Implementation of Data Science Methods in Armed Forces Budgeting: Challenges and Opportunities

Tetiana Zatonatska, Vadym Pakholchuk, Alim Syzov and Daryna Vorontsova

Taras Shevchenko National University of Kyiv, Volodymyrska Street 60, Kyiv, 01033, Ukraine

Abstract

This paper explores the challenges and opportunities of incorporating data science methods into military budgeting, with a focus on practical advice in the area of forecasting defense budgets and expenditures. The study uses aggregated monthly expenditure data from military bases and applies various forecasting models, including smoothing, linear regression, prophet and XGBoost. The three best performing models are selected to build a meta-model that improves the accuracy of the forecasts. The results suggest that incorporating data science methods into military budgeting can lead to better decision making and resource optimisation.

Keywords [1]

Data Science, armed forces budgeting, forecasting, defense budget, defense expenditures

1. Introduction

Budgeting in the armed forces is a complex process that requires careful consideration of various factors such as operational requirements, personnel needs and procurement costs. In recent years, there has been an increasing focus on incorporating data science methods into the budgeting process, as these methods can help to improve decision making and optimize resource allocation. This thesis explores the challenges and opportunities associated with incorporating data science methods into armed forces budgeting.

Budgeting in the armed forces is a complex and critical process that requires careful consideration of a range of factors, including operational requirements, personnel needs and procurement costs. The importance of accurate budgeting in the armed forces cannot be overstated, as it can have a direct impact on operational readiness, strategic planning and, ultimately, the success of military operations. However, the Armed Forces budgeting process is not without its challenges, such as the need to allocate resources efficiently, effectively and on time. In recent years, there has been a growing interest in incorporating data science methods into the Armed Forces budgeting process. Data science methods have the potential to improve decision-making and resource allocation by providing insights into complex data sets and identifying patterns and trends that may not be immediately apparent. In this chapter, we will explore the challenges and opportunities associated with incorporating data science methods into Armed Forces budgeting. We will provide an overview of data science methods that can be used in the budgeting process, discuss the importance of data collection and analysis, explore modeling and forecasting techniques, and examine the potential benefits and challenges of incorporating data science methods into the Armed Forces budgeting process.

This thesis aims to build a system of models for use in the armed forces budgeting process and to develop recommendations for the use of data science tools in armed forces budgeting. The scientific hypothesis can be attributed to the improvement of forecasting accuracy in the budgeting process when implementing data science technologies.

In summary, the incorporation of data science methods into military budgeting has the potential to transform the way budgets are developed and allocated, ultimately improving mission readiness and

Information Technology and Implementation (IT&I-2023), November 20-21, 2023, Kyiv, Ukraine
EMAIL: tzatonatska@gmail.com (T. Zatonatska); vadym.pakholchuk@univ.kiev.ua (V. Pakholchuk); alik_sizov@ukr.net (A. Syzov); v.darisha@ukr.net (D. Vorontsova)
ORCID: 0000-0001-9197-0560 (T. Zatonatska); 0000-0002-9657-6148 (V. Pakholchuk); 0000-0002-6400-3951 (A. Syzov); 0000-0003-0227-6935 (D. Vorontsova);

success. However, this transformation requires careful consideration of the challenges and opportunities associated with these methods. The following sections of this chapter explore these challenges and opportunities in more detail, and provide guidance and recommendations for the successful integration of data science methods into Armed Forces budgeting.

2. Literature review

In today's world of limited resources, how to properly manage and control them is a serious issue. For a long time, economists, politicians and citizens themselves, as their literacy grew, had concerns about economic growth and the effectiveness of government. Thus, different methods of measuring these parameters, which are indicators of government productivity measurement, were born, and the most advanced of them use data science tools for optimisation.

One study concluded that current outdated budgeting, forecasting and planning methods require revision [1]. These methods are time-consuming and demanding on staff, and are also flawed as they frequently contain incomplete and inaccurate data. Changes in business strategies or exogenous factors can render these budgeting plans ineffective. The authors advocate employing driver-based budgeting, which connects a business's performance with its activity. It can enhance complete openness by assessing various business drivers that may affect its performance. By utilizing this system and integrating predictive modeling using machine learning, it can considerably facilitate the development of budgets and plans for the company's future. Furthermore, such models tend to be both more precise and impartial than human counterparts, as well as being more cost-effective to utilize.

The incorporation of machine learning in such systems has resulted in an elevated risk of data breaches in recent times. This development is crucial as the sensitive data being used in these algorithms can now be stolen, which has significant implications for private businesses and governments. According to one study, it is imperative that a specified machine learning model be employed to safeguard this type of data [2]. The authors have developed a "differentially-private filtering scheme" to safeguard confidential information, such as budgeting, in machine learning algorithms while querying. One of the most crucial aspects of government budget allocation is funding Research and Development. In the contemporary world, where knowledge is pivotal to a nation's growth and well-being, there is a growing consensus on the necessity of increased R&D investment. However, there remains a level of uncertainty when assessing the returns on R&D expenditure. Jang H. has developed a machine learning model that applies a robust optimization technique to assess the risks associated with predicted R&D output values [3]. The model's implementation resulted in a 13.6% improvement in budget allocation towards R&D, making it a valuable tool for achieving systematic and efficient budget allocation in data science. Arguably the most important part of any business plan today is budgeting, as it directly explains what possibilities there are in terms of financial capacities. Faccia A. suggests the use of big data and data science tools can make these processes more efficient in terms of time, finances, and productivity [4]. The author came up with a statistical model that would draw from the pool of data of the business and create a balanced budget based on indicators and assumptions provided in the model. This creates a practical standardization approach to budgeting processes, which is important across all sectors of a business big or small. The author concludes that while the use of big data and data science can make life easier for the business in terms of budget, the output is only going to be as good as the data put in.

In a separate investigation , analysts evaluated the government's use of these measures to enhance planning, budgeting, and management processes in Brazil [5]. The researchers contend that the volume of data being processed and stored is colossal and requires a significant amount of computational power. Even after all this, it remains a challenging task to utilize it for machine learning purposes. The authors have concluded that implementation of an optimized model for data extraction, transformation, and loading can expedite the analysis of heterogeneous data and enhance the functioning of Brazil's Ministry of Planning, Budgeting, and Management's SIAPE system. The SIAPE system controls payroll information for all federal public-sector employees. In the future, authors aim to enhance the process by reducing its demands on the electrical systems. This will decrease costs for the government.

Municipal budgeting is an intricate process. Currently, several budgeting systems incorporate Key Performance Indicators of a city to establish metrics that are fundamental for the city's growth. The study's authors developed an autoregressive linear model that uses public KPIs, including crime rates and healthcare KPIs, to optimize municipal budgets for public safety and healthcare [6]. According to the authors, this model enables the analysis of the current situation in these sectors, creates forecasts, and identifies the most significant performance indicators. For future research, the employment of nonlinear models is recommended alongside the extension of the geographical scope of the analysis.

