithms could enable the automatic detection of irregularities in defense budgets. In conclusion, natural language processing methods have immense potential to be applied in the analysis of defense budgets, both in the public and private sectors.

Additionally, natural language processing methods could help to compare the defense budgets of different countries and provide more accurate data for the economic evaluation of military activities by governments. Furthermore, natural language processing applications could be used to generate reports on defense budget expenses to better inform legislators and the general public about defense spending. Furthermore, more sophisticated techniques involving natural language processing could be applied to the analysis of defense budgeting by automatically generating documents and aiding with data analysis. To further develop the application of natural language processing in defense budget analysis, context-sensitive methods should be developed to better determine the meaning of military-related texts and extract more detailed information.

3. Methodology

In the current research, the following methods of general scientific research were applied: empirical (familiarization with the basics of Data Science methods), theoretical, observation, experiment, analysis and synthesis, induction and deduction, as well as mathematical methods, statistical information processing methods, and parametric models.

First, a corpus of texts related to defense budgeting was compiled. This corpus included documents such as news reports, articles, speeches, and other sources of related information on the topic.

Next, a software program was developed to analyze the corpus and identify the relevant words and phrases that were relevant to the topic, according to the classification provided. This program employed natural language processing techniques such as

tokenization, lemmatization, and parts-of-speech (POS) tagging to identify and classify the words and phrases.

Using the output from this process, the program analyzed and identified the key elements of defense budgeting including military budgeting analysis, budget execution analysis, budget forecasting analysis, budget optimization analysis, and strategic analysis.

Finally, the system was tested for accuracy using the human expert feedback as we don't know the ground truth about the true labels and similarities in the datasets.

Overall, the methodology used enabled the development of a system that can identify and analyze the key elements of defense budgeting, providing useful information and insights to support decision-making.

4. General theoretical information and the concept

The similarity of texts or documents is one of the most important problems in natural language analysis. Document similarity search is used in many areas such as recommender systems, plagiarism detection, analysis of legal documents, etc.

We can consider two documents to be similar if they are semantically similar and define the same concept or are duplicates.

Metrics. To get a machine learning algorithm to determine the similarity between documents, we need to define a way to mathematically measure the similarity, and it must be comparable so that the algorithm can tell which documents are most similar and which are less. We also need to transform text from documents into a measurable form (or mathematical object, usually vector form) so that we can perform calculations on it.

Thus, turning a document into a mathematical object and determining a measure of similarity are, first of all, the two steps required to make machine learning algorithms perform this task. We propose to consider different approaches to solving this problem.

Some of the most common and efficient ways to calculate similarity are cosine distance, Euclidean, and Jaccard.

Cosine distance. Cosine distance/similarity is the cosine of the angle between two vectors, which gives us the angular distance between the vectors. The formula for calculating the similarity of cosines between two vectors A and B:

$$cos(\theta) = \frac{A \cdot B}{|A| \cdot |B|} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \cdot \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (1)$$

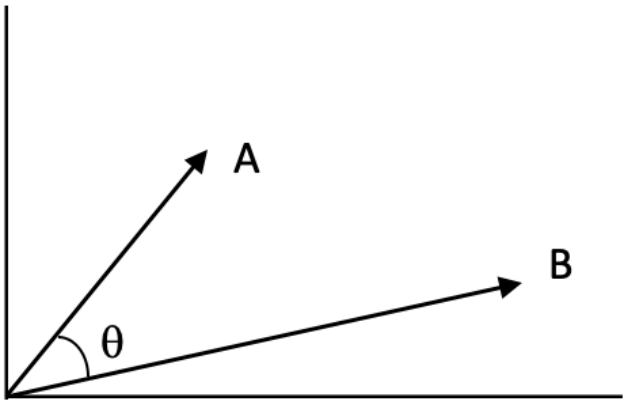

Figure 1: Representing the cosine distance between two vectors in 2D space

The cosine is 1 for theta = 0 and -1 for theta = 180, which means that for two overlapping vectors, the cosine will be the highest and lowest for two exactly opposite vectors. For this reason, it is called similarity. You can think of 1-cosine as a distance.

Euclidian distance. Euclidean distance is one form of Minkowski distance at p=2. It is defined as follows:

$$d(a,b) = d(b,a) = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2} \quad (2)$$

We can use cosine or Euclidean distance if we can represent the documents in vector form. The Jacquard measure should be used if we are considering our documents as just collections of words without any semantic meaning.

Cosine and Euclidean distance are the most widely used approaches, so we will use them in our models below. At the same time, we will use embeddings for the vector representation of text. These are vector representations of text in which words or sentences with similar meanings or contexts have similar representations (vectors).

Base model concepts. As methods for conducting the research, a set of the most popular models was chosen, including TF-IDF, word2vec, doc2vec, and BERT. These are listed in order of increasing complexity of implementation and hypothesis testing. However, each deserves attention, as even the simplest approaches often show good results.

Concept of TF-IDF. One of the most popular and simplest methods in classical machine learning is TF-IDF. This is a combination of term frequency (TF) and inverse document frequency (IDF) for a document. It assigns weights to each word in a document based on its occurrence not only in the document itself but also in the general number of documents with that term in the whole corpus. This reflects its distribution, and thus the importance of determining the text's topic.

$$TFIDF_{t,d} = TF_{t,d} \times IDF_{t,d} = TF_t \times log\frac{N}{df_t} \quad (3)$$

Interpreting this metric can be done as follows: TF-IDF is highest for term t when it appears many times in a small number of documents. TF-IDF will be lower for terms t when they appear in a document fewer times or appear in many documents. TF-IDF will be the lowest when the term is present in all documents.

Concept of Word2vec. Word2vec maps words into a vector space, as the name implies. Word2vec takes a corpus of words as input data and creates vectors of words as output. Word2vec has two main learning algorithms: a continuous bag of words (CBOW) and a continuous skip-gram (Chakraborty & Bhattacharjee, 2020). The difference is shown in the image below, and can generally be defined as predicting a word from its context or predicting the context of a word.

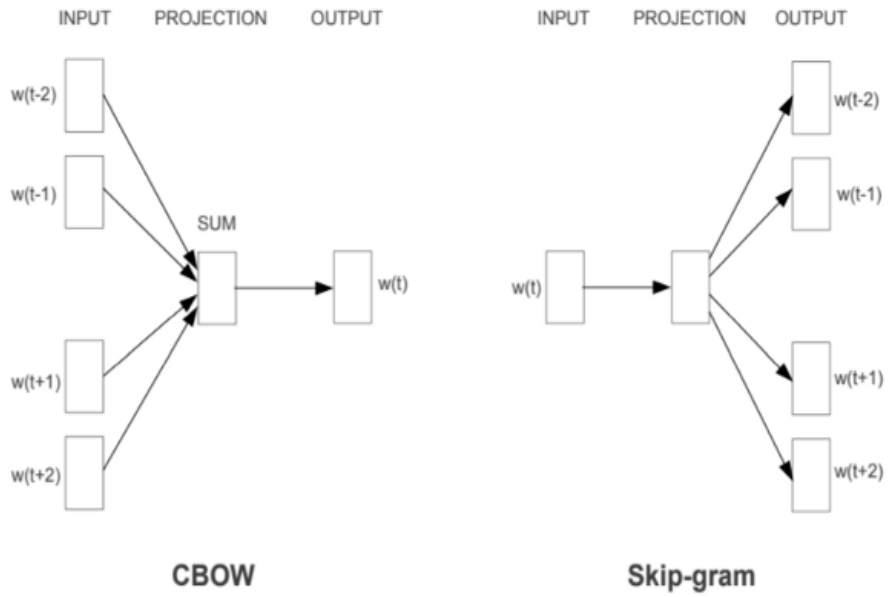

Figure 2: Bag of words and skip-gram models

Using this vector data, we can transform each word of our document corpus into a 100-dimensional vector. We have vectorized all of the text descriptions of the Armed Forces' strategic goals and budget programs. However, the model only returns word vectors. To vectorize documents, we need to represent them as a collection of vectors. In the case of angle approximation, taking the average of all the word vectors in the text is appropriate to get its representation.

However, varying document lengths can also affect the conduct of such operations. Hence, one of the best ways to do this would be to use the weighted average of the word vectors using the TF-IDF method. This to some extent solves the problem of varying text

lengths and also preserves the semantic and contextual meaning of the words. After such transformations, we can use pairwise cosine distances to compute similar documents, as we did in the TF-IDF model.

Concept of Doc2vec. Doc2vec is an unsupervised learning algorithm for creating vector representations of sentences/paragraphs/documents. It is an adaptation of word2vec. Doc2vec can represent an entire document as a vector. This way, we don't need to average the word vectors to create a document vector (Li, 2010).

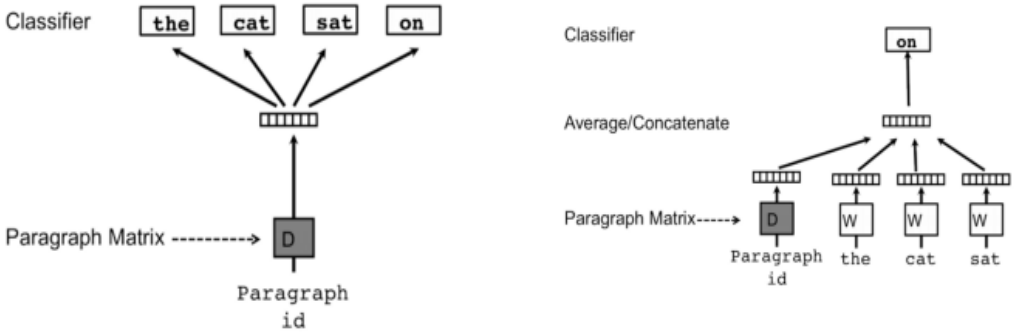

Figure 3: The distributed bag-of-words representation of a paragraph as well as its reversed version for doc2vec.

Concept of BERT. One of the leading methods currently used for modern Natural Language Processing tasks is BERT (Bidirectional Encoder Representations from Transformers). Developed by Google, BERT is a modern pre-training method for Natural Language Processing that is trained on large corpora of unsupervised text data like Wikipedia and Book Corpus. BERT employs the Transformer architecture and an attention mechanism to learn word vectors (Hájek, 2018).

BERT consists of two pre-training stages: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). During BERT training, three different embeddings are utilized: token embeddings, segment embeddings, and position embeddings.

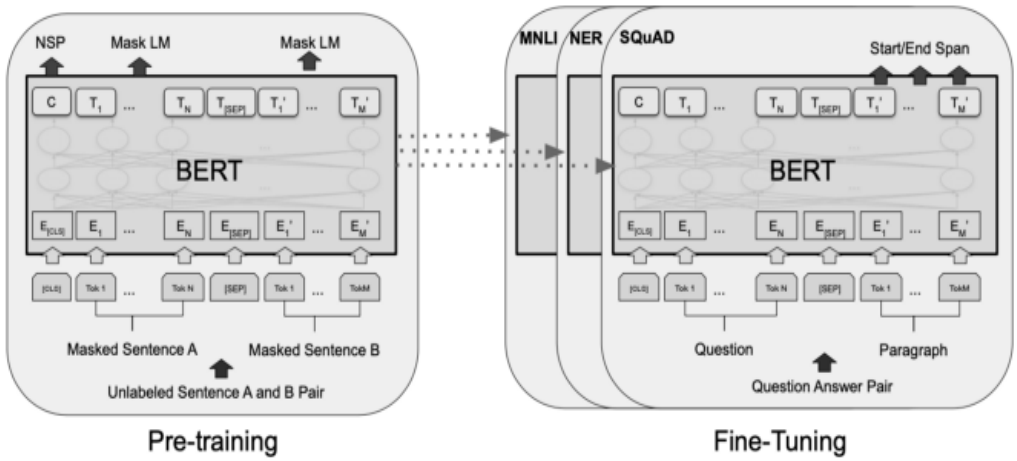

Figure 4: BERT architecture

We will be using the pre-trained BERT model from Huggingface for our corpus processing. We download the base BERT model that consists of 12 layers (Transformer blocks), 12 attention layers, 110 million parameters, and 768 hidden layers. However, we applied the fine-tuned model with the weights “youscan/ukr-roberta-base” on Ukrainian Wikipedia**.**

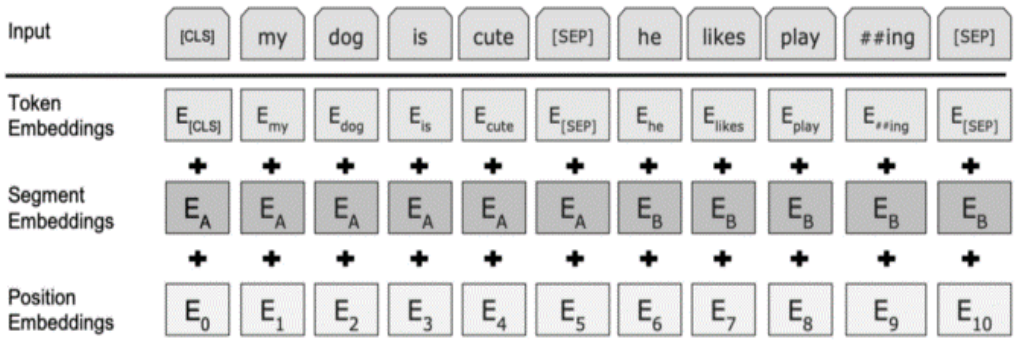

Figure 5: Vector representation of text in BERT

5. Results of research and experiments

To present our research results, we will describe the data used in the experiments, which can be reused to verify the results. We will also discuss the approach to model construction and evaluation of the model results in the form of a confusion matrix.

Data model

As data for the research, the textual part of the description of programs, objectives, and directions of activity was used, which were determined by the Strategic Defense Bulletin of Ukraine. Another significant part was composed of data from the passports of the budget programs of expenses of the Ministry of Defense of Ukraine for different years in the section

of codes of the program classification of expenses. For each code, a list of directions for which expenses are made has also been determined.

Model development

For the development and validation of models, open-source tools were used. Firstly, the software part was implemented in Python, using free libraries: scikit-learn, spaCy, nltk, transformers, etc. Infrastructurally, all computations were performed on a local computer or in a Google Collaboratory cloud environment. However, for reproducing the experiment, we have prepared the corresponding repository on GitHub with a fixed development environment.

Measurement of the results

Thus, information on the achievement of the planned goals, tasks, and results of budgetary programs, as well as goals of state policy for ministries and departments now necessarily include data on the goals of state policy and indicators of their achievement. However, it should be noted that the achievement of goals is expressed in quantitative indicators and does not provide a breakdown and comparison of sources for achieving the above-mentioned strategic goals. For example, the goal of acquiring the capabilities of the Armed Forces of Ukraine for assured deterrence of military aggression, defense of the state, and participation in maintaining peace and security include elements such as:

- Number of units of restoration of armaments and military equipment;
- Number of purchased components for ammunition;
- Number of built and reconstructed buildings;
- Number of cases when housing was obtained for servicemen or cash compensation was paid;
- Number of a brigade and international trainings;
- Number of hours of training for one aircraft crew and number of days of training for one ship at sea;

The comparison matrix below clearly shows the correlation between the improvement of the management system of the Armed Forces of Ukraine and the development of the necessary infrastructure for the organization of effective army management. Also quite well-identified tasks associated with the preparation of the army.

Table 1

Comparison of strategic goals with programs using TF-IDF

Budget program	Strategic goals				
	1	2	3	4	5
2101010	0.242008	0.056054	0.123157	0.074834	0.066047
2101020	0.06078	0.11417	0.059488	0.159239	0.069585
2101150	0.07537	0.12109	0.030639	0.126857	0.042486
2101190	0.065279	0.072228	0.099794	0.162422	0.102912
2101210	0.089781	0.115941	0.083427	0.137616	0.087196
2105010	0	0.055737	0	0.051396	0

The most successful result was achieved precisely for the 2101010 programs, related to the leadership and military management of the Armed Forces of Ukraine. It was

very accurately correlated with the goal of effective defense management and army leadership. An interesting result was also for the most significant program 2101020 - providing the Armed Forces. We managed to correlate it with the goals related to the maintenance of personnel, and also the development of material and technical base and infrastructure. It should be noted that we received zero values precisely for the programs related to the provision of the State Transport Service, which is not reflected in the strategic goals. But the conclusion from our analysis is that this method relies on the frequencies of using key terms, but may miss the semantic sense of goals if they are formulated too generally.

Table 2

Comparison of strategic goals and tasks 2101020 using word2vec

Task	Tasks in program 2101020							
	1	2	3	4	5	6	7	8
1.1.	0.47705	-0.01697	0.43750	0.25428	0.40695	0.16309	0.14487	0.17340
1.2.	0.20069	0.0371	0.20804	0.26194	0.22359	0.01470	0.24857	0.35374
1.3.	0.53632	-0.10965	0.53726	0.41430	0.56833	0.32287	0.24076	0.42315
1.4.	0.38564	0.06017	0.37229	0.35051	0.44340	0.11438	0.28490	0.44025
1.5.	0.52233	-0.13649	0.51066	0.46888	0.40523	0.19673	0.31650	0.38400
1.6.	0.53989	-0.00085	0.66263	0.50821	0.67536	0.37539	0.06104	0.47046
2.1.	0.65395	-0.0934	0.59164	0.51852	0.62837	0.37488	0.30827	0.43864
2.2.	0.04032	-0.04164	0.25885	-0.04607	0.15882	-0.05992	-0.15871	0.18893
2.3.	0.23805	0.18687	0.2663	0.19052	0.25265	0.37282	0.09439	0.35351
2.4.	0.56521	0.06559	0.48563	0.43944	0.50559	0.39591	0.26960	0.43357
2.5.	0.26727	-0.15417	0.44765	0.28177	0.39953	0.21405	0.03939	0.35993
2.6.	0.27551	0.02231	0.42327	0.23239	0.23767	0.13378	-0.07082	0.22818
3.1.	0.38474	-0.04965	0.38572	0.27809	0.21000	0.10106	0.17688	0.20286
3.2.	0.31977	-0.13309	0.43832	0.36574	0.40853	0.30311	0.12705	0.21825
3.3.	0.52050	-0.13457	0.52833	0.53385	0.52685	0.32006	0.3087	0.51098
4.1.	0.36428	-0.27166	0.56158	0.23082	0.40504	0.03308	-0.00212	0.41066
4.2.	0.53773	-0.09794	0.52111	0.18747	0.48641	0.2062	0.25527	0.33461
4.3.	0.47859	0.12076	0.42574	0.21796	0.44983	0.31343	0.02363	0.47773
4.4.	0.43830	0.09991	0.45803	0.41748	0.38699	0.32468	0.39904	0.29314
4.5.	0.57116	-0.07585	0.46798	0.35238	0.47929	0.21505	0.24377	0.47404
4.6.	0.59816	-0.12025	0.66878	0.40525	0.50595	0.38107	0.14572	0.29419
5.1.	0.41355	-0.09024	0.30641	0.35769	0.27658	0.25561	0.28117	0.22640
5.2.	0.51739	-0.07918	0.47012	0.60526	0.49679	0.43958	0.33584	0.47600
5.3.	0.16916	-0.22702	0.16751	0.20490	0.20519	0.29637	0.40582	0.14584
5.4.	0.58528	-0.17945	0.53376	0.53334	0.63023	0.35809	0.25245	0.45090
5.5.	0.40857	-0.11308	0.43892	0.46149	0.40168	0.26268	0.15380	0.51693
5.6.	0.21652	-0.01548	0.38202	0.32292	0.45508	0.27594	0.19915	0.46573
5.7.	0.40797	0.15798	0.53031	0.52966	0.37882	0.33918	0.10780	0.35536

Taking into account the results presented in Table 2, it is worth noting the intersection of strategic task 5.3, related to achieving the ability of military units and subunits to rapidly deploy in threatening directions, implement tasks independently of the main forces, and task 7, which entails updating and modernization of mobile command points, life support systems of stationary command points, repair of premises and equipment of working places of the operational stock. Goal number two does not have common global goals, since they are related to payments of one-time monetary assistance in the event of death and disability.

However, for many other goals, a significant specification is required, since the terms “ensuring”, and “personnel” are encountered too often, and some goals are similar to each other.

The results of this modeling came out quite ambiguous since it is obvious that some goals are very close for most programs, but this is explained by a similar general context. Although it turned out quite accurately identifies the goals of housing construction, motivation of the personnel, the introduction of a professional contract army, development of weapons and military equipment, and also leadership of the Armed Forces.

Table 3

Comparison of strategic goals and expenditures of using doc2vec

Task	Expenditures in program 2101020					
	7	8	9	10	11	12
1.1.	0.499582	0.561908	0.550095	0.569777	0.611908	0.466576
1.2.	0.588981	0.618015	0.642168	0.655437	0.652241	0.485467
1.3.	0.510976	0.539857	0.559634	0.571945	0.607111	0.467814
1.4.	0.531586	0.595911	0.621792	0.627173	0.588228	0.489907
1.5.	0.567988	0.629112	0.646249	0.625536	0.754146	0.338862
1.6.	0.541215	0.606805	0.562156	0.578074	0.559763	0.896888
2.1.	0.633827	0.666766	0.638383	0.650102	0.607704	0.647732
2.2.	0.912591	0.928676	0.930481	0.949818	0.886077	0.873553
2.3.	0.671123	0.629871	0.684401	0.687007	0.546914	0.544061
2.4.	0.609456	0.585527	0.652405	0.63627	0.484719	0.512284
2.5.	0.65825	0.711314	0.684022	0.718494	0.684411	0.689997
2.6.	0.582291	0.629573	0.640124	0.619932	0.532046	0.523344
3.1.	0.783368	0.631545	0.667342	0.67135	0.664037	0.501536
3.2.	0.801945	0.613885	0.636391	0.639732	0.607939	0.59735
3.3.	0.60701	0.562256	0.568473	0.592695	0.595261	0.636408
4.1.	0.900857	0.879537	0.898161	0.896762	0.795849	0.840514
4.2.	0.711481	0.792752	0.78627	0.784656	0.814861	0.555519
4.3.	0.605605	0.596577	0.644492	0.650863	0.56131	0.53775
4.4.	0.933084	0.930592	0.9244	0.937637	0.890443	0.891942
4.5.	0.68407	0.736484	0.723175	0.737439	0.699477	0.552989
4.6.	0.405965	0.65218	0.549739	0.540869	0.631013	0.54467
5.1.	0.767111	0.81502	0.829927	0.840854	0.84015	0.651231
5.2.	0.682064	0.77779	0.798249	0.815563	0.775561	0.758127
5.3.	0.635231	0.632956	0.629205	0.62644	0.538229	0.662575
5.4.	0.576035	0.641883	0.590183	0.610267	0.603012	0.88292
5.5.	0.572286	0.691894	0.685003	0.702265	0.727041	0.5777
5.6.	0.590051	0.635674	0.66342	0.677971	0.612442	0.579411
5.7.	0.876314	0.885739	0.856438	0.852058	0.832015	0.867883

The matrix given above testifies that some goals are very non-specific and, consequently, it is very difficult to accurately classify and attribute them to a particular task. However, the digitalization of activities and the introduction of information technologies have been well-identified. Tasks 3.1 and 3.2 should also be noted, which are well correlated with expenditure type 7, which is responsible for armaments or military equipment. No less important, but worth attention is the identification of expenses related to increasing compatibility with NATO forces.

Table 4

Comparison of strategic goals and program expenditures using BERT

Task	Tasks of program 2101020						
	1	2	3	4	5	6	7
1.1.	0.650841	0.809275	0.775602	0.690156	0.775734	0.709425	0.681196
1.2.	0.714152	0.746669	0.670013	0.639065	0.711457	0.643466	0.729826
1.3.	0.622907	0.673193	0.680382	0.668679	0.712899	0.780568	0.632537
1.4.	0.696227	0.741491	0.814631	0.761795	0.7668	0.738501	0.654366
1.5.	0.700455	0.762433	0.738132	0.68155	0.741317	0.68315	0.735019
1.6.	0.716989	0.771381	0.806691	0.704533	0.79421	0.686321	0.692223
2.1.	0.621287	0.769362	0.754073	0.575456	0.738937	0.614945	0.619564
2.2.	0.659433	0.735129	0.812206	0.662479	0.742009	0.78904	0.65008
2.3.	0.568149	0.757282	0.836259	0.717073	0.76238	0.646341	0.646641
2.4.	0.679855	0.739786	0.714818	0.661308	0.697978	0.683997	0.699712
2.5.	0.689365	0.687489	0.73541	0.701417	0.729842	0.744897	0.619225
2.6.	0.638779	0.769699	0.819981	0.686148	0.792692	0.641821	0.630986
3.1.	0.650841	0.809275	0.775602	0.690156	0.775734	0.709425	0.681196
3.2.	0.633182	0.778294	0.768059	0.700972	0.762283	0.722051	0.706131
3.3.	0.655214	0.751822	0.838065	0.720264	0.771419	0.697393	0.641031
4.1.	0.697231	0.801641	0.798247	0.781194	1	0.677852	0.739885
4.2.	0.701208	0.669328	0.732812	0.789508	0.743674	0.754952	0.624369
4.3.	0.636802	0.754787	0.823362	0.710258	0.762162	0.617667	0.660042
4.4.	0.658178	0.788476	0.821771	0.730868	0.762916	0.654685	0.758833
4.5.	0.722132	0.778592	0.742612	0.676078	0.763263	0.650652	0.687019
4.6.	0.597531	0.760153	0.833015	0.713335	0.790015	0.615363	0.627138
5.1.	0.660549	0.775333	0.827598	0.829643	0.831727	0.635528	0.782949
5.2.	0.549837	0.6862	0.771135	0.655958	0.713802	0.550901	0.593941
5.3.	0.661917	0.744142	0.784723	0.641079	0.750657	0.753363	0.649679
5.4.	0.682564	0.729755	0.74567	0.644616	0.741368	0.609572	0.647195
5.5.	0.597531	0.760153	0.833015	0.713335	0.790015	0.615363	0.627138
5.6.	0.569216	0.702809	0.765816	0.663734	0.703314	0.540686	0.617163
5.7.	0.66768	0.783619	0.769247	0.664578	0.736599	0.796131	0.706561

6. Discussion

The four case studies offer an interesting collection of results for different approaches: TFIDF, word2vec, doc2vec and transformers (BERT). We constructed a clear workflow and architechture for solving the problem of defense budget expenditures and strategic goals comparison in case where there is no direct relations. Results of the research approved the that there is a sematic relation among strategic goals or tasks and defense budget programs. The cosine distance for text embeddings indicates different relations for different parts of strategic documents.

In recent stydies Georgescu (2020) and Bruno (2019) both demonstrate the potential of natural language processing (NLP) in defense-related applications suche as automatic analysis of cybersecurity-related documents or optimization of US weapon systems. Howere, the area of defense budfet analysis was undeveloped. Recently Damaševičius (2023) provides a comprehensive overview of AI's impact on various fields, including economics, finance, and innovation. But it is necessary to emphasize that almost all recent researched describes application of large-langiage models (LLMs), while in our

work we provide usefull examples of basic and ligt-weight models. They could be used completely offline and do not require significant computational resiurces.

The results should be interpreted with the limitations of the method and particular document samples in mind. It could be necessary to use larger military documents database in order to create bigger and richer corpus of embeddings. That was one of the reason for some similarities in dataset.

Other possible influences, only regional document database, for example, could be considered. In addition, there exist some similarity in goals description in ukrainian language. Unfortunately, there is a small number of available trasformer models for ukrainian language and we could achive better results especially in English. Further research may be able to accumulate enough document database and apply modern large language models.

7. Conclusion

We have conducted 72 experiments for various combinations of data, distance metrics, and text vectorization methods. In general, we can consider the experiments successful, and also formulate certain conclusions:

- Many strategic goals, objectives, cost types, etc. described using similar terms and definitions. In most cases, the difference between targets is rather blurry, as they can only be distinguished by single individual terms in the description;
- The subject of military texts is not very easy to understand algorithms. Therefore, to properly vectorize texts and improve accuracy, it is necessary to increase the educational corpus of texts to increase the number of words, diversify their context;
- We provided for the measurement of distances between vectors using Euclidean distances, and cosine distances, as well as using various methods for evaluating words and phrases, but cross-validation for such combinations is possible only in cases of expert evaluation of the results;
- A potentially more efficient way to use this approach may be to build a multiclass classifier, but such a model requires a significant corpus of texts and words. You also need to label the training data;

The measures we have identified and the models we have considered, if implemented effectively, will certainly have a positive effect and will help improve the quality of the processes for analyzing costs and operations, comparing them with the goals of budget programs. At the same time, subject to the introduction of such a system, the state will be able to conduct such an analysis not only within the Ministry of Defense but also for other departments.

Thus, the analysis of the possibilities of using modern methods of natural language processing and their application in the public sector made it possible:

- ✓ To formulate an innovative approach using modern information technologies for a comparative analysis of the proximity of the elements of strategic goals and the costs of budget programs;
- ✓ To formulate proposals for the development of a model of a universal classifier of operations according to their textual description;
- ✓ An algorithm for training and retraining the obtained models is proposed.

In sum, our models and empirical results complement and expand the existing literature on the strategic analysis of defense budgets. We believe that future research needs to pay closer attention to the strategic goals and tasks in defense budgets in order to achive the necessary capabilities level for countries Armed Forces.

Acknowledgement: This work was supported by the Hasso Plattner Foundation through the grant LBUS-UA- RO-2023, financed by the Knowledge Transfer Center of the Lucian Blaga University of Sibiu.

The article results is to be implemented in the national Ukrianian project "Openness and transparency of budgeting in the defence and security sector of Ukraine" (Ministry of Education and Science of Ukraine, ID: 186964, 04.11.2021, 00002-1).

8. References

Bochkay, K. (2014) Enhancing empirical accounting models with textual information (Doctoral dissertation). Rutgers University-Graduate School-Newark.

Bruno, N., Jun, T., Tessier, H. (2019) Natural language processing and classification methods for the maintenance and optimization of US weapon systems. In systems and information engineering design symposium, SIEDS, https://doi.org/10.1109/sieds.2019.8735587

Chakraborty, B. and Bhattacharjee, T. (2020) A review on textual analysis of corporate disclosure according to the evolution of different automated methods.Journal of Financial Reporting and Accounting ,18(4):757-777.

Damasevicius, R. (2023) Artificial intelligence techniques in economic analysis, Economic Analysis Letters, 2(2): 52–59. https://doi.org/10.58567/eal02020007

Davies, J., Arana-Catania, M., Procter, R., van Lier, F.A. and He, Y. (2021) Evaluating the application of NLP tools in mainstream participatory budgeting processes in Scotland, In Proceedings of the 14th International Conference on Theory and Practice of Electronic Governance, pp. 362-366.

Devlin, J., Chang, M. W., Lee, K. and Toutanova, K. (2018) Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805.

El-Haj, M. et al. (2019) In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse.Journal of Business Finance & Accounting, 46(3-4): 265-306.

Fernandez-Cortez, V. et al. (2020) Can artificial intelligence help optimize the public budgeting process? Lessons about smartness and public value from the Mexican federal government. In 2020 Seventh International Conference on eDemocracy & eGovernment, ICEDEG, pp.312-315.

Fisher, I. E., Garnsey, M. R. and M. E. Hughes, M. E. (2016) Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research, Intelligent Systems in Accounting, Finance and Management, 23(3): 157-214.

Hájek, P. (2018) Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns.Neural Computing and Applications, 29: 343-358.

Henry, E. and Leone, A.J.. (2016) Measuring qualitative information in capital markets research: Comparison of alternative methodologies to measure disclosure tone. The Accounting Review, 91(1): 153-178.

Georgescu, T.-M. (2020) Natural language processing model for automatic analysis of cybersecurity-related documents. Symmetry, 12(3): 354. https://doi.org/10.3390/sym12030354

Le, Q., Mikolov, T. (2014) Distributed representations of sentences and documents. In International conference on machine learning, PMLR, pp. 1188-1196.

Li, F. (2010) Textual analysis of corporate disclosures: A survey of the literature. Journal of accounting literature, 29(1): 143-165.

Loughran, T. and McDonald, B. (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of finance, 66(1): 35-65.

Loughran, T. and McDonald, B. (2016) Textual analysis in accounting and finance: A survey. Journal of Accounting Research, 54(4):1187-1230.

Luo, Y.and Zhou, L. (2020)Textual tone in corporate financial disclosures: a survey of the literature.International Journal of Disclosure and Governance, 17(2-3): 101-110.

Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013) Efficient estimation of word representations in vector space, arXiv preprint arXiv:1301.3781.

Siano, F. and Wysocki, P. (2021) Transfer learning and textual analysis of accounting disclosures: applying big data methods to small (ER) datasets. Accounting Horizons, 35(3): 217-244.

Valle-Cruz, D., Fernandez-Cortez, V. and Gil-Garcia, J. R. (2022) From E-budgeting to smart budgeting: Exploring the potential of artificial intelligence in government decision-making for resource allocation, Government Information Quarterly, 39(2): 101-144.

Wujec, M. (2021) Analysis of the Financial Information Contained in the Texts of Current Reports: A Deep Learning Approach.Journal of Risk and Financial Management, 14(12): 582.

Zhang, M.C. et al. (2019) Text data sources in archival accounting research: Insights and strategies for accounting systems' scholars.Journal of Information Systems, 33(1):145-180.

Zuiderwijk, A. et al. (2021) Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda.Government Information Quarterly, 38(3):101577.