PUBLISHER

Canadian Institute of Technology
Address: St. Xhanfize Keko, No. 12 Tirana, Albania

Contact: editorialoffice@cit.edu.al

TABLE OF CONTENTS

WELCOME REMARKS

Prof. Dr.
ISMAIL KOCAYUSUFOĞLU
Editor-in-Chief

Dear Readers,

As the Chief Editor of CRJ, it is my distinct pleasure to welcome you to our May' 24 issue. This edition marks a significant step in our ongoing commitment to providing you with the latest advancements, insightful research, and thought-provoking discussions within our field.

The month of May often symbolizes renewal and growth, and we have embraced this spirit in our latest issue. Our editorial team has worked diligently to curate a collection of articles that not only reflect the current trends and innovations but also anticipate future directions in our discipline. This issue is particularly special as it features a diverse array of topics, showcasing the multidisciplinary nature and global reach of our journal.

We are proud to feature contributions from emerging scholars; whose fresh perspectives and innovative methodologies are a testament to the vibrant future of our field. Their work not only enriches our current issue but also underscores our commitment to supporting the next generation of researchers.
I extend my heartfelt gratitude to our authors, reviewers, and editorial board members. Their unwavering dedication and expertise are the backbone of our journal, ensuring that we maintain the highest standards of academic excellence.

Finally, I encourage you, our readers, to engage with the content, share your thoughts, and join the conversation. Your feedback is invaluable and helps us to continually improve and evolve.

Thank you for your continued support and readership. We look forward to your engagement with this issue and hope that it inspires, informs, and challenges you in equal measure.

Warm regards,

Prof. Dr. Ismail KOCAYUSUFOĞLU
Rector

EMPIRICAL ANALYSIS OF THE TWIN DEFICITS HYPOTHESIS IN THE REPUBLIC OF NORTH MACEDONIA

Merale Fetahi-Vehapi [1]**,** ***Fatbardha Jonuzi*** [2] **,*****Florenta Jonuzi*** [3]

[1] Department of Finance and Accounting, Faculty of Economy, University of Tetovo, North Macedonia, merale.fetahi@unite.edu.mk, ORCID: 0000-0002-0572-891X

[2] Kicevo Municipality, North Macedonia, fatbardhajonuzi@gmail.com, ORCID: 0009-0009-1852-2054

[3] Master student in the Department of Finance, Banking and Insurance, Faculty of Economy, University of Tetovo, North Macedonia, f.jonuzi211595@unite.edu.mk, ORCID:0009-0001-1713-449X

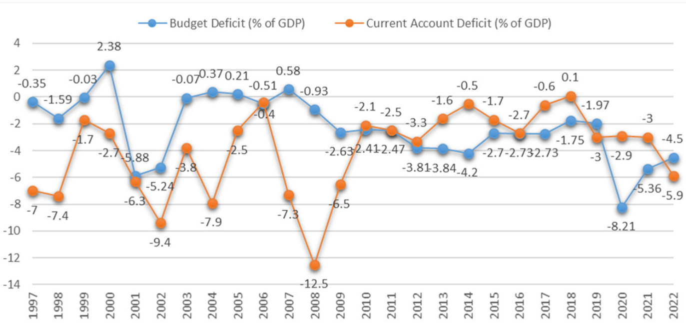
Abstract
The twin deficits hypothesis is widely considered one of the most frequently employed phenomena in the economic literature. An econometric analysis of the twin deficit hypothesis is of special importance in understanding the perspective on macroeconomic stability in the Republic of North Macedonia. This paper aims to empirically test the validity of this hypothesis in the Republic of North Macedonia. To do so, we utilized quarterly data on Macedonia's budget deficit, the current account deficit, exchange rate, interest rate, GDP, government expenditure, and money supply, starting from the first quarter of 2001 to the fourth quarter of 2022. Through the application of the ARDL model, the study found that between the variables taken into analysis, there exists a short and long-run relationship. More specifically exchange rate, government expenditure, and GDP result in improvement on the budget deficit, both in the short run and long run. While current account deficit, interest rate, and money supply result in worsening the budget deficit, both in the short run and long run.

Keywords: North Macedonia; twin deficits; budget deficit; current account deficit; ARDL model

1. INTRODUCTION

Economies with both fiscal and current account deficits are commonly known as having "twin deficits". According to the twin deficit hypothesis, there exists a long-term positive relationship between the budget and the current account deficit, with the budget deficit causing the current account deficit. The importance of this phenomenon rose in the 1980s because of the rapid expansion of the twin deficits in the United States and many other countries around the world.
Analyzing the empirical relationship between budget and current account deficits is important for EU candidate countries and those with aspirations for potential candidacy. Over the past two decades, North Macedonia has faced concurrent budget and current account deficits:
· In 2022, North Macedonia registered a budget deficit amounting to 4.50% of its GDP. Over the period from 1997 until 2022, the budget deficit averaged -2.32% of GDP, reaching its highest point of 2.38% of GDP in 2000 and its lowest point of -8.21% of GDP in 2020, and
· In 2022, North Macedonia reported a current account deficit amounting to 6% of its GDP. The average current account to GDP ratio in North Macedonia from 1997 to 2022 was -4.04%, reaching its peak at 0.10% of GDP in 2018 and its lowest point at -12.5% of GDP in 2008.

Figure 1.
Budget Deficit as a percentage of the GDP and Current Account Deficit as a percentage of the GDP



Source: Authors based on data from the trading economics

In addition to the introduction. The following sections of the paper are structured in the following manner: the following section provides a literature review, followed by the description of data and methodology in Section three. Section four presents and deliberates on the empirical findings. Lastly, the conclusion is presented in Section five.

*Corresponding author:

2. LITERATURE REVIEW

In this section, we analyze relevant empirical literature, discussing four testable hypotheses regarding the twin deficit.

The first hypothesis, following Keynesian theory, proposes that an increase in the budget deficit will lead to a similar increase in the current account deficit. However, variations among countries and the evolution of their respective economies may lead to different levels of impact. Several empirical studies provide evidence in support of this perspective. For example, Harko (2009) estimates the causal link between twin deficits and other macroeconomic variables by using multivariate time series data for Pakistan. The estimates from the VAR model indicate that the causal relationship of deficit flows comes from the budget deficit to prices, then to interest rates, capital flows, exchange rates, and finally to the trade deficit. Additionally, there is evidence to suggest that reducing budget deficits could contribute to moderating price levels. Tarawalie (2014) found evidence supporting the Twin Deficits Hypothesis for Sierra Leone through the use of a multivariate model instead of direct comparison between the budget deficit and the current account deficit.

The second hypothesis, known as the Recardian Equivalence Hypothesis, suggests that a change between taxes and budget deficits has no effect on the real interest rate, investment quantity, or the current account balance. Using monthly data and the bounds-testing approach to co-integration, Ratha (2009) finds that the twin-deficits theory holds for India in the short run. But, there's no such correlation in the long run. Hence, it's concluded that the Keynesian perspective dominates in the short run, while the neo-classical theory dominates in the long run. Rahman and Mishra (2001) found that budget and current account deficits have no possibility of reverting to a long-run equilibrium relationship in the United States during the period 1946–1988.

The third hypothesis suggests a one-way causality, implying that the budget deficit is influenced exclusively by the current account deficit. Based on this statement, Alkswani (2002) examines the relationship between twin deficits in the Saudi Arabian economy employing annual data and argues that in the oil economy, neither the REH nor the Keynesian proposition is valid. Alkswani contends that export revenue affects government income, expenditure, and the exports of goods and services. They conclude that the two deficits are positively correlated, with causality running from the trade deficit to the budget deficit. Empirical evidence from Saeed and Khan (2012) indicates long run relationship between the two deficits in Pakistan, with causality running from the current account deficit to the budget deficit during the period 1972 to 2008. Sobrino (2013), analyzing quarterly data from 1980 to 2012, rejects the twin deficits hypothesis for Peru. His findings strongly imply reverse causality, suggesting that the current account deficit affects the fiscal deficit.

The last hypothesis addresses the two-way causality between the budget deficit and the current account deficit. Lau and Tang (2009) find that there is a bi-directional causality between the budget deficit and the current account deficit in Cambodia. Pahlavani and Saleh (2009) investigated the twin deficits for the Philippines and Mukhtar, Zakaria, and Ahmde (2007) for Pakistan. Both studies confirm a two-way causality among budget deficit and current account deficit. Lau and Baharumshah (2006) argue that interest rates, exchange rates, and budget deficits play an important role in explaining the current account balance. Omoniyi et al. (2012), studied the Twin Deficits Hypothesis in Nigeria from 1970 to 2008 and presented findings suggesting a two-way causality between budget deficits and trade deficits.

Regarding our country, we have the studies of:

- Rilind Ademi and Zana Beqiri Luma (2023) tested the twin deficit hypothesis's validity of North Macedonia by analyzing quarterly data on the budget and trade deficit from 2006 to 2021. Their use of Granger causality and VAR model yielded partially supportive results for the twin deficit hypothesis. The observed weak relationship among budget and trade deficit is not unexpected for a small open economy like North Macedonia, characterized by a fixed exchange rate and foreign direct investments primarily influenced by business conditions rather than the interest rates.
- Vesna Bucevska (2020) tested the validity of the twin deficit hypothesis by applying multiple econometrics techniques, including Granger causality, VAR model, and VECM model. Her analysis indicates that efforts aimed at reducing current account imbalances via fiscal policy may not be effective in the short run.
- Vesna Stojcevska and Mite Miteski (2016) investigated empirically whether the proposition of a positive relationship between the budget balance and the trade balance, i.e. the twin deficit hypothesis, holds in the case of the Republic of Macedonia. The study was conducted using the VAR model, following the methodology established by Kim and Roubini (2007). The key empirical findings indicate a positive relationship among the selected trade and fiscal variables, suggesting a potential relationship among budget balance and trade balance.

3. DATA AND METHODOLOGY

This study attempts to explore the twin deficits hypothesis in the context of North Macedonia utilizing quarterly data. The study spans from the first quarter of 2001 to the fourth quarter of 2022 and relies on secondary data sourced from annual reports and various publications of the National Bank of North Macedonia.