Fisher, Garnsey, and Hughes' article offers a thorough assessment of natural language processing (NLP) in accounting, auditing, and finance. They introduce the concept of NLP and its potential applications within these fields before reviewing relevant literature. The article delves into topics including sentiment analysis, fraud detection, and financial statement analysis. The aforementioned article outlines the challenges and limitations of implementing NLP in these fields, while also offering a roadmap for future research [7]. In their own article , Valle-Cruz, Fernandez-Cortez and Gil-Garcia investigate the potential of artificial intelligence (AI) in government decision-making for the allocation of resources [8]. Their work includes a comprehensive review of e-budgeting and smart budgeting, followed by an examination of the current status of AI in government budgets. The article explores the challenges and opportunities related to the implementation of AI in this context. It concludes with a discussion of future research directions.

Article evaluates the use of NLP tools in mainstream participatory budgeting processes in Scotland by Davies, Arana-Catania, Procter, van Lier, and He [9]. The article begins with a comprehensive literature review of natural language processing (NLP) and participatory budgeting. The authors then present their study based on the analysis of online comments from a participatory budgeting platform. In conclusion, the article discusses the prospective role of NLP in supporting participatory budgeting and the necessity for further research in this area.

The article authored by Tiron-Tudor, Donțu, and Bresfelean delves into the ways that emerging technologies can contribute to the digital transformation of accountancy firms. In their review of literature pertaining to technologies like cloud computing, blockchain, and artificial intelligence, the authors consider the potential impacts on the accounting profession. The article presents case studies demonstrating the implementation of these technologies by accountancy firms and concludes with an exploration of the challenges and opportunities associated with digital transformation [10].

Eltweri, Faccia, and Khassawneh discuss the applications of big data in finance, specifically in relation to fraud detection and risk management within the real estate industry. The authors critically evaluate the pertinent literature on big data analytics, machine learning and artificial intelligence, providing illustrative instances of how these technologies can be employed to uncover deceit and monitor perils in real estate transactions. Furthermore, the article deliberates the obstacles and constraints of using big data within this framework, culminating with a discourse on future research pursuits [11]. Aboagye-Otchere, Agyenim-Boateng, Enusah and Aryee, conduct an appraisal of big data exploration in accounting. The authors examine the body of literature on big data analytics in finance reporting, auditing and fraud detection, and consider both the benefits and the challenges associated with these applications. Further, the article presents a thorough critique of past research and outlines an agenda for future investigation in this field [12].

The literature review indicates that data science methods, including time-series analysis, machine learning, natural language processing, and statistical models, can be effectively employed for defense budgeting. The studies demonstrated the application of these techniques in various countries and regions. The authors emphasized data science's significance in supplying better decision-making support and precision in forecasting and decision-making. The study identifies the challenges and prospects of incorporating data science in defense budgeting, encompassing the implementation of explainable AI and privacy-preserving machine learning methods. In conclusion, data science has a crucial role in budgeting, as evidenced by its implementation in both business and government sectors. It helps to generate functional standardized data and produce impartial results through advanced machine learning models. However, more technological advances in budgeting data science tools are still necessary.

3. Overview of Data Science Methods

In this section, an overview of data science methods that may prove useful in the budgeting process is presented. Topics to be covered include machine learning, data mining, and predictive analytics, with examples provided of how these methods have been employed in other industries.

Data science methods allow the Armed Forces to utilize data gathering, analysis, modeling, and machine learning to improve budgeting processes. This is achieved through processing vast datasets, obtaining insights, and developing accurate models to facilitate informed decision-making.

The budgeting process can benefit significantly from collating and analyzing data. This enables budget analysts to acquire a thorough understanding of the resources necessary to support mission readiness, identify less-obvious patterns and trends, and gain further insights into the budgeting process. Moreover, by avoiding subjective evaluations and utilizing objective language, the resulting reports are comprehensible, concise, and logically structured. Incorporating historical data to model and forecast can aid budget analysts in anticipating future needs, including personnel and procurement costs. This can enhance decision-making and resource allocation in the budgeting procedure.

Machine learning improves decision-making by analyzing complex datasets and developing accurate predictive models. This transformation of Armed Forces budgeting leads to better resource allocation, requirement planning and efficiency. However, meeting the challenges of data quality and resource requirements requires systematic processes and role adjustments. The implementation of data science techniques in Armed Forces budgeting may necessitate the creation of novel tools and technologies to facilitate their use, including dashboards and data visualization tools.

The implementation of data science techniques in Armed Forces budgeting may necessitate the creation of novel tools and technologies to facilitate their use, including dashboards and data visualization tools. Adopting such techniques has the capacity to revolutionize budget allocation and development, thereby enhancing mission readiness and success. Nevertheless, the effective fusion of these techniques demands scrupulous examination of the associated challenges and prospects. The ensuing parts of this chapter will delve into these difficulties and opportunities in more detail, proffering direction and suggestions for the triumphant amalgamation of data science methods into budgeting for the Armed Forces.

4. Data Collection and Analysis

One of the primary obstacles to incorporating data science techniques in the budgeting of armed forces is the scarcity of data. This segment will address the significance of data accumulation and evaluation, as well as techniques for gathering and scrutinizing data associated with armed forces budgeting. We will also examine the importance of data quality and governance while presenting examples of optimal data collection and analysis strategies.

Collecting and analyzing data is imperative in Armed Forces budgeting to obtain and scrutinize important information, which refines budget accuracy and thoroughness.

Various data sources, such as historical data, surveys, and operational reports, can be used in the budgetary process. Historical data can identify trends and patterns in prior expenditure to develop basic estimates for future disbursements. Surveys may be utilized to obtain information concerning personnel requirements, training demands, and other factors that affect the budget. Operational reports can furnish information on mission readiness and operational requirements, which are indispensable inputs to the budgeting process.

After collecting data, it must be analyzed to inform decision-making and identify insights. Various data analysis techniques, including descriptive analytics, exploratory analytics and predictive analytics can be employed during the budgeting process. The summarizing and description of collected data, providing a high-level outlook while detecting any anomalies or outliers are included in descriptive analytics. Exploratory analytics includes the visualization of data and the exploration of relationships between different variables to gain a deeper understanding of the underlying trends and patterns present in the data. Predictive analytics involves the implementation of statistical models and machine learning algorithms to forecast future expenditure and resource requirements based on historical data.

In addition, data analysis can contribute to the identification of cost savings and operational efficiencies within the budgeting process. For instance, through the analysis of historical data, analysts might be capable of locating areas of immense spending or inefficiencies in the procurement process. By remedying these issues, the Armed Forces could decrease overall spending and dispense resources more proficiently.

To examine our hypothesis of enhancing departmental expense forecasts, we acquired and processed primary data from payment orders of one department for the almost six-year time frame spanning 01.04.2016 to 01.01.2022. A monthly granularity level was deemed appropriate, as it aligns with the planning horizon and economic content encompassed in the allocation plans of subordinate departments. However, it is important to note that the distribution of expenses in a calendar month has a specific structure and clear seasonality. This is determined by the governing documents, which take into account the payment of monetary allowances and wages, expenses that make up a significant proportion of the total financial resources, and are paid on certain dates. We have also observed a complex seasonality throughout the year, as evidenced by the provided graph.

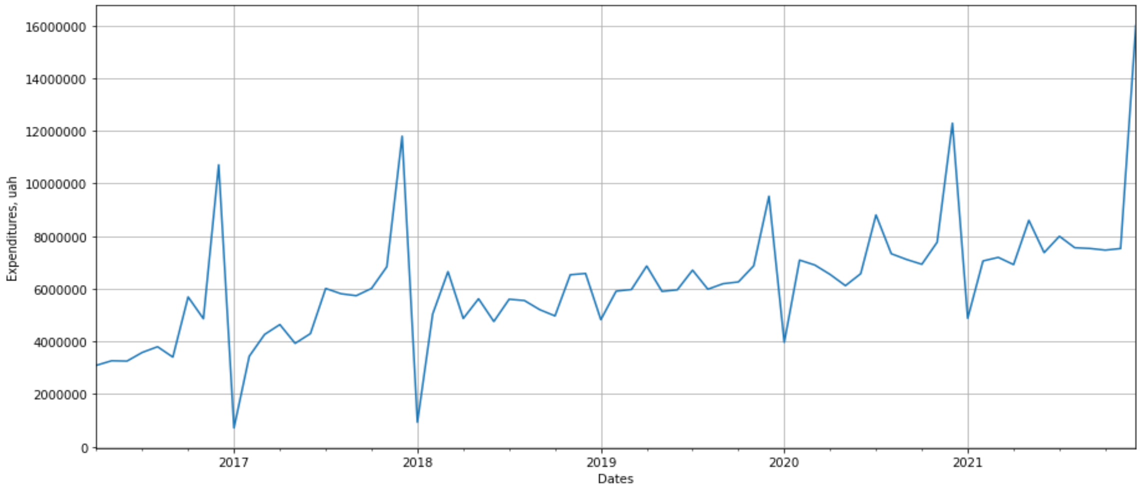

Figure 1: Monthly military base expenditures for 2016 -2021 years

Effective budgeting in the Armed Forces requires comprehensive data collection and analysis. Budget analysts can produce more precise and thorough budgets, allocate resources more efficiently, and identify opportunities for cost savings and streamlining through the examination of pertinent information.

5. Modeling and Forecasting results

After collecting and analyzing data, models and forecasts can be formulated. This section will examine the utilization of modeling and forecasting in budget allocation for armed forces. Furthermore, illustrations of their applications in other industries will be provided. Varied modeling techniques will be explored, comprising regression analysis and neural networks. We will examine the significance of scenario analysis in forecasting and illustrate its applications in the context of budgeting for the Armed Forces. Data science methods for modeling and forecasting constitute essential elements of the budgeting process. The objective of modeling is to formulate mathematical or statistical models that can represent the associations between diverse variables in the budgeting process. These models can predict future expenditures and resource requirements by analyzing historical data and identify potential cost savings and efficiency areas.

The budgeting process encompasses various modeling techniques, such as linear regression, time-series analysis, and machine learning algorithms. Linear regression models can identify the connections between various factors, such as mission requirements and staff needs, and make predictions based on those connections. Time-series analysis examines trends and patterns in historical data, providing insight for future spending forecasts based on those trends. Advanced machine learning techniques, including decision trees and neural networks, enable the creation of sophisticated models capable of identifying nonlinear associations between variables and generating highly precise forecasts.

One classic mathematical method that belongs to a group of linear smoothing models is the moving average model. Although this model is mainly unsuitable for long-term predictions, it is adept at smoothing trends and deviations.

$$\widehat{y_t} = \frac{1}{k}\sum_{n=0}^{k-1} \quad y_{t-n} \tag{1}$$

The model smoothing results are presented below. Notably, the confidence intervals visibly increase during characteristic periods, especially towards the end of each year as depicted in the graph. Such periods are structurally dissimilar in the series; hence, the substantial variance increase is explicable.

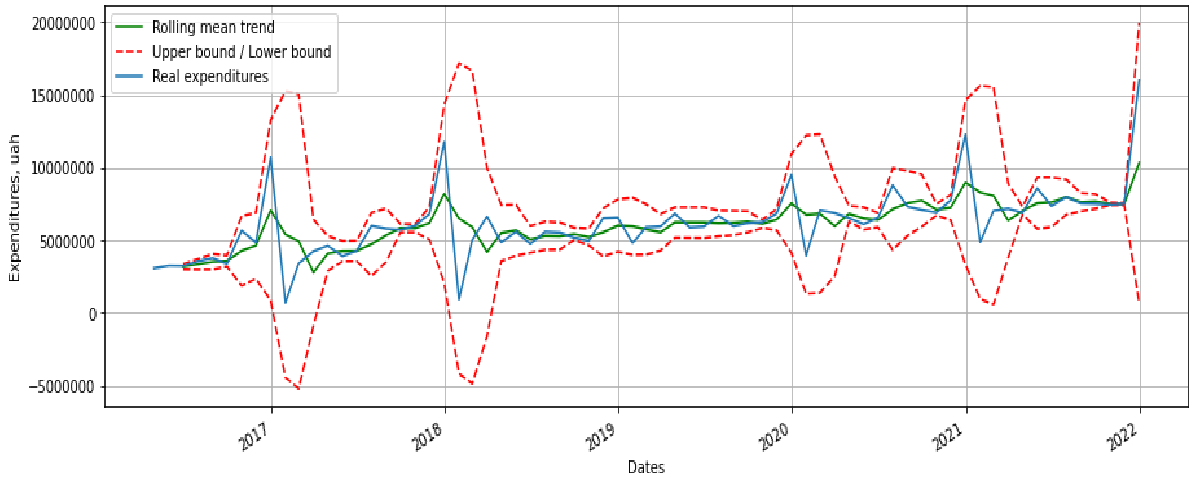

Figure 2: Rolling mean forecasting with window size 3 months

In practice, it is challenging to make a long-term forecast as, for one step forecasting, the preceding value needs to be an actual and known quantity. Frequently, enhancing this model involves applying certain weights that reflect the significance of the most recent values in predicting the next one. Furthermore, usually, more importance is given to the values that are closest to the forecast.

A more advanced approach to the previous solution is the exponential smoothing model, which weighs the most recent n values and accounts for the exponential reduction of the model coefficients.

$$\widehat{y_t} = \alpha \cdot y_t + (1-\alpha) \cdot \widehat{y_{t-1}} \tag{2}$$

The presented model value is a result of the weighted average of both the current true and previous model values. The smoothing factor, α, determines the rate at which we will discard the last available true observation. A smaller α corresponds to a greater influence of previous model values, leading to a more smoothed-out series. Exponentiality is concealed within the recursive function. Each iteration involves multiplying the previous model value by (1-α), which also contains (1-α), and so on until the very start. The figure below illustrates several versions of exponential smoothing that utilize different α parameters. The presented model value is a result of the weighted average of both the current true and previous model values. The smoothing factor, α, determines the rate at which we will discard the last available true observation. A smaller α corresponds to a greater influence of previous model values, leading to a more smoothed-out series. Exponentiality is concealed within the recursive function. Each iteration involves multiplying the previous model value by (1-α), which also contains (1-α), and so on until the very start. The figure below illustrates several versions of exponential smoothing that utilize different α parameters.

$$l_x = \alpha y_x + (1-\alpha)(l_{x-1} + b_{x-1}) \tag{3}$$

$$b_x = \beta(l_x - l_{x-1}) + (1-\beta)b_{x-1} \tag{4}$$

$$\widehat{y_{x+1}} = l_x + b_x \tag{5}$$

The presented model value is a result of the weighted average of both the current true and previous model values. The smoothing factor, α, determines the rate at which we will discard the last available true observation. A smaller α corresponds to a greater influence of previous model values, leading to a more smoothed-out series. Exponentiality is concealed within the recursive function. Each iteration involves multiplying the previous model value by (1-α), which also contains (1-α), and so on until the very start. The figure below illustrates several versions of exponential smoothing that utilise different α parameters. The result of smoothing and forecasting is shown in the figure below, which illustrates that at sufficiently high parameter values, the series is described very poorly, but still allows us to estimate the level and trend. At the same time, properly conducted cross-validation will allow us to select the optimal parameters for the model.

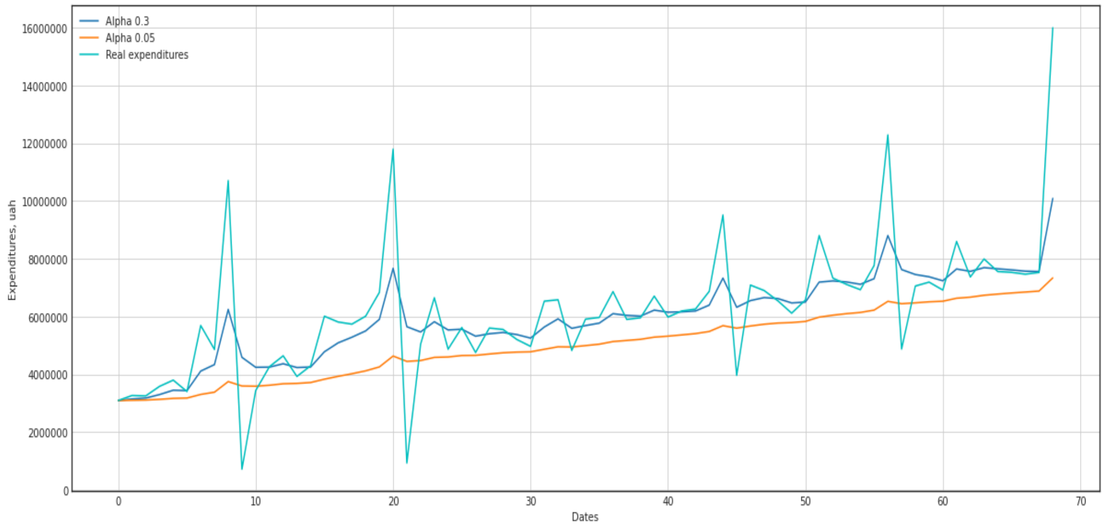

Figure 3: Exponential smoothing models forecasting

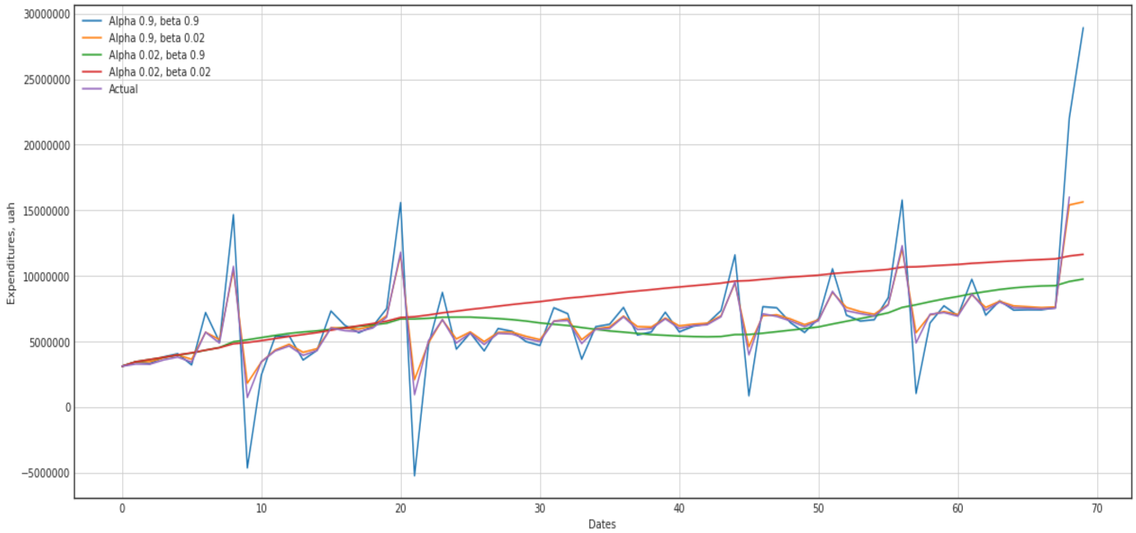

Figure 4: Double exponential smoothing models forecasting

The presented model value is a result of the weighted average of both the current true and previous model values. The smoothing factor, α, determines the rate at which we will discard the last available true observation. A smaller α corresponds to a greater influence of previous model values, leading to a more smoothed-out series. Exponentiality is concealed within the recursive function. Each iteration involves multiplying the previous model value by (1-α), which also contains (1-α), and so on until the very start. The figure below illustrates several versions of exponential smoothing that utilize different α parameters.

$$l_x = \alpha y_x + (1-\alpha)(l_{x-1} + b_{x-1}) \tag{3}$$

$$b_x = \beta(l_x - l_{x-1}) + (1-\beta)b_{x-1} \tag{4}$$

$$s_x = \gamma(y_x - l_x) + (1-\gamma)s_{x-L} \tag{6}$$

$$\widehat{y_{x+m}} = l_x + mb_x + s_{x-L+1+(m-1)modL} \tag{7}$$

Now we can move on to building a seasonal ARIMA model (SARIMA), which has the general form ARIMA *(p,d,q)x(P,D,Q)S*, where:

p - order of the autoregressive component AR(p);
d - degree of differencing of the original data;
q - order of the moving average component MA(q);
P - order of the seasonal autoregressive component SAR(P);
D - degree of differencing of the seasonal component;
Q - order of the seasonal moving average component SMA(Q);
S - seasonality (month, quarter, year).

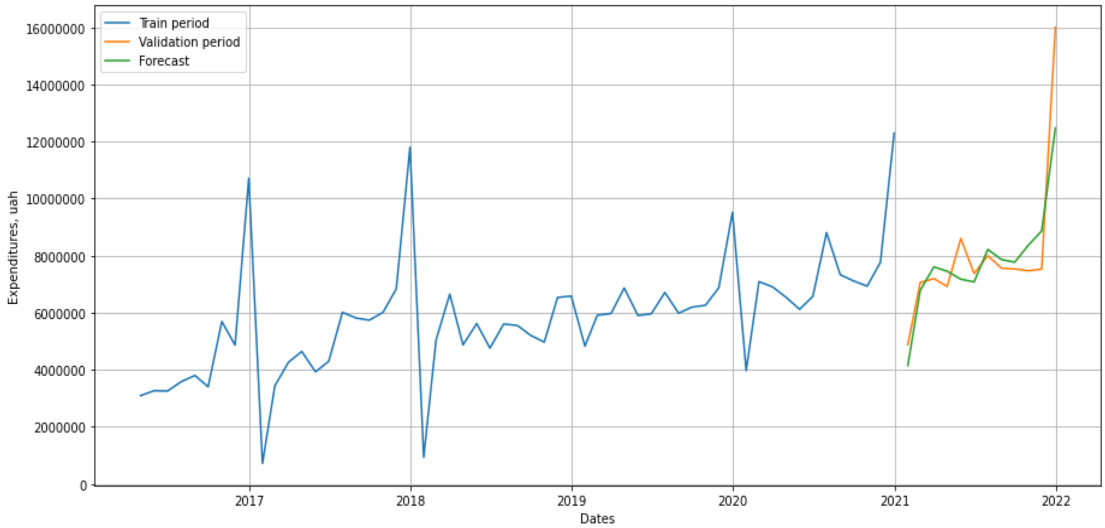

Figure 5: Triple exponential smoothing model forecasting (Holt-Winters)

We have considered parameters ranging from 0 to 5 and applied the Akaike criterion to assess the quality of the model in terms of the amount of information lost. Our analysis has led to the following model:

$$ARIMA\ (p, d, q) \times (P, D, Q)\ S = ARIMA\ (0, 0, 0) \times (3, 1, 2)12 \quad (8)$$

The graph below shows the results of our model testing. Based on the approximation parameters, the model has successfully captured the seasonality. However, it has a significant bias, resulting in poor explanation of variance in the data. Notably, the model has accurately captured anomalous peak values during this period.

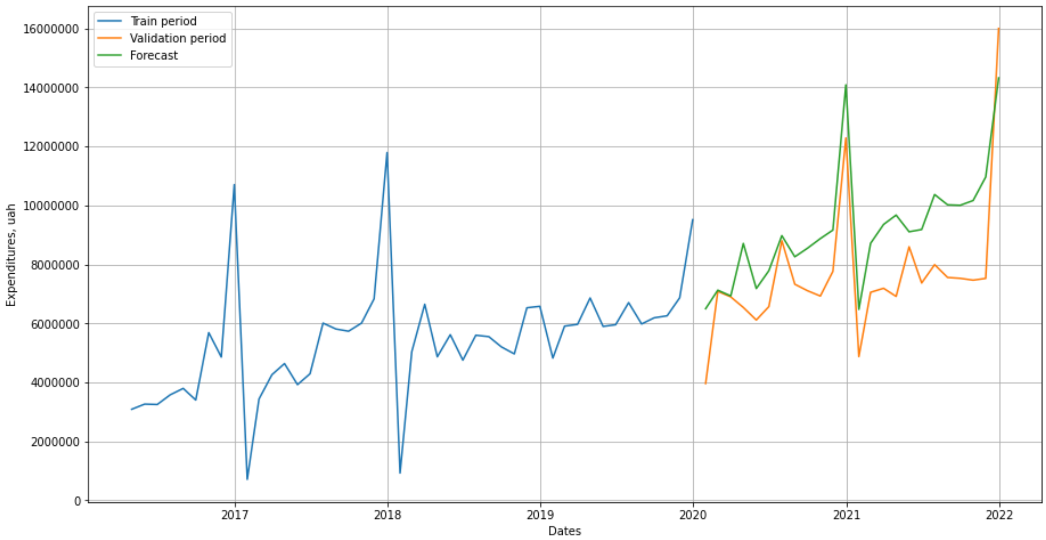

Figure 6: SARIMA forecasting

In 2017, Sean Taylor and Benjamin Letham from Facebook proposed an innovative approach for time series forecasting. It is fully based on simple linear models, but supplemented and extended with additional components. These include *g(t)* - a non-linear trend model that allows for considering trends in the context of forecast horizons, *s(t)* - seasonality, and *h(t)* - holidays and important events, as well as an error term *e*.

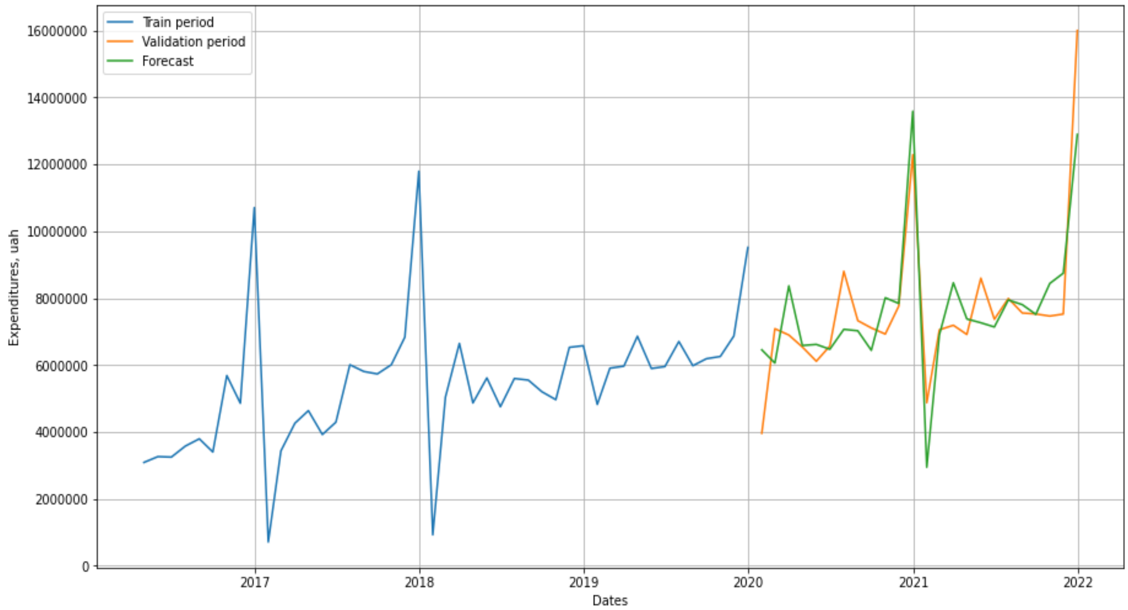

Figure 7: Prophet forecasting

The figure above displays the outcomes of the model, revealing findings closely resembling actual data. It is worth observing that the Prophet precisely captured peak values via the seasonal component. This unique methodology should be taken into account for our ensemble, since it differs greatly from prior options, and thus furnishes unconventional insights.

However, time series models have not gained widespread popularity due to the need to train multiple models in parallel, store and retrain them when a large number of series need to be forecasted. Therefore, classical machine learning models are the commonly preferred approaches.

Firstly, in order to predict the linear correlation between expenses and time, we should contemplate utilizing linear regression as it potentially has less variation and subsequently less bias which can compensate for the involvement of non-linear models. Of course, the downside of this technique is the requirement for synthetic or expert modeling of factors that can enhance the comprehensiveness of the model. During the factor engineering phase, we encoded the year and month with the corresponding period's average value. Additionally, we selected the optimal number of lag values using a selection method that reflects the current data's dependence on previous values in the series. We chose lag values commencing from periods 11 to 22. By optimizing the sum of squared deviations of predicted values from observed values, we will derive a function with a general form.

$$y_t = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + e_t \tag{9}$$

The figure above presents a comparison between predicted values and validation data, indicating a satisfactory forecast accuracy. This can be attributed to the model's reflection of the general trend and seasonality of actual data. An advantage of this approach is the ability to interpret the results of models not only by their forecast quality, but also by the significance of factors affecting the final result. One of the current methodologies that can be attributed to SOTA (State of the Art) is the utilization of SHAP (SHapley Additive exPlanations). This approach implements game theory to elucidate the outcomes of any machinelike model, relating the optimum distribution of model weights to explanations using classical Shapley values from game theory.

The above graph displays items ordered by the total SHAP values for all examined samples, to illustrate the distribution of each factor's influence on the model outcomes. The function's values are shown by color, with red indicating higher values and blue indicating lower ones. The results indicate that the mean expenditure figures for months and years are most impactful and increase the overall result. Conversely, lagged values shift it in the opposite direction.

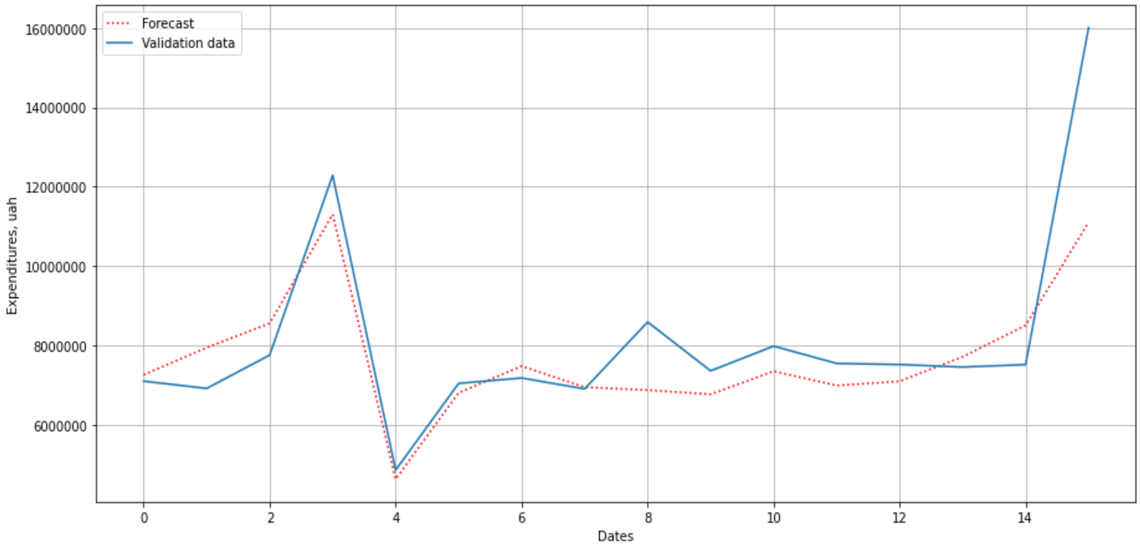

Figure 8: Regression forecasting

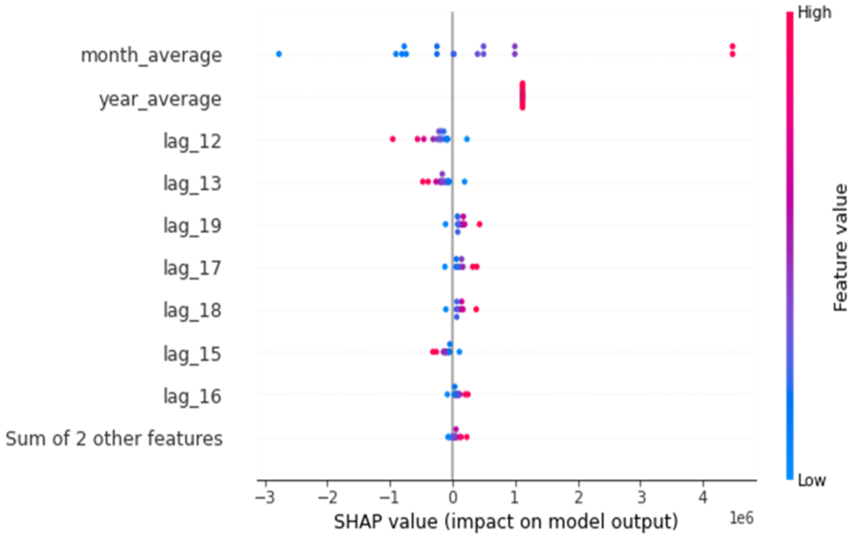

Figure 9: SHAPE values for the regression model

Another model that has gained popularity in machine learning is XGBoost. At its core, XGBoost utilizes a gradient-boosting algorithm to create model ensembles. Gradient boosting is widely utilized in machine learning for both classification and regression applications. It constructs a prediction model as an ensemble of simpler and less powerful models, including decision trees. The ensemble model is trained sequentially, unlike bagging. At each iteration, the prediction deviation on the training sample is calculated. The next model to be included in the ensemble will consider these deviations. Thus, by adding the forecast of a new decision tree to the ensemble forecast, we can decrease the average deviation of the model, which is the target optimization task. Training continues until the overall prediction error of the model continues to decrease or until the early stopping criteria are fulfilled.

The graph indicates that the boosting model has a slightly biased prediction accuracy with a small variance. Therefore, we propose including it in our "metamodel". For this case, we presented confidence intervals at a significance level of 0.05. Notably, values for the 3 and 15 periods constitute anomalies within this series. However, in the former scenario, there is a stronger inclination towards the prediction corridor, whereas the latter situation presents a significant distinction. This variance can be clarified by the existence of considerable residuals of financial resources and accounts that have been reallocated among subordinate units during recent months, leading to a surge in expenditure volume. Furthermore, expenditures remained constant throughout the year. To choose the best three models for the ultimate "metamodel," we utilized quality metrics: the mean absolute error (MAE) and the mean absolute percentage error (MAPE), which were computed using the formulas given below.

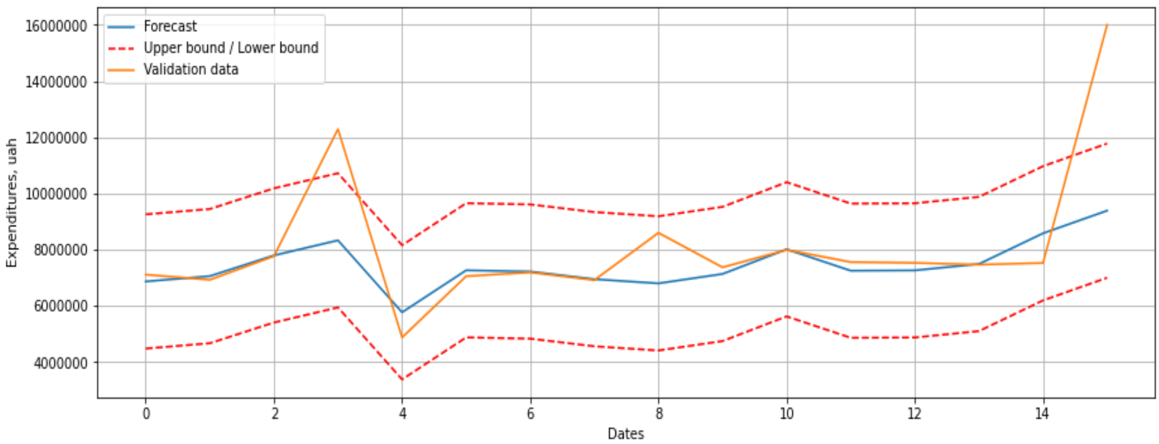

Figure 10: XGBoost model forecasting

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y_i}| \quad (10)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y_i}}{y_i}\right| \quad (11)$$

The accuracy metrics calculation results for the models are displayed in the following table. It is recommended to select a maximum of three basic models that have shown optimal performance for the meta-model. In this instance, we have opted for the Holt-Winters triple exponential smoothing linear model, linear regression, and gradient boosting model. This is attributable not only to the superior MAE and MAPE indicators but also to the fundamental congruence of the models' approaches, which serve to mutually reinforce and mitigate overall forecast bias and variance.

$$\hat{y_t} = 0.5 \times \widehat{y_{hwm}} + 0.2 \times \widehat{y_{lr}} + 0.3 \times \widehat{y_{xgb}} \quad (12)$$

Table 1

MAE and MAPE result comparison for different models

Metrics	Holt-Winter	SARIMA	Prophet	Regression	XGBoost
MAE	606972	1682813	812210	595491	567103
MAPE	8.42%	24.50%	12.49%	7.81%	7.16%

To enhance the prediction of indicators, a cross-validation method was proposed for selecting weights and verifying final metrics. We examined the weight distribution for 50 combinations in the range of 0 to 1. Upon comparison, the best result yielded a mean absolute error of 452235 UAH and a mean absolute percentage error of 5.51%. This outcome significantly outperformed the independent forecasts of the models. The formula used to forecast our models, along with their respective weights, is presented below. Furthermore, the final forecast of the model has been illustrated in the figure provided. The methodologies and models outlined, with their successful implementation, are poised to enhance budgetary planning and forecasting processes, as well as departmental expense allocation and cost optimization. The implementation of these innovative methodologies is facilitated by the development of cutting-edge information-analytical systems, which necessitates minimization of transaction costs. Forecasting is an essential aspect of the budgeting process, allowing analysts to project future spending and resource needs. Through utilizing data science approaches to create more precise and thorough forecasts, the Armed Forces can allocate resources with greater effectiveness, diminish waste and inefficiencies, and advance mission readiness.

One instance of employing modeling and forecasting in the budgeting process of the Armed Forces involves producing prognostic models that anticipate the expenditure for equipment maintenance and repair. This is achieved by scrutinizing past data on incurred expenses for maintenance and repair, and applying machine learning algorithms to assess the determinants that influence these costs. Experts can subsequently construct models to forecast forthcoming maintenance and repair costs derived from equipment usage and additional variables. These models enable the development of more precise maintenance and repair budgets, mitigating the risk of unforeseen costs and equipment malfunctions.

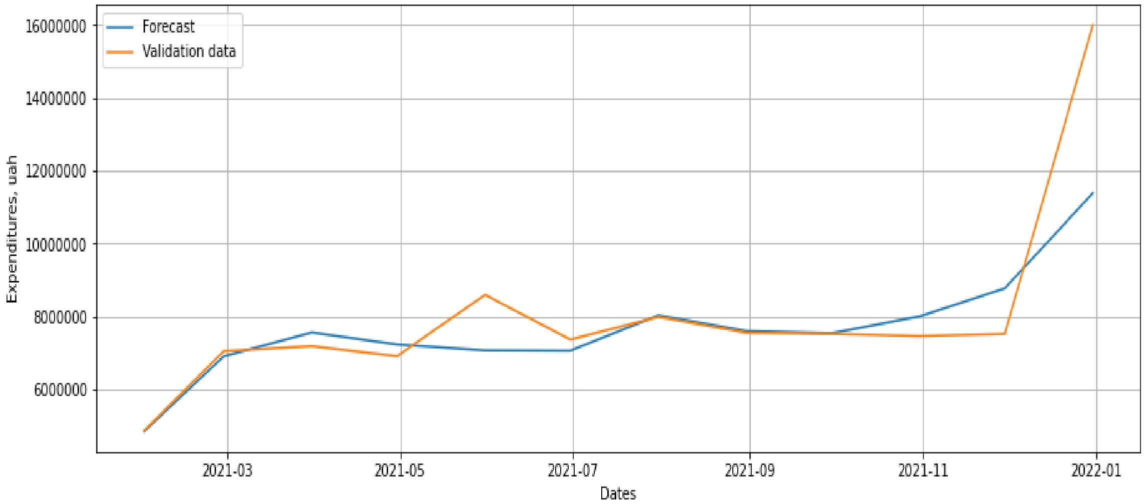

Figure 11: Metamodel forecasting

Modeling and forecasting are essential constituents of the data science methods utilized in the Armed Forces' budgeting process. Through applying these techniques, the Armed Forces can generate highly detailed and accurate forecasts, allocate resources more efficiently, minimize wastage and inefficiencies, and augment mission readiness.

6. Discussion

This section examines the challenges and opportunities linked to integrating data science techniques in military budgeting. Stakeholder participation and change management are crucial, and we will discuss these topics in detail. Additionally, we will explore the advantages of incorporating data science methods, such as elevated efficiency and accuracy, and also the potential hurdles, such as specialized expertise requirements and difficulties with data quality. Additionally, we will explore the advantages of incorporating data science methods, such as elevated efficiency and accuracy, and also the potential hurdles, such as specialized expertise requirements and difficulties with data quality. Additionally, we will explore the advantages of incorporating data science methods, such as elevated efficiency and accuracy, and also the potential hurdles, such as specialized expertise requirements and difficulties with data quality. The implementation of data science techniques in the budgeting process of the Armed Forces brings both advantages and disadvantages. One of the essential challenges is obtaining vast amounts of high-quality data. The implementation of data science techniques in the budgeting process of the Armed Forces brings both advantages and disadvantages. Data science techniques are only as accurate and effective as the data they use. Therefore, the Armed Forces must ensure they can access reliable, comprehensive, and prompt data, in order to optimize their budgeting procedure. Significant investment may be required in data collection and management systems.

A further challenge is the necessity of specialized skills and expertise. Data science methods require professionals to possess specialized skills in statistical analysis, data mining, and machine learning. The Armed Forces should invest in training and hiring personnel with these specialized skills to optimally incorporate data science methods into their budgeting process. Moreover, implementing data science methods within the Armed Forces budgeting process may face cultural and organizational challenges. Some stakeholders may resist change due to their familiarity with traditional budgeting methods. As a result, the Armed Forces must put in place extensive communication and educational initiatives to encourage endorsement and support for data science methods. Despite the challenges, integrating data science techniques into the budgetary processes of the Armed Forces presents significant opportunities. These methods facilitate more precise and comprehensive forecasts, which can enhance resource allocation and mission readiness. By identifying potential areas of cost savings and increasing efficiency, the use of data science methods can aid the Armed Forces in reducing waste and improving the value of their operations. Additionally, the integration of data science techniques can enable the Armed Forces to maintain a competitive edge in a constantly evolving technology-driven environment. As data volume and complexity grow, data science approaches will become more essential for successful budgeting and decision-making in the Armed Forces.

The integration of data science procedures into Armed Forces budgeting presents a range of challenges and opportunities. Although challenges may arise, the advantages of precise predictions, enhanced allocation of resources, and reduced expenses highlight the potential for data science techniques to transform the military's budgeting approach.

7. Conclusion

The integration of data science methods into military budgeting processes presents opportunities to optimize resource allocation, reduce costs, and bolster mission readiness. Advanced data science techniques allow for sophisticated analysis of extensive datasets, enabling military forces to generate more accurate forecasts to inform budget planning. Data-driven insights can identify areas where operational costs may be reduced without compromising effectiveness. Overall, incorporating data science into budgetary procedures has considerable potential to enhance fiscal efficiency, resource adequacy, and mission preparedness for the military. With investments in data infrastructure, management systems, and personnel training, the challenges of implementing these advanced technical capabilities can be overcome. The promise of data science approaches suggests they could fundamentally transform military budgetary processes and decision-making.

Further research could explore questions such as what data infrastructure and management systems optimally support military budgetary processes or can data science budgetary systems be securely implemented to protect sensitive financial and operational data. Investigating these and other issues can help guide effective adoption of data science methods for military budgeting.

8. References

[1] K.H. Naveen, “Predictive model building for driver-based budgeting using machine learning,” in Proc. of International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 567-575.

[2] W. Li et al., “Privacy Budgeting for Growing Machine Learning Datasets,” in Proc. of IEEE INFOCOM 2021-IEEE Conference on Computer Communications, 2021.

[3] H. Jang, “A decision support framework for robust R&D budget allocation using machine learning,” Decision Support Systems, vol. 121, p. 1-12, 2019.

[4] A. Faccia, “Big Data-driven Budgeting and Business Planning,” Preprints 2020 doi: 10.20944/preprints202009.0747.v1.

[5] R.C. Huacarpuma et al., “Big data: A case study on data from the Brazilian ministry of planning, budgeting and management,” in Proc. of IADIS Applied Computing Conference (IADIS), 2013.

[6] B. Shen et al., “KPI-Driven Predictive ML Models Approach Towards Municipal Budgeting Optimization,” arXiv preprint arXiv:1510.05767v2 [cs.LG], 2015.

[7] I. E. Fisher, M. R. Garnsey and M. E. Hughes, 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) (2016): 157-214.

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

[9] J. Davies, M. Arana-Catania, R. Procter, R., F.A. van Lier, Y. He, 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, 2021, pp. 362-366.

[10] A. Tiron-Tudor et al., “Emerging Technologies’ Contribution to the Digital Transformation in Accountancy Firms,” Electronics, vol. 11, no. 22, p. 3818, 2022.

[11] A. Eltweri et al., “Applications of Big Data within Finance: Fraud Detection and Risk Management within the Real Estate Industry,” in Proc. of E-Business and E-commerce Engineering (EBEE), 2021.

[12] F. Aboagye-Otchere et al., “A Review of Big Data Research in Accounting,” Intelligent Systems in Accounting, Finance and Management, vol. 28, no. 4, pp. 268–283, 2021.