The variables employed in the empirical analysis are defined as follows:

• CAD = Current account deficit, as the sum of net exports of goods and services and net factor income.
• BD = The budget deficit, as the difference of government revenue and government expenditure.
• IR = Interest rate represented by the nominal deposit rate, (as %).
• ER = The exchange rate represents the comparative value of one currency expressed in relation to another currency (EUR/MKD).
• GDP = GDP at constant prices (REAL GDP) - reference year 2005 (in million denars).
• GE = Government expenditure at current prices (as % of GDP).
• M1 = an indicator of the money supply reflecting the liquidity within the economy.

The model has the following form:

$$BD = \beta_0 + \beta_1 CAD + \beta_2 IR + \beta_3 ER + \beta_4 GDP + \beta_5 GE + \beta_6 M1 + \varepsilon$$

The model has the following form:

Table 1

Variables and their expected signs based on the theories

Variable	Notation	Units	Expected Relation
Budget deficit	BD	Million MKD	
Current account deficit	CD	Million MKD	Positive (+)
Interest rate	IR	%	Positive(+) / Negative (-)
Exchange rate	ER	MKD	Positive (+)
Gross Domestic Production	GDP	Million MKD	Negative(-)
Government expenditure	GE	% of GDP	Negative (-)
Money supply	M1	Million MKD	Positive (+)

Source: Author's illustration

4. EMPIRICAL RESULTS

Since we are dealing with time series data, it's crucial to assess its properties. To assess the unit root among the variables, the ADF Test is performed.
The unit root tests presented in Table 2 reveal a combination of I(0) and I(1) order of integration. Some variables have zero-order integration, such as the budget deficit, current account deficit, interest rate, and, government expenditure. Whereas the other variables have a first-order integration like exchange rate, GDP, and money supply (M1).

Table 2

Findings from of the ADF Test

Variables	Lags	Level		First-Difference		Result
		t	*5%*	*t*	*5%*	
BD	2	-2.901	-2.902	-11.631	-2.903	I(0)
CAD	4	-3.391	-2.904	-4.975	-2.904	I(0)
IR	1	-4.416	-2.901	-5.491	-2.902	I(0)
ER	1	-2.758	-2.901	-8.591	-2.903	I(1)
GDP	4	-1.402	-3.535	-5.509	-2.905	I(1)
GE	4	-4.550	-2.907	-2.049	-2.904	I(0)
M1	0	3.090	-2.901	-2.991	-2.901	I(1)

Source: Researcher's calculation using Stata14

Since there are both order-zero and first-order stationary variables in the series, the ARDL model was used as a more suitable model to perform the empirical analysis. The ARDL bounds technique was applied to assess the cointegration or the long-run relationship by comparing F-statistics with the specific critical values for the period from 2001 to 2022.
According to the results of Table 3, it surmised the existence of a long-run relationship among budget deficit, current account deficit, interest rate, exchange rate, government expenditure, GDP, and money supply. With the F statistic surpassing the critical value of the upper bounds (I(1)) at significance levels, we can reject the null hypothesis.

Table 3

ARDL Bound Test of Cointegration

Number of regressors	Test statistic	Critical Value Bounds							
		10%		5%		2.5%		1%	
k	**F- statistic**	I (0)	I (1)	I (0)	I (1)	I (0)	I (1)	I (0)	I (1)
6	9.330	2.12	3.23	2.45	3.61	2.75	3.99	3.15	4.43
k	**t-statistic**								
6	-5.701	-2.57	-4.04	-2.86	-4.38	-3.13	-4.66	-3.43	-4.99

Source: Researcher's calculation using Stata14

As we have detected a long-run relationship. In progress, it is going to be checked the effect of the current account deficit, interest rate, exchange rate, GDP, government expenditure, and money supply on budget deficit in the long run.

Table 4 presents the findings for the long-run coefficients of our ARDL model. According to the findings, the budget deficit is positively affected by the current account deficit, interest rate, and GDP. While it is negatively affected by the exchange rate, government expenditure, and money supply. A million denar rise in the current account deficit results in a 17.47 million denar rise in the budget deficit in the long run, ceteris paribus. A 1% rise in the interest rate results in a 983 million denars rise in the budget deficit in the long run, ceteris paribus. A million denar rise in the GDP results in a 0.299 million denar rise in the budget deficit in the long run, ceteris paribus. A 1% rise in the exchange rate results in a 17851 million denar reduction in the budget deficit in the long run, ceteris paribus. A million denar rise in government expenditure results in a 996 million denars reduction in the budget deficit in the long run, ceteris paribus. A million denar rise in the money supply results in a 0.047 million denars reduction in the budget deficit in the long run, ceteris paribus.

Table 4
The long-run coefficients estimated by the ARDL model

	Coefficient	Standard Errors	t	p
CAD	17.46636	5.812726	3.00	0.004
IR	983.3241	430.1644	2.29	0.026
ER	-17851.24	4429.19	-4.03	0.000
GDP	0.2988884	0.0791987	3.77	0.000
GE	-996.4996	382.5934	-2.60	0.012
M1	-0.0471629	0.0141642	-3.33	0.002

Source: Researcher's calculation using Stata14

To analyze the short-run movement's tendency on long-run equilibrium, the ECM was used. Table 5 presents the findings for the short-run coefficients of our ARDL model. Furthermore, the short-run results show that there exists a significant effect of the previous budget deficit, the current account deficit, the lags of the previous interest rate, the lags of the previous exchange rate, the current GDP, and the lags of the previous GDP, the lags of the previous government expenditure, and the current money supply and the lags of the previous money supply.

Table 5
The short-run coefficients estimated by the ARDL model

Variables	Coefficient	Standard Errors	t	p
$D(BD_{t-1})$	-0.3063539	0.1145497	-2.67	0.010
D(CAD)	-8.971048	3.308936	-2.71	0.009
D(IR)	-540.3945	719.5372	-0.75	0.456
$D(IR_{t-1})$	-510.0563	702.0378	-0.73	0.471
$D(IR_{t-2})$	-2448.754	815.2928	-3.00	0.004
$D(IR_{t-3})$	-1528.178	642.8127	-2.38	0.021
D(ER)	-272.8927	5522.632	-0.05	0.961
$D(ER_{t-1})$	561.249	5067.7	1.00	0.323
$D(ER_{t-2})$	16834.57	4891.849	3.44	0.001
$D(ER_{t-3})$	11629.65	5217.378	2.23	0.030
D(GDP)	0.190993	0.1063222	1.80	0.078
$D(GDP_{t-1})$	0.4707283	0.1157968	4.07	0.000
$D(GDP_{t-2})$	0.6437658	0.122321	5.26	0.000
$D(GDP_{t-3})$	0.558236	0.1088065	5.13	0.000
D(GE)	814.6386	498.9487	1.63	0.109
$D(GE_{t-1})$	1764.947	458.2273	3.85	0.000
$D(GE_{t-2})$	1750.688	381.8274	4.59	0.000
$D(GE_{t-3})$	856.9683	293.2239	2.92	0.005
D(M1)	-0.3874849	0.1247578	-3.11	0.003
$D(M1_{t-1})$	-0.2862412	0.1297733	-2.21	0.032
ADJ(ECM)	-1.023593	0.1795419	-5.70	0.000

Source: Researcher's calculation using Stata14

The short-run coefficient of the previous budget deficit has a negative effect on the current budget deficit, i.e. 1 million denar rise of the first lags of the budget deficit results in a 0.306 million denar reduction in the current budget deficit, ceteris paribus.

The budget deficit is negatively affected by the short-run coefficient of the current account deficit, i.e. 1 million denar rise in the current account deficit results in an 8.97 million denar reduction in the budget deficit, ceteris paribus.

The budget deficit is negatively affected by the short-run coefficient of the previous interest rate, along with the second and third lags of the interest rate, i.e. 1% rise of the second lags interest rate results in a 2448.754 million denar reduction in the budget deficit, ceteris paribus; and 1% rise of the third lags of the interest rate results in a 1528.178 million denar reduction in the budget deficit, ceteris paribus.

The budget deficit is affected positively by short-run coefficient of the previous exchange rate, along with the second and third lags of the exchange rate, i.e. a million denar rise of the second lags of the interest rate results in a 16834.57 million denar rise in the budget deficit, ceteris paribus; and a million denar rise of the third lags of the exchange rate results in a 11629.65 million denar rise in the budget deficit, ceteris paribus. The short-run coefficient current GDP and the lags of the previous GDP have a positive effect on the budget deficit, i.e. a million denar rise in the current GDP results in a 0.1909 million denar rise in the budget deficit, ceteris paribus; a million denar rise of the first lags of the GDP results in a 0.4707 million denar rise in the budget deficit, ceteris paribus; a million denar rise of the second lags of the GDP results in a 0.6476 million denar rise in the budget deficit, ceteris paribus; a million denar rise of the third lags of the GDP results in a 0.5582 million denar rise in the budget deficit, ceteris paribus.

The short-run coefficient of the previous government expenditure, such as the first, second, and third lags of the government expenditure have a positive effect on budget deficit; a 1% rise of the first lags of the government expenditure results in a 1764.947 million denar rise in the budget deficit, ceteris paribus; 1% rise of the second lags of the government expenditure results in a 1750.688 million denar rise in the budget deficit, ceteris paribus; a 1% rise of the third lags of the government expenditure results in a 856.968 million denar rise in the budget deficit, ceteris paribus.

The budget deficit is negatively affected by short-run coefficient of the current money supply and the lags of the previous money supply; i.e. a million denar rise of the current money supply results in a 0.3875 million denar reduction in the budget deficit, ceteris paribus; a million denar rise of the first lags of the money supply results in a 0.2862 million denar reduction in the budget deficit, ceteris paribus.

Finally, the negative ECM coefficient (−1.023593) and the high speed of adjustment (-5.70) have worked to bring equilibrium to the economy, with exogenous and endogenous shocks contributing to its restoration over a long period.

Table 6 presents the results of diagnostic tests conducted to confirm that our model is free from issues such as serial correlation, heteroskedasticity, normality, multicollinearity, and correct specification.

Table 6
Diagnostic Tests Results

Tests	Chi-squared value	prob
Autocorrelation Test		
– Breusch-Godfrey	3.864	0.0510
Heteroskedasticity Test		
– Breusch-Pagan	1.04	0.3085
Normality Test		
– Skewness/Kurtosis	3.98	0.1365
Multicollinearity Test		
– VIF	7.27	
Model Specification Test		
– Ramsey RESET	2.44	0.0751

Source: Researcher's calculation using Stata 14

The probability of all tests is higher than 0.05 or 5% which means that our model is free from autocorrelation problems, and hetero-scedasticity problems and the model is normal. Ramsey RESET test value (0.0751) is greater than 0.05 so this means that there is no functional misspecification in the model or other words the function of the model is correct. In terms of VIF test for multicollinearity, the resulting value is below 8, suggesting that multicollinearity is not present.

5. CONCLUSION

This paper aimed to empirically assess the validity of the twin deficit hypothesis in North Macedonia. To accomplish this goal, we employed quarterly data for the time period from 2001 to 2022 on Macedonia's budget deficit, the current account deficit, exchange rate, interest rate, GDP, government expenditure, and money supply. Our analysis was conducted using the ARDL model.

The empirical findings suggest the existence of short-run and long-run relationships among the budget deficit, current account deficit, exchange rate, interest rate, GDP, government expenditure, and money supply.

In the short run, the budget deficit is negatively affected by the previous budget deficit, the current account deficit, the previous interest rate, the current money supply, and the previous money supply. On the other hand, the budget deficit is positively affected by the previous exchange rate, the previous government expenditures, and current GDP and previous GDP.

In the long run, the budget deficit is positively affected by the current account deficit, interest rate, and GDP. While it is negatively affected by the exchange rate, government expenditure, and money supply.

The findings of the analyzed studies and this paper's results highlight the importance of government policymakers exercising caution in managing fiscal policies. Specifically, they should focus on controlling public expenditures, as these tend to outweigh benefits, leading to increased budget deficits and state debt.

It's also important to prioritize productivity growth and enhance competitiveness in foreign markets to achieve long-term positive impacts and reduce trade deficits. By focusing on initiatives such as investment in innovation, infrastructure, and education, countries can improve their ability to compete globally, leading to increased exports and decreased reliance on imports. This strategic approach can contribute to a more balanced trade position over time.

References

Alkswani, M. A. (2002). The Twin Deficits Phenomenon in Petroleum Economy: Evidence from Saudi Arabia. Discussion Papers of Department of Economics, King Saud University. Web site: https://www.mafhoum.com/press2/79E15.pdf

Anoruo, E. a. (1998). Current Account and Fiscal Deficits: Evidence from Five Developing Economies of Asia. Journal of Asian Economics, 9(3), 487-501. doi:https://doi.org/10.1016/S1049-0078(99)80099-2

Bucevska, V. (2020). Testing the Twin Deficit Hypothesis: Evidence from the Republic of North Macedonia. Business Systems Research, 11(3), 45-62. doi:https://doi.org/10.2478/bsrj-2020-0026

Darrat, A. F. (1988). Have Large Deficits Caused Rising Trade Deficits? Southern Economic Journal, 54(4), 879-887. doi:http://dx.doi.org/10.2307/1059523

Harko, A. N. (2009). Twin Deficits Causality Link – Evidence from Pakistan. International Research Journal of Finance and Economics, 24, 54-70. Web site: https://www.researchgate.net/publication/255481139_Twin_Deficits_Causality_Link-Evidence_from_Pakistan

Lau, E. a. (2006). "Twin Deficits Hypothesis in SEACEN Countries: A Panel Data Analysis of Relationships between Public Budget and Current Account Deficits. Applied Econometrics and International Development, 6(2), 213-226. Web site: https://www.usc.es/economet/journals1/aeid/aeid6215.pdf

Lau, E. a. (2009). Twin deficits in Cambodia: Are there Reasons for Concern? An Empirical Study. Discussion Papers of Monash University, Australia, 11(9). Web site: https://www.monash.edu/__data/assets/pdf_file/0011/925481/twin_deficits_in_cambodia_are_there_reasons_for_concern_an_empirical_study.pdf

Mukhtar, T. Z. (2007). An Empirical Investigation for the Twin Deficit Hypothesis in Pakistan. Journal of Economic Cooperation, 28(4), 63-80. Web site: https://www.researchgate.net/publication/253599258_AN_EMPIRICAL_INVESTIGATION_FOR_THE_TWIN_DEFICITS_HYPOTHESIS_IN_PAKISTAN

Pahlavani, M. a. (2009). Budget Deficits and Current Account Deficits in the Philippines: A Casual Relationship? American Journal of Applied Sciences, 6(8), 1515-1520. doi: https://doi.org/10.3844/ajassp.2009.1515.1520

Perera, A., & Liyanage, E. (n.d.). An Empirical Investigation of the Twin Deficit Hypothesis: Evidence from Sri Lanka. Staff Studies, 41(1 & 2), 41-87. doi: http://dx.doi.org/10.4038/ss.v41i1.4683

Rahman, M. a. (2001). Cointegration of U.S. Budget and Current Account Deficits: Twins or Strangers?". Journal of Economics and Finance, 16(2), 119-127. doi: https://doi.org/10.1007/BF02920112

Ratha, A. (2011). Twin Deficits or Distant Cousins? Evidence from India. Working Papers of Department of Economics, St. Cloud State University, USA. Web site: https://repository.stcloudstate.edu/econ_wps/5

A QUALITY MANAGERIAL APPROACH ON THE RELATIONSHIP BETWEEN THE DIGITAL DEVELOPMENT LEVEL AND THE CYBER SECURITY INDEX

Enriko Ceko [1]

[1] Business Administration and Information Technology Department, Faculty of Economy, Canadian Institute of Technology, enriko.ceko@cit.edu.al, ORCID: 0000-0002-3372-2785

Abstract
The main goal for the realization of this study was to verify the level of relations between digitalization and cyber security. The path followed for the realization of this study was the creation of the idea for the study and the relationship between the issues that were considered, the search for the most appropriate and freshest literature, the collection of data for this study, the raising of hypotheses, the processing of data and performing the regression analysis, extracting the results, verifying the main hypothesis, and from them reaching the relevant conclusions and recommendations. The main recommendation of this study is that investment in digitalization also affects investments in cyber security and both together help the development of countries and their integration between each other and beyond, while the relations between digitalization, cyber security and quality management are embodied at ISO standards, especially at ISO 38500, and ISO 27000 family of standards.

Keywords: digitalization, cyber security, regression analysis, ISO standards, quality management.

INTRODUCTION

In order to provide citizens and organizations with quick and simple services, digitalization entails substituting manual procedures with digital ones. In the digital age we live in today, customers are strongly linked to the usage of digital tools, which they use to read newspapers, interact with others, conduct transactions, purchase goods and services, purchase books, make financial decisions, and reserve hotels and holidays. Consequently, digitalization has had a profound impact on the economics of many nations worldwide, especially in certain economic sectors. Technology is evolving and everything will change with it. It is increasingly becoming an indisputable element in raising customer service standards and cutting operating expenses. Innovation is unavoidable and essential. These changes are occurring at an unstoppable rate due to new digital technology. Its revolutionary effect on business and the economy extends far beyond the potential for savings and optimization through the use of information technology advancements (Zërat. 2017). However, security measures must also be used to support digitalization processes, with cyber security being a crucial component.

The protection of computer systems and networks against equipment theft, computer program damage, electronic data theft, and misuse of the services they offer is known as computer security, cyber security, or information technology security (IT security) (Schatz, Daniel; Bashroush, Rabih; Wall, Julie 2017). The growing use of computer systems, the Internet, wireless network standards like Bluetooth and Wi-Fi, and "smart" devices like televisions, smart phones, and other gadgets that make up the "Internet of Things" are all contributing factors to the field's increased importance. One of the biggest issues facing the modern world is cyber security because of its complexity in terms of both politics and technology (Stevens, 2018).

Cybersecurity and digitalization have a significant positive impact on economic and sustainable development in every nation, region, and world at large. Both developed and developing nations invest in cyber security and related digital developments and technologies because failing to do so will cause a gulf between them and hinder communication and international relations, particularly in the areas of business, finance, and economics.

MATERIALS AND METHODS

1. Digitalization

The practice of employing digital tools by individuals, companies, and organizations to increase productivity, value, and speed of work is known as digitalization. For businesses in particular, digitalization is crucial since it may save time and open up new growth prospects.

*Corresponding author:

There are 5.35 billion people using the internet in 2024, equating to 66.2% of the world's total population. Internet users have grown by 1.8% over the past year, with 97 million new users coming online for the first time during 2023 (Kemp. 2024)). This makes the digital world the biggest it has ever been. Businesses are in the forefront of keeping up with the latest advancements in the digital world as we adapt to the technology practices and tools that make up daily life (UT. 2023).

Since technology advancements in the 1990s and 2000s, many organizations have implemented digitalization; yet, its potential applications and advantages are still untapped.

Digitalization is the process of transforming a business's internal operations and opening up new avenues for value and income generation. It entails converting analogue information that already exists into digital formats and a few procedures that businesses can carry out more effectively by utilizing the newest instruments and technologies.

Digitalization is a process that involves more than just utilizing computers to record data and going paperless. When fully implemented, digitalization encompasses a network of procedures that leverage digital data to employ technology to expedite activities like accounting, invoicing, and inventory control. Any company that wants to compete in the modern business world must adopt digitalization, especially if they want to cut costs and save time. Whether you're moving from paper to electronic invoicing or completely revamping your sales process for internet optimization, digitization can have a lot of advantages (UT. 2023).

1.1. Digital transformation

More extensive changes are involved in digital transformation than in digitalization. It suggests using digital technology into every aspect of a company's operations. To capitalize on digitization, digital transformation modifies business models, operational procedures, and corporate cultures. Employee remote work enabled by cloud-based solutions is an example of digital transformation.

1.2. How businesses use digitalization

Diverse company models employ various forms of digitalization. While some businesses may wish to go digital in order to improve customer service, others may choose to do so in order to update their inventory management system.

1.3. Digitalization of products and services

If you are a product or service seller, digitization can help you expand your business. For instance, it can facilitate the creation of digital supply chains with real-time data collection at each stage and facilitate the tracking and management of items (UT. 2023). Having data at your fingertips allows you to quickly spot patterns, problems, bottlenecks, and inefficiencies and find solutions. The following are some instances of digitized goods and services:

- Enabling consumers to pay for goods via an app;
- Using a central digital database to store and retrieve patient medical records;
- Adding a barcode or QR code to products.
- Putting your company at the forefront of your product or service online;
- using a point of sale (POS) device linked to a POS system;
- Digitizing internal processes (UT. 2023)

If the organization trades in real goods, there are several ways in which internal business operations can be digitalized. Inventory management may be exceedingly challenging, especially when done manually. Investing in a digital inventory management system facilitates inventory management and helps to streamline procedures. After that, the goods may be packaged for shipping, precise information on products that have been stored can be obtained, and damaged goods can be located. Issuing digital invoices rather than printing paper invoices helps streamline this process. Because digital invoices are paperless, they are easier to examine and process and are better for the environment. In addition, digital invoices are less prone to errors and require no physical storage space.

The ability of invoicing and inventory systems to interact directly, allowing inventory management to be completed without analogue interaction, is one of the many advantages of digitization with digital invoices. The work-in-process inventory counts are subtracted from the invoiced amount.

Lastly, accounting software helps streamline a company's financial reporting. Digitalization allows for the recording of transactions, the making of entries, and the completion of a final analysis with a few laptop touches.

1.4. Digitizing customer interactions

In a recent survey, 88% of participants said that they thought a company's customer experience was just as essential as its goods or services (SR. 2022). This serves as another evidence that firms should do everything within their power to guarantee that their relationships with clients are fulfilling and constructive. This may be accomplished at scale with the aid of digitalization, which optimizes every digital point of contact between a company and its clients. Software for customer relationship management, for instance, can assist in keeping track of each and every customer encounter. By having access to it, interactions can be tailored to offer individualized services. Additionally, CRMs gather and examine customer-related data that can be utilized to create more effective customer service plans (UT. 2023).

1.5. Digitalization of the supply chain

Managing supply chains by hand could be challenging. Because supply networks are so complex, interdependent, and variable, managing them can be difficult. Thankfully, the procedure has become considerably simpler due to recent advancements in supply chain management technologies. There are various ways to incorporate digital technology into the supply chain:

•The utilization of software for inventory management,
• Demand forecasting,
• Last-mile delivery,
• On-demand storage (UT. 2023)

1.6. The benefits of digitalization

When properly implemented, digitization can:
• Save time by minimizing or doing away with labor-intensive manual data input procedures.
• Enhances business preparedness by empowering you to foresee obstacles and devise strategies to overcome them.
• Enhances workflows through process automation and the reduction or elimination of inefficiencies caused by humans.
• Makes data-driven decisions easier by spotting trends and averting possible issues.
• Lowers errors through process automation that eliminates human error.
• Increases productivity by optimizing available resources.
• Lowers operating costs by lowering the number of workers needed for manual tasks.
• Boosts output by enhancing both team and individual worker productivity.
• By applying digital technologies and strategies around customer-facing procedures, enhance customer engagement and service.
• Enhances data analysis by streamlining the gathering and storing of data and by offering insights to direct business decision-making.
• Facilitates automation by fostering a manual intervention culture that is restricted to tiresome or repetitive dialogues.
• Facilitates quick decision-making by verifying new procedures and reviewing and refining current ones.
• Boost income through the development of sophisticated data-driven automation and efficiency-based sales and marketing systems (UT. 2023).

1.7. Disadvantages of digitalization

Among the possible obstacles to corporate digitalization are:
• Technical difficulties brought on by a lack of skilled workers and the requirement for sufficient labour proficiency in implementing new technologies.
• Decision-making that is stifled by risk aversion, ingrained legacy processes, and silencing. Employee and management resistance to implementing digitally led systems and processes is a symptom of cultural problems.
• Security and privacy issues pertaining to the cost of cyber security infrastructure, the possibility of malevolent cyberattacks, and regulatory data compliance (UT. 2023).

2. Cyber security

Cybersecurity is the discipline of defending programmes, networks, and systems from online threats. These cyberattacks typically target sensitive data access, alteration, or destruction; ransomware extortion of users' funds; or interference with regular corporate operations (Cisco, 2024). Today's increasingly sophisticated attackers and the fact that there are more devices than people make it particularly difficult to implement effective cybersecurity safeguards. Many levels of security are layered over computers, networks, programmes, and data that one wants to keep safe in a good cybersecurity strategy. To effectively defend against cyberattacks, an organization's people, processes, and technology must work in concert with one another.

People: Users need to be aware of and follow fundamental data security guidelines, like selecting secure passwords, being cautious when opening emails, and regularly backing up their data (Cisco, 2024).

Processes: Businesses need to have a plan in place for handling both successful and attempted cyberattacks. A reputable framework can serve as a reference and provide guidance on how to recognize attacks, safeguard systems, recognize and respond to risks, and recover from successful attacks.

Technology: In order to provide businesses and individuals with the computer security capabilities they need to defend themselves against cyberattacks, technology is needed. Endpoint devices, which include PCs, smart gadgets, and routers; networks; and clouds are the three primary entities that need to be secured. Next-generation firewalls, DNS filtering, antivirus software, malware protection, and email security solutions are among the common technologies utilized to safeguard these companies (Cisco. 2024).

2.1. The importance of cyber security

Everyone benefits from cutting edge cyber security solutions in today's connected environment. Individually, a cyber security breach may lead to identity theft, extortion attempts, or the loss of private information such as family photos. Everybody depends on vital infrastructure, including hospitals, power plants, and financial services firms. The smooth operation of our society depends on the security of these and other organizations (Cisco, 2024).

2.2. Some of the types of cyber security threats

The act of sending bogus emails that appear to be from reliable sources is known as phishing. The intention is to steal private information, including credit card numbers and login credentials. As per Cisco (2024), this is the most prevalent kind of cyberattacks:
1. An adversary may employ social engineering as a strategy to coerce you into disclosing private information. They might ask for money or access to private information. Any of the aforementioned dangers can be paired with social engineering to increase your likelihood of downloading malware, clicking on links, and believing unscrupulous sources.
2. Malicious software includes ransomware. By preventing access to files or the computer system until the ransom is paid, it is intended to extort money. Restoring the system or recovering the files is not assured by paying the ransom.
3. Malware is a category of software intended to damage or obtain unauthorized access to a computer (Cisco, 2024).

3. Quality management and Digitalization (ISO 38500:2024)

Most businesses depend on information technology (IT) to run successfully, both as a supporting function and as a component of their ability to transform themselves. In order to satisfy the demands and expectations of the organization's stakeholders, IT can significantly improve results and facilitate the development of new business models. The ISO 38500 standard, specifically designed for digitization, was created by the International Standards Organization (ISO 2024, a). In order for their organizations to fulfil their mission in a way that their stakeholders anticipate, governing bodies need direction on the responsible, innovative, sustainable, and strategic use of IT, data, and digital capabilities. To accomplish excellent governance, the governing body can use three instruments together with related governance and management practices of IT:
1. IT governance principles: Using these guidelines to ensure that IT is used responsibly and strategically can make an organization more flexible and adaptable.
2. IT Governance Model: This model outlines the primary IT governance duties and their relationships across the entire organization, facilitating decision-making and assigning of roles for all facets of IT use.
3. IT Governance Framework: This helps make sure that the important governance actions are taken into account and applied to the way the organization uses IT. It outlines the components that the organization uses to regulate its IT arrangements.

This document's target audience includes: — all organizations, ranging from small to large, irrespective of the degree of their use of IT; — public and private corporations, government agencies, and non-profit organizations (ISO. 2024, a). It also offers direction to people who advise, inform, or support governing entities.

Among them are the following:
• External business or technical specialists, such as legal or accounting specialists,
• Retail or industrial associations,
• Professional bodies;
• Internal and external service providers (including consultants);
• Auditors (ISO. 2024, b).

Although ISO 38500 was created to direct board oversight of IT, it contains a plethora of helpful guidelines for digital leadership and change (Toomey, 2017).
Digital transformation is a long-term, ongoing process that disrupts and changes society, markets, business, and government rapidly. It is made possible by the creative application of new digital technologies. It upends many preconceived notions, alters established power structures, and generates new value propositions and value perceptions. (Toomey, 2017).
Digital leadership is the ability of corporate executives to recognise and seize opportunities for revolutionary business growth and value through the transformational use of IT in a way that is acceptable, efficient, and successful. For most businesses, a successful digital transformation will include large adjustments to people (abilities, culture, responsibilities), processes, technology, and organization (structure, market). Without the full executive team and the Board working together, success will be elusive. When applied to business strategy and the means of delivering it, the Evaluate, Direct, and Monitor focus areas guarantee that the Board is appropriately involved in navigating the business strategy through times of disruption; choosing the best technology to realise corporate strategy; ensuring that the rest of the organization is aware of the path being taken; and ensuring progress towards the goal (Toomey. 2017).

4. Quality management and cyber security (ISO 27001 Family)

In the current market, businesses want to give their clients trust and show that they are dedicated to protecting the security of the data they handle. To do this, an organization can gain a competitive edge by obtaining certification of an ISO standard or security regulation, which attests to the proper management of security needs in information processing operations.

4.1. Cybersecurity and ISO standards

These days, cybersecurity is truly taking off, but why? It is undeniable that controls are necessary to ensure the security of devices, communication networks, and information assets given the rising number of security events and attacks involving information and IT systems that businesses are facing. The idea of cybersecurity is born out of this requirement. These kinds of assaults seek to gain access to, alter, or delete private information that belongs to organizations (Martin, 2022).

Effective cybersecurity measures are difficult to deploy because cybercriminals are constantly coming up with new ways to carry out their assaults because to the vast array of tools and technology they use. Nonetheless, there is a method for putting data and information security measures into place that partly simplifies and normalizes the process of putting these IT security measures into place.

These are the information security and cybersecurity-related ISO standards and rules. The International Standards Organization creates and disseminates ISO standards (ISO). The International Electrotechnical Commission (IEC) and ISO are the world's foremost authorities on standardizing. They create worldwide standards to standardize particular procedures in fields like information security through technical committees made up of ISO and IEC member organizations (Martin, 2022).

These standards now form an essential part of businesses' compliance systems, giving them respect and recognition across the globe. Organizations that implement ISO standards benefit from a distinct advantage over their competitors. This is because the certified standards undergo periodic reviews and audits to ensure compliance, thereby enhancing the organization's reputation with stakeholders like shareholders and clients. ISO standards are categorized into families and given sequential numbers that correspond to the areas they cover. This allows standards related to comparable problems to be grouped together. These rules and standards seek to define methods, guidelines, policies, skill development, etc. in relation to the domains they cover (security, continuity, and quality, etc.).

4.2. ISO 27000 Family

ISO 27000 is a set of ISO standards. To manage information security within an organization, this set of information security standards (Martin, 2022). It lays out requirements and principles for putting an Information Security Management System (ISMS) in place. ISO 27001 is the primary standard in this set and serves as the series benchmark. Requirements for the creation, execution, upkeep, and ongoing enhancement of an Information Security Management System are outlined in this standard. Based on the Deming Cycle, often known as PDCA (Plan-Do-Check-Act), is the process of continuous improvement. The remaining family standards aid in the implementation of the ISMS by providing guidance. PDCA, often known as the Deming Cycle, is the foundation of the continuous improvement process (Plan-Do-Check-Act). The other standards in the family act as a roadmap and assistance for putting the ISMS into practice. An additional notable standard is ISO 27001/02. This is a best practice guide that outlines the necessary controls and control objectives for information security (Martin, 2022)

ISO 27031 is a member of the same family, however it has a different objective. This non-certifiable standard acts as a roadmap and offers a range of techniques and protocols for determining elements that enhance an organization's ICT readiness to ensure and strengthen business continuity. In other words, the primary goal of this standard is to guarantee service continuity and the ability of the organization to resume operations in the event of a disaster and return to a pre-agreed operational condition.

From the ISO 27000 family, ISO 27701 is comparable to the standard mentioned above. Additionally, it lays out guidelines for handling, safeguarding, and monitoring the privacy of the company's personal data in accordance with laws and rules like the General Data Protection Regulation (GDPR). It contains recommendations for safeguarding the privacy and confidentiality of personal data handled by an organization, based on the rules, controls, and goals of the ISO 27001 security standard. It should be noted that only with ISO 27001 certification can this standard be met (Martin, 2022).

I an article on the relations between cyber security and ISO 27001 author verifies by a regression analysis the strong relations between two characters (Ceko. 2023).

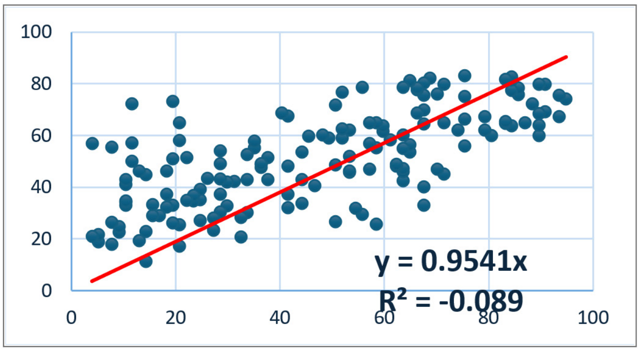
5. Methodology and methods

The process taken to carry out the research project, including coming up with the idea for the study and figuring out how the issues under consideration related to each other, finding the most recent and appropriate literature, gathering data for the study, formulating hypotheses, processing the data and running regression analysis (done by excel), extracting the findings, confirming the main hypothesis, and drawing pertinent conclusions and recommendations from them. The study's principal recommendation is that investments in digitalization have an impact on cyber security investments, and that both positively contribute to national development and intergovernmental integration.

Hypothesis 0: There is no strong relationship between the Level of Digital Development and the Cyber Security Index.

Hypothesis 1: There are strong relationships between the Level of Digital Development (DDL) and the National Cyber Security Index (NCSI).

Graph 1. *Correlation between the Level of digital development and the Cyber Security Index.*



Source: Author of this study.

Tables 1, 2 and 3. *Regression analysis between Level of Digital Development and Cyber Security Index.*

SUMMARY OUTPUT (Table 2)	
Regression Statistics	
Multiple R	0.937599
R Square	0.879092
Adjusted R Square	0.872556
Standard Error	18.69788
Observations	154

ANOVA	(Table 3)				
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	388916.2	388916.2	1112.426	8.2E-72
Residual	153	53490.45	349.6108		
Total	154	442406.7			

(Table 4)	*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
DDL	0.921394	0.027625	33.35305	4.41E-72	0.866817	0.975971	0.866817	0.975971

Source: Author of this study.

RESULTS

• Through regression analysis, it is proven that there are strong relationships between the Level of Digital Development and the Cyber Security Index, since $R^2 = 0.879092 > 0.50$.
• The results show that Digital Development Level at 87.9092% have the explanation of indication of Cyber Secyrity Index and vice a versa.
• Relations between DDL and CSI are strong (r = 0.937599).
• Regression equation y = 0.9541x
• $R^2 = 0.879092$
• Correlation coefficient "r" = 0.937599.
• Hypothesis H1 – There is a strong relationship between DDL and CSI has been verified.
• By ANOVA Fllog> Fcrit, F Significance F (probability getting these results) < α = 0,05.
• H1 has been verified, with a significance level of 0.05 or a level of reliability = 95 %.
• The H1 has been verified with a significance level 0.1 or a level of reliability = 90% too, and delivers the same result of $R^2 = 0.879092$.
• Coefficients are the values of the correlation coefficient.

CONCLUSIONS AND RECOMMENDATIONS

1. Digitalization is replacing manual processes with digital ones so that citizens and organizations can receive quick and easy services. In the current digital era, consumers are closely associated with the use of digital technologies, which they employ to read newspapers, communicate with others, carry out transactions, buy products and services, buy books, make financial decisions, and book travel and accommodations.
2. As a result (of 1), digitization has significantly impacted the economies of numerous countries throughout the world, particularly in specific economic sectors. Everything is changing in tandem with the evolution of technology. It is undeniably becoming a necessary component for improving customer service standards and reducing operating costs. Innovation is necessary and inevitable. The rapid pace of these changes is attributed to modern digital technology. Beyond the potential for savings and optimization through the application of information technology breakthroughs, it has a revolutionary effect on business and the economy.
3. In every country, every region, and every part of the world, cybersecurity and digitalization have a major positive influence on sustainable and economic development. Investments in

cyber security and related digital innovations and technologies are made by both developed and developing countries since not doing so will widen the gap between them and impede international connections and communication, especially in the fields of commerce, finance, and economics.

4. The process of digitalization involves transforming pre-existing analogue data into digital formats and a few operations that companies can perform more efficiently by employing the newest tools and technologies. It is a process that goes beyond using computers to store information and eliminating printed documents, and when completely executed, entails a web of processes that use digital data to make use of technology in order to accelerate tasks such as inventory control, invoicing, and accounting. Any company that wants to compete in the modern business world must adopt digitalization, especially if they want to cut costs and save time.

5. Real goods trading companies can invest in a digital inventory management system, which makes inventory management easier and streamlines processes. After that, products can be located and packaged for shipping with accurate information on stored goods. This process can also be streamlined by issuing digital invoices instead of paper ones, which are easier to examine and process, better for the environment, less error-prone, and require no physical storage space. While another benefit of digitization with digital invoices is the direct interface between inventory and invoicing systems, which makes it possible to complete inventory management without analogue interaction. The invoiced amount is deducted from the work-in-process inventory counts, while accounting software facilitates the financial reporting of a business, making clear transaction recording, entry, and final analysis can all be completed thanks to digitalization. Customer relationship management software can help maintain a record of every interaction with customers. Access to it allows interactions to be customized to provide personalized offerings, and also collect and analyze client-related data that can be used to develop customer care strategies that are more successful. Handling supply networks by yourself could be difficult. Managing supply networks can be challenging due to their complexity, interdependence, and variability. Thank goodness, new developments in supply chain management systems have made the process much easier. Software for last-mile delivery, demand forecasting, inventory management, on-demand storage, and other related tasks can be used to do this.

6. The innovative use of new digital technologies can enable digital transformation, a protracted, continuous process that quickly upends and modifies markets, industry, government, and society. Numerous preexisting beliefs are turned on their head, existing power structures are changed, and new value propositions and value perceptions are created.

7. The ability of corporate executives to identify and grasp chances for transformative business growth and value through the innovative use of IT in a way that is successful, efficient, and acceptable is known as digital leadership. For the majority of firms, significant changes to people (abilities, culture, responsibilities), processes, technology, and organization (structure, market) are necessary for a successful digital transformation. Success will elude you if the Board and the entire executive staff don't collaborate. The Evaluate, Direct, and Monitor focus areas ensure that the Board is appropriately involved in navigating the business strategy through times of disruption, selecting the best technology to realise corporate strategy, keeping the rest of the organization informed of the path being taken, and ensuring progress towards the goal when it comes to business strategy and the means of delivering it.

8. The majority of organizations rely on information technology (IT) to perform properly, both as a necessary component of their capacity to change and as a supporting role. IT may greatly enhance performance and ease the creation of new business models in order to meet the needs and expectations of the organization's stakeholders. The International Standards Organization developed the ISO 38500 standard especially for digitalization. Governing bodies want guidance on the responsible, innovative, sustainable, and strategic use of IT, data, and digital capabilities so that their organizations may fulfil their mission in a way that their stakeholders anticipate.

9. The discipline of protecting systems, networks, and programmes from internet attacks is known as cybersecurity. These hacks usually aim to interfere with ordinary business operations, extract money from users via ransomware, or access, change, or destroy sensitive data. It is especially challenging to put in place efficient cybersecurity measures because of today's more skilled attackers and the fact that there are more devices than people. A strong cybersecurity plan layers many layers of security over computers, networks, programs, and data that one wishes to keep safe. An organization's people, processes, and technology must cooperate in order to effectively protect against cyberattacks. As a result of these standards, businesses are now respected and recognized globally and are an integral element of their compliance processes. Businesses that use ISO standards have a clear competitive edge over others. This is due to the fact that the organization's reputation with stakeholders, such as shareholders and clients, is improved when the certified standards are subjected to recurring inspections and audits to guarantee compliance. The families of ISO standards are assigned sequential numbers that match the

domains they cover. This makes it possible to group standards pertaining to similar situations together. These regulations and guidelines aim to provide procedures, guidelines, policies, training programs, and other things concerning the areas they address (security, continuity, quality, etc.).
10. Digitalization is the practice of substituting digital procedures for manual ones in order to offer citizens and organizations quick and simple services.
11. Security considerations must be included in digitalization processes, with cyber security being a key consideration.
12. The protection of computer systems and networks against equipment theft, damage, or misuse, as well as against malfunctions or abuse of the services they offer, is known as computer security, cyber security, or information technology security (IT security).
13. Users need to be aware of and follow fundamental data security guidelines, like selecting secure passwords, using caution when opening email attachments, and regularly backing up their data.
14. Businesses ought to have a plan in place for handling both attempted and successful cyberattacks. A reputable framework can serve as a reference and provide guidance on how to recognize attacks, safeguard systems, recognize and respond to risks, and recover from successful attacks.
15. Technology is necessary to provide people and organizations with the computer security tools they need to fend off cyberattacks. Endpoint devices, which include PCs, smart gadgets, and routers; networks; and clouds are the three primary entities that need to be secured. Next-generation firewalls, DNS filtering, malware protection, antivirus software, and email security solutions are examples of common technologies used to safeguard these institutions.
16. Everyone benefits from cutting-edge cyber security solutions in today's linked society. Individually, a cyber security breach may lead to identity theft, extortion attempts, or the loss of private information such as family photos. Critical infrastructure is a necessity for everyone, and the smooth operation of our society depends on the security of these and other institutions.
17. There is a strong relationship between the level of digital development and the cyber security index verified by means of regression analysis where $R^2 = 0.879092 > 0.50$
18. One of the study's suggestions is that investments in digitalization have an impact on cyber security investments, and that both positively contribute to national development and intergovernmental integration.

Conflict of interests

The author declares no any conflict of interests.

References

1. Carlos Martín. 2022. ISO standards and regulations for improving cybersecurity. https://www.globalsuitesolutions.com/iso-standards-and-regulations-for-improving-cybersecurity/. Visited 20th April 2024.
2. Ceko Enriko. 2023. On relations between Cyber Security Index and ISO 27001 Standard Index in Western Balkan countries. CIT Review Nov issue 2023. https://crj.cit.edu.al/wp-content/uploads/2014/10/crj-nov-issue-2.pdf. Visited 4th May 2024.
3. Cisco. 2024. What Is Cybersecurity? https://www.cisco.com/c/en/us/products/security/what-is-cybersecurity.html. Visited 20th April 2024.
4. International Standards Organization. 2024, a. ISO/IEC 38500 Third edition 2024-02 (en), Information technology — Governance of IT for the organization Technologies de l'information — Gouvernance des technologies de l'information pour l'entreprise, https://cdn.standards.iteh.ai/samples/81684/61bc5f7eee154e2cad5dbdb72f95d208/ISO-IEC-38500-2024.pdf. Visited 20th April 2024.
5. International Standards Organization. 2024, b. ISO/IEC 38500:2024. Information technology. Governance of IT for the organization. https://www.iso.org/standard/81684.html. Visited 20th April 2024.
6. Kemp Simon, 2024. Internet Use in 2024. https://datareportal.com/reports/digital-2024-deep-dive-the-state-of-internet-adoption#:~:text=There%20are%205.35%20billion%20people,the%20first%20time%20during%202023. Visited 4th May 2024.
7. Mark Toomey. 2017. Leading the RoboNDIS campaign. Seeking change and justice for people harmed by bad behaviour in the NDIS. https://www.linkedin.com/pulse/iso-38500-standard-guide-digital-transformation-mark-toomeyISO 38500 - a Standard to Guide Digital Transformation. Visited 20th April 2024.
8. NCSI Report. 2023. https://ncsi.ega.ee/ncsi-index/. Visited 20th April 2024.
9. Salesforce Report. 2022. Nearly 90% Of Buyers Say Experience a Company Provides Matters as Much as Products or Services. https://www.salesforce.com/uk/news/stories/customer-engagement-research/Visited 4th May 2024.
10. Schatz, Daniel; Bashroush, Rabih; Wall, Julie (2017). "Towards a More Representative Definition of Cyber Security". Journal of Digital Forensics, Security and Law (në anglisht). 12 (2). ISSN 1558-7215.
11. Stevens, Tim (2018-06-11). "Global Cybersecurity: New Directions in Theory and Methods" (PDF). Politics and Governance. 6 (2): 1–4. doi:10.17645/pag.v6i2.1569.
12. The Upwork Team. Mar 3, 2023. What Is Digitalization in Business? Basics and Best Practices. https://www.upwork.com/resources/digitalization-in-business. Visited 20th April 2024.
13. Zërat. 2017. Ç'farë është digjitalizimi. https://zeri.info/zerat/127535/cfare-eshte-digjitalizimi/. Visited 20th April 2024.

APPENDIXES

Tables 4. *Ranking of countries according to the Level of Digital Development and the Cyber Security Index.*

Rank	Country	NCSI	DDL
1	Belgium	94.81	74.07
2	Lithuania	93.51	67.34
3	Estonia	93.51	75.59
4	Czech Rep	90.91	69.21
5	Germany	90.91	80.01
6	Romania	89.61	59.84
7	Greece	89.61	64.02
8	Portugal	89.61	68.46
9	UK	89.61	79.96
10	Spain	88.31	72.21
11	Poland	87.01	65.03
12	Austria	85.71	75.76
13	Finland	85.71	78.35
14	S. Arabia	84.42	63.89
15	France	84.42	77.29
16	Sweden	84.42	81.51
17	Denmark	84.42	82.68
18	Croatia	83.12	64.63
19	Slovakia	83.12	65.44
20	Netherlands	83.12	81.86
21	Serbia	80.52	59.81
22	Malaysia	79.22	62.19
23	Italy	79.22	67.26
24	Ukraine	75.32	55.96
25	Latvia	75.32	66.23
26	Ireland	75.32	75.18
27	Switzerland	75.32	82.93
28	Bulgaria	74.03	62.06
29	DominicRep	71.43	45.21
30	Russia	71.43	65.12
31	Singapore	71.43	79.93
32	Morocco	70.13	46.88
33	Canada	70.13	75.96
34	Korea Rep	68.83	82.23
35	Bangladesh	67.53	33.11
36	India	67.53	40.02
37	Hungary	67.53	64.25
38	Slovenia	67.53	69.74
39	Israel	67.53	75.50
40	Norway	67.53	80.19
41	Cyprus	66.23	68.83
42	Australia	66.23	77.61
43	Luxembourg	66.23	78.40
44	Georgia	64.94	53.50
45	Thailand	64.94	56.63
46	USA	64.94	81.05
47	Paraguay	63.64	42.58
48	Philippines	63.64	45.99
49	Indonesia	63.64	47.41
50	Azerbaijan	63.64	54.78
51	Argentina	63.64	60.43
52	Japan	63.64	78.69
53	Peru	62.34	48.23
54	Albania	62.34	48.74
55	Türkiye	61.04	58.29
56	Chile	59.74	61.44
57	Uruguay	59.74	63.86
58	Benin	58.44	25.83
59	NRMacedonia	58.44	55.36
60	Qatar	58.44	64.99
61	Egypt	57.14	46.93
62	Moldova	57.14	56.79
63	Bahrain	57.14	65.17
64	Zambia	55.84	29.66
65	Iceland	55.84	78.64
66	Nigeria	54.55	31.76
67	Ecuador	53.25	45.57
68	Tunisia	53.25	46.26
69	Colombia	53.25	52.08
70	Belarus	53.25	62.33
71	Brazil	51.95	59.11
72	China	51.95	62.41
73	New Zealand	51.95	76.81
74	Uganda	50.65	26.71
75	Panama	50.65	48.43
76	Malta	50.65	71.74
77	Costa Rica	49.35	58.87

78	Kazakhstan	48.05	60.18
79	Ghana	46.75	40.68
81	Oman	45.45	59.51
82	Côte d'Ivoire	44.16	33.54
83	Sri Lanka	44.16	43.02
84	Mauritius	44.16	53.57
85	Pakistan	41.56	32.23
86	Kenya	41.56	37.14
87	Jamaica	41.56	48.18
88	Brunei	41.56	67.5
89	UAE	40.26	68.87
91	Kyrgyzstan	37.66	42.96
92	Mexico	37.66	51.46
93	Vietnam	36.36	47.69
94	Uzbekistan	36.36	49.00
95	S. Africa	36.36	49.24
96	Armenia	35.06	55.06
97	Montenegro	35.06	57.79
99	Rwanda	33.77	30.23
100	Algeria	33.77	42.81
101	Trnd&Tbg	33.77	52.6
103	Ethiopia	32.47	20.70
104	Cameroon	32.47	28.28
105	Bolivia	31.17	42.09
107	Nicaragua	29.87	32.70
108	Botswana	29.87	41.96
109	Nepal	28.57	30.58
110	Namibia	28.57	37.28
111	Venezuela	28.57	43.14
112	B&H	28.57	49.31
113	Jordan	28.57	54.07
114	Malawi	27.27	23.20
115	Vanuatu	27.27	28.10
116	Tonga	25.97	43.40
117	Tanzania	24.68	26.96
118	Guatemala	24.68	35.43
119	El Salvador	24.68	39.17
120	Cambodia	23.38	34.59
121	Bhutan	23.38	36.90

122	Honduras	22.08	35.09
123	Suriname	22.08	51.50
126	Chad	20.78	17.28
127	Sudan	20.78	25.50
128	Grenada	20.78	58.00
129	Bahamas	20.78	65.10
131	Mali	19.48	26.00
132	Senegal	19.48	33.04
133	Iran	19.48	51.04
134	Barbados	19.48	73.10
135	Lao PDR	18.18	32.37
136	Belize	18.18	37.10
137	Mongolia	18.18	46.41
139	Cuba	16.88	29.10
140	Zimbabwe	15.58	28.97
141	Syria	15.58	33.40
142	Mauritania	14.29	11.30
143	Madagascar	14.29	22.80
144	Fiji	14.29	44.90
145	Afghanistan	12.99	19.50
146	Saint Lucia	12.99	46.30
147	Seychelles	11.69	50.30
148	Antg&Brbd	11.69	57.10
149	SK&Nevis	11.69	72.40
151	Samoa	10.39	33.00
152	Myanmar	10.39	34.29
153	Tajikistan	10.39	34.56
154	Libya	10.39	41.10
155	Guyana	10.39	42.91
156	Angola	9.09	22.69
157	Mozambique	9.09	24.88
158	Yemen	7.79	18.00
160	Haiti	7.79	26.46
161	SV&Grnd	7.79	55.40
166	Congo	5.19	18.91
167	Kiribati	5.19	21.70
168	Slmn Isl	3.90	21.10
169	Dominica	3.90	56.90

Source: NCSY Report. 2023.

THE TECHNOLOGICAL REVOLUTION'S IMPACT ON BUSINESS MANAGEMENT

***Jaroslav Kollmann* [1]*, Daniel Chamrada* [2]*, Pavel Cepák* [3]**

[1] Department of Economics, The Faculty of Operation and Economics of Transport and Communications, University of Žilina, Slovakia, kollmann@vste.cz, ORCID: 0000-0003-4110-145X

[2] Department of Economics, The Faculty of Operation and Economics of Transport and Communications, University of Žilina, Slovakia, chamrada@vste.cz, ORCID: 0000-0002-3934-7840

[3] The Faculty of Corporate Strategy, Department of Management, Institute of Technology and Businesses in České Budějovice, Czech Republic, cepak@vste.cz

Abstract

The rapid advancement of technology in the era of the fourth industrial revolution has significantly transformed organizational management practices. This study examines the influence of technological progress on organizational management, specifically analyzing the effects of data analysis and digital management methods in response to these changes. The main objective is to determine how these technological advancements impact enterprises of various sizes in the Czech Republic. The research employed comparative and analytical approaches, utilizing survey data collected from numerous organizations. Concrete methods used in the study include a comprehensive questionnaire survey, comparative analysis, and statistical correlation analysis to assess the significance of technologies based on the classification of enterprise size. The findings demonstrate substantial disparities in the implementation and advantages of technology contingent upon the magnitude of the organization, with smaller and larger enterprises exhibiting enhanced flexibility and deliberate assimilation of novel technologies. The main conclusion is that proficient administration of technology results in competitive advantages and operational efficiencies. Companies are advised to customize their technology strategies based on their available resources and market requirements. Further analysis explores the connections between investments in technology and other organizational elements such as financial resources and personnel, highlighting the importance of developing coherent plans that successfully incorporate technology into the overall structure of the organization.

Keywords: Technology Integration, Industry 4.0, Organizational strategy, Business Competitiveness

INTRODUCTION

Organizations have undergone tremendous changes in the last few decades, mainly due to the constant advances in technology and digitalization. Organizations want to build a competitive advantage in a highly competitive environment, but it is important to note that some are stumbling because they are not keeping up with the current trends that are necessary for a business to thrive today. The technological revolution has not only caused fundamental changes in the way businesses operate but also in their management strategies and operations.

Data analytics and digital management have become two of the main pillars of this revolution. This is the fourth industrial revolution, which is specific to its rapid development and has the potential for positive changes in the organizational and management framework of the enterprise. Unfortunately, however, there is a lack of a unified perception and approach to the implementation of Industry 4.0, and organizations face technological, operational, and organizational challenges and are forced to cope with many uncertainties (Suleiman et al., 2022). The technological revolution is also impacting the human resources in a company and their wages, with Lee and Clarke (2019) stating that earlier in the 2nd and 3rd industrial revolutions, people were concerned about mechanization reducing employment levels, but it has been shown that the overall gains from technological advancements are not only benefiting organizations but also the workforce as their wages are increasing, making some occupations less profitable.

Given the ever-evolving nature of the industry and rapid technological innovation (Min, 2022), it is essential that companies use data analytics that consider the effects of integrating different technologies. This approach is essential for the successful implementation of innovative research, development, and management strategies (Kim et al., 2023). However, according to Wambsganss et al. (2023), the introduction of new technologies poses

*Corresponding author:

significant challenges for enterprises in terms of their technology strategy and new business models (Van Zeebroeck, Kretschmer, and Bughin, 2023), which is because enterprises must deal with the unfamiliarity of previously distinct technologies. However, according to Huamanchumo (2021), technological innovation facilitates strategy formulation. If an organization wants to get ahead of the competition and gain a competitive advantage (Hernandez and Salcedo, 2020), it needs to use technology platforms that require unique management strategies that many technology platform owners do not address (Venter and Grobbelaar, 2023). According to Fosfuri, Giarratana, and Sebrek (2020), organizations that focus on a broad product strategy and provide tailored services to their customers and are as flexible as possible thrive more (Rosa et al., 2022). Artificial intelligence, or whether it has positive or negative effects on businesses, is a much-discussed topic nowadays. Rosales et al. (2020) observe that a wide range of industries are implementing AI technologies, and with proper talent training, AI can create more jobs and opportunities. Artificial intelligence has a significant role to play, especially because of big data, intelligent manufacturing, cyber-physical systems, and real-time scheduling algorithms (Zeba et al., 2021). Organizations are using AI, analyzing big data, and extending digital government because they are more adaptable, efficient, and responsive to customer wants and needs (Newman, Mintrom, and O'Neil, 2022). As technology continues to evolve, educating employees within the enterprise is important to positively influence the implementation of new technology and reduce the risk of implementing it (Jin et al., 2023). Rahmani et al. (2022) assert that e-learning greatly enhances employee education by providing a variety of methods to improve employee knowledge, skills, and attitudes using the latest technology. The popularity of e-learning continues to grow due to its ability to provide learning opportunities, a necessary way to prepare people for data-driven and connected industrial systems (Gowripeddi et al., 2021), anywhere and anytime. Identifying and addressing the factors that can affect its success while taking full advantage of its potential benefits and avoiding potential drawbacks is essential to ensuring the successful implementation of e-learning. According to Ozturkler (2021), due to rapid technological advancement, training activities are very important to maintain competitive advantage and a good position in the market environment. Gajek et al. (2022) also confirm this fact, arguing that employees must receive technology education, particularly in the area of process safety. Based on research, Scepanovic (2019) concludes that the scale and complexity of the technological revolution are unlike anything humanity has experienced so far, and this technological revolution represents the way people communicate, live, work, and even think. According to Quinn (2019), the technological revolution is also characterized by price reversals. Employee adoption of mobile technology is necessary for competency development, especially for knowledge transfer (Ku-Ciapski, 2019). The mere presence of remarkable technological advancements cannot guarantee social and economic progress, but it is necessary to spend resources on modern communication and information technologies to continuously improve quality and ensure sustainable development (Kalenyuk, Bo-hun, and Djakona, 2023). According to Russo (2020), almost half of the companies surveyed want to invest in technology to remain competitive in the market and be able to exploit new opportunities. It is essential for governments to create a strong institutional structure, especially through initiatives such as education, welfare, and training programs, which are essential to exploiting the new technical opportunities that innovation brings with it (Focacci and Perez, 2022).

Automation technologies, such as sensors, robots, and programmable devices interconnected with the internet, characterize Industry 4.0 (Giustozzi, Saunier, and Zanni-Merk, 2018). Large enterprises implement automation technologies at a higher level than medium and small enterprises, which makes sense given the larger pool of skilled workers (Pap-ulova, Gazova, and Sufliarsky, 2022). According to Garcia-Loro et al. (2021), Industry 4.0 has affected almost all layers of society and has also changed the way we interact with the world, and the world in turn interacts with us. According to Mesárová, Kordos, and Sokol (2019), Industry 4.0 positively affects the economy and society in general, but many organizations that are transitioning to Industry 4.0 take little account of the human factor (Angelopoulou, Mykoniatis, and Boyapati, 2020), which is the biggest risk of the industry 4.0 implementation process (Masár and Hudáková, 2020). The industry 4.0 movement now permeates the entire production system, from the labor to the social, economic, political, and legal dimensions of life (Fuster, 2019) and shows great potential (Li and Huang, 2021). These developments include the comprehensive digitization of industrial processes and products through the application of cyber-physical systems and self-motivation, both in production facilities and in logistics and operational processes (Figueiras et al., 2021). Jaskov (2020) asserts that Industry 4.0 enhances performance and productivity, with potential impacts on business downturns at both macro and micro levels. Nowadays, many manufacturers and organizations are transitioning to Industry 4.0 with great success, which also has a huge impact on the dynamics of the labor market (Su et al., 2022). They have recognized the benefits of digitized production and have started to implement highly mechanized and data-driven processes that allow them to offer superior services and products to their customers (Angelopoulou, Mykoniatis, and Boyapati, 2020). The global industry is currently undergoing significant transformations due to Industry 4.0. This revolutionary concept is reshaping not only business practices but society. Every organization is feeling the impact of physical, virtual, and combined technologies, which are becoming increasingly important in all aspects of production management. The primary goal of incorporating Industry 4.0 into production management is to modernize, automate, and digitize the manufacturing process. The aim is to achieve more efficient and productive production while maintaining lower costs (Richnák, 2021; Elafri et al., 2022). Ameywal et al. (2022) use technologies like artificial intelligence, the internet of things, machine learning, cloud computing, big data, digitalization, deep learning, and cyber-physical systems to manage the production flow, thereby minimizing production time and ensuring on-time delivery. According to Taner and Parlak Biçer (2021), the use of technology and project management ensures the successful operation of the business and contributes to the organisation's development. Therefore, it is also necessary to design the right communication technology to use all technological management capabilities and prevent problems (Reyez et al. 2020).

The purpose of this paper is to investigate and determine the impact of technology in organizations, as well as to provide enterprises with information about technology.

Organizations must react and adapt to the wide range of factors that influence today's world. These factors can, of course, include technological equipment and general access to technology. According to Gong et al. (2021), technology plays a significant role in the process of creating products and services. It is critical for organizations to understand the future trend of technology development to improve and innovate the products in their portfolio.

The subsequent matter under scrutiny expands upon the preceding one and explores the individual associations between the components being investigated inside organizations, as well as the interconnections between them. Li, An, and Liu (2021) emphasize the significance of identifying essential characteristics, particularly in the context of planning techniques that enhance internal processes inside a company. Therefore, institutions and organizations can attain superior outcomes, enhance their competitiveness in the market, and accomplish their objectives.

MATERIALS AND METHODS

The objective of this study is to evaluate the influence of technical progress on organizational management and examine how these effects differ according on the size of the firm. Furthermore, it examines the interconnectedness of different organizational aspects that are impacted by technology.

The theoretical part of the study employed data from research articles to ascertain the importance of technology in firms and investigate methods for efficiently overseeing operations and relationships within the framework of the fourth industrial revolution. The text explores how firms are now adopting and implementing technological trends in their corporate processes, while also examining the possibility for future success.

Based on the literature review, the following hypotheses were formulated:

H1: There is a relationship between technology and firm size categorization.

H2: There is an interrelationship between factors examined in organizations.

In 2023, the practical component involved the examination and collection of data from companies located in the Czech Republic. The study chose a varied sample of enterprises that operate in the market environment of the Czech Republic. This sample includes organizations of various sizes and from different industries. Almost all businesses strive to integrate and utilize technology in their production processes to consistently enhance their performance and outcomes.

The present study entailed the selection of data from many firms functioning within the market environment of the Czech Republic. The sample was heterogeneous, including businesses of various magnitudes and spanning across multiple sectors. Almost all businesses endeavor to include and employ technology in their production processes to continually enhance their performance and results. A questionnaire survey was developed internally to gather dat. The questionnaire survey comprised eight questions focusing on the theoretical understanding of technology and the technological revolution in relation to market-operating enterprises.

To address the research questions, several analytical techniques were employed:

Conducted to identify and compare technology adoption levels and benefits among micro, small, medium, and large enterprises. Analysis of Variance (ANOVA) was used to test for significant differences.

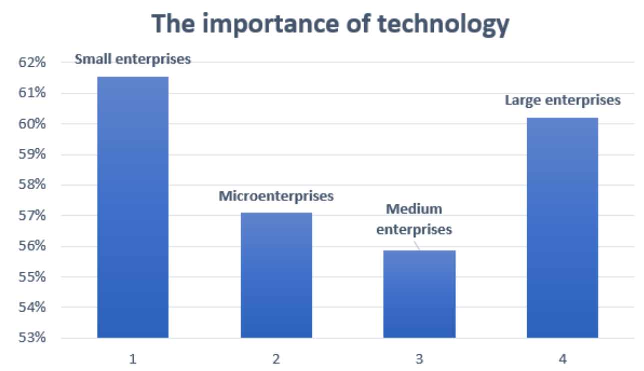
Correlation Analysis: Utilized to explore the relationships between technology adoption and other organizational factors such as capital, human resources, and investment. Pearson correlation coefficients (r) were calculated to determine the strength and direction of these relationships.

RESULTS

As previously mentioned in the literature study, technology has a significant impact on nearly every industry and is present worldwide. In the Czech Republic, the situation is identical. The Czech Republic, despite its modest size, recognizes the importance of digitalization and technical breakthroughs for enterprises of all sizes. These advancements not only benefit individual businesses but also have a substantial impact on the country's economy. By studying and elucidating the findings, this knowledge can be invaluable for all firms aiming to adopt or currently using diverse technologies to improve their position in an already fiercely competitive environment.

The first hypothesis examined the significance of the technologies utilized by each of the entities under study, considering their respective sizes, which encompass micro, small, medium, and large enterprises. The findings were analyzed using the averaging approach, which helped get a fundamental knowledge of how various firms interpret and use technology in their distinct surroundings. The average metrics for the significance of microenvironmental components offer varying viewpoints on how these institutions interact with technological advancements.

Graph 1. *Importance of technology by size of enterprise*



Source: Own

According to the conducted research, small enterprises demonstrate a significant level of openness to technological advancements, as indicated by a score of 0.62. This receptiveness allows them to be more adaptable in responding to market fluctuations and efficiently utilizing new technical tools. This technological capability offers a benefit in terms of both flexibility and innovation. However, it is important to mention that micro enterprises have a value of 0.57%, which is somewhat less than that of small enterprises. Nevertheless, it is not necessarily the case that they have failed to keep up with technical advancements. On the contrary, their attitude may be characterized as cautious and conventional. Me-dium enterprises have a technological relevance score of 0.56 (56%), which is the lowest among all enterprises. Despite the average degree of technology utilization in this sector, there is still significant opportunity for growth. Both large and small enterprises have a technology relevance score of 0.60 (60%), indicating a strong emphasis on investing in technology and creative techniques to maintain competi-tiveness in their respective sectors.

In summary, the technology perspectives offer valuable insights into various strategies and technologies that can significantly influence the future competitiveness and performance of enterprises operating in specific markets and industries.

Correlation analysis

The data underwent analysis using a correlation study to investigate the links among the four components. An inquiry was carried out to examine the interdependence among the elements. The relationships between the factors were clarified, and after analyzing the correlations, clear results and recommendations were given for the dependencies that were assessed.

Table 1: *Correlation analysis*

	Capital	Technology	Human Resources	Investment
Capital	1			
Technology	0,091132	1		
Human Resources	-0,010790	0,18228044	1	
Investice	0,416119	0,22886291	0,102919439	1

Source: Own

Technology-Capital (0.0911)
The correlation between technology and capital is positive, but it is characterized by a low level of strength. From this, it may be inferred that businesses with greater wealth are more inclined to invest in technological equipment, although the correlation is not significant. However, it is not an absolute belief that enterprises with limited money cannot effectively implement technological innovation.

Capital-Human Resources (-0.0108)
The link between human resources and capital is negative. Consequently, human resources and capital are not interdependent, and the quantity of capital does not have a substantial impact on human re-sources. This implies that even businesses with less funding may effectively oversee their human re-sources.

Technology-Human Resources (0.1823)
There is a favorable correlation between technology and human resources; however, the relationship is not very significant. Therefore, it implies that enterprises that allocate money to technology also pri-oritize their human capital, indicating that businesses that currently invest in technology expect their employees to receive training and possess the skills necessary to effectively utilize the technology.

Capital-Investment (0.4161)
There exists a direct and robust association between capital and innovation. Consequently, organizations with higher capital are inclined to make larger investments. Among the data analyzed, this correlation stands out as the strongest, suggesting that enterprises with greater capital tend to allocate more resources to various sectors, including technology, compared to organizations with lower capital.

Human Resources-Investment (0.1823)
Furthermore, a positive association exists between human resources and investment. However, it's important to recognize that this specific relationship is relatively weaker compared to the other elements under analysis. Therefore, investment has a minimal impact on human resources, and the association between these elements does not exhibit significant value. Ultimately, these ideals might have detri-mental effects on the firm. It is essential to allocate financial resources to invest in human resources for the organization's continuing development, both in terms of material growth and human well-being.

Technology-Investment (0.2289)
The most recent association between investment and technology is positive. Like the previously analyzed categories, there is a positive link, indicating that enterprises that invest in technological equipment and technology in general are also likely to invest more money overall. In summary, technical progress affects the company's entire investment.
In general, there appear to be some moderate and weak correlations among these variables. The findings indicate that capital can impact investment decisions, and organizations that allocate money towards technology also emphasize the growth and development of their human resources. To determine the precise effects and relationships among these components, a comprehensive study and statistical analysis would be necessary. This would offer more precise and in-depth understanding of the correlated factors and the nature of their relationship.

DISCUSSIONS

Depending on the technologies and their use within the organizations, the research was conducted through a questionnaire survey that the organizations completed, and the results can be drawn from this to answer the research questions.
RQ1: What is the importance of technology in relation to firm size categorization?

Based on the calculation of the average of the questionnaire survey by firm size categorization, technologies by firm size categorization have the highest significance for small and large enterprises. Micro and medium enterprises have medium levels of technology salience. According to Perdan et al. (2022), this may be due to the limited resources that these categories of enterprises have. Therefore, it is crucial to emphasize the importance of leadership and governance, and to clarify the significance and applicability of technology. The government can then help these enterprises minimize the cost of tech-nology adoption through subsidies, grants, or tax breaks. Joseph Ng (2023) confirms this fact, arguing that medium enterprises face significant challenges in implementing strategies due to limited resources, potentially leading to underutilization of technology. Thus, an important aspect in terms of the relevance of technology according to the categorization of firm size is their resource capacity, and it is essential that all types of enterprises invest their resources in technological equipment to continuously develop their potential.

RQ2: What are the interrelationships between the factors examined in organizations?

The linkages within the firm are a key aspect for organizations to achieve sustainable development and gain competitive advantage in the current uncertain times, with sustainable competitive advantage correlating with organizational performance (Astuti, Datrini, and Chariri, 2023). The largest correlation examined for companies is between capital and investment, which may ultimately affect technology investment, as companies with more capital also invest more in technological equipment. According to Dai, Hou, and Li (2021), there is a need to make targeted investments to upgrade regional industry and guide organizations to invest their capital in continuous upgrading. Conversely, the analysis reveals that the relationship between capital and human resources has the lowest value, indicating that even or-ganizations with limited capital can effectively manage their human resources, provided they set up their processes correctly. Fernández-Alles et al. (2022) support this by stating that limited resources, including capital, may not negatively impact human capital, which can compensate for these constraints and serve as a solid foundation for organizational prosperity.

Based on these results and the discussion of the results, it can be concluded that this article pro-vides unique and valuable perspectives and contributions to the local economy. Firstly, it emphasizes the notable discrepancies in the adoption of technology and its advantages among firms of varying sizes, offering a more distinct comprehension of how the allocation of resources impacts the integration of technology. This comprehension can assist policymakers and corporate executives in developing more efficient plans for the implementation of technology, specifically for small and medium-sized firms. Furthermore, the study highlights the crucial significance of leadership and governance in promoting the adoption of technology, emphasizing the necessity of favorable policies and incentives.

The main constraint of this study is its concentration on businesses in the Czech Republic, which may restrict the applicability of the results to other areas or nations with distinct economic circumstances and technology environments. Furthermore, the study is dependent on data obtained from self-reported responses from a questionnaire survey, which may be influenced by response biases. Subsequent studies could broaden the range by conducting a comparison examination with businesses in different geographical areas and integrating more unbiased criteria to evaluate the adoption of technology and its effects.

The study incorporates both theoretical and practical data to facilitate additional research. As a result, it has the potential to expand the scope of investigation to include other organizations and in-dustries operating in the Czech Republic market. All organizations operating in the market environment can apply its findings, as the thesis makes a dual contribution, both practically and theoretically. Therefore, the facts, analysis, results, and conclusions suggest that the thesis has achieved its objective.

CONCLUSIONS

Technology is of the utmost importance and should not be undervalued. Suleiman et al. (2022) have observed that the technological revolution has brought up several uncertainties and challenges for en-terprises. The current transformation, marked by swift progress in data analytics and digital management, requires a cohesive strategy for adopting Industry 4.0 technology. It is essential to fully understand the problems of technology from both a theoretical and practical perspective. This is because technology is vital for all businesses that operate in a market setting and strive for success. Holding the necessary technological tools and a thorough comprehension of their capabilities can have a favorable influence on an organization's competitiveness and sales. Enterprises must develop strong interconnections among the many components within the organization to ensure that the multiple links are mutually supportive and firmly formed. Organizations are increasingly utilizing technology to enhance their competitiveness and efficiency in the market. When considering the adoption of technology based on organization size, both small and large enterprises exhibit the highest level of importance placed on technology. Micro- and medium-sized firms have a moderate degree of technological importance. Of the variables examined, the highest value is linked to capital and investment, indicating that companies with higher financial re-sources are more inclined to invest, especially in technical equipment. The point where capital and human resources overlap exhibits the lowest value, indicating that effective business practices can obviate the necessity for significant capital investment while still enabling the organization to flourish in the market. The results of this study highlight the significance of implementing a deliberate strategy while adopting technology. Both small and large firms can obtain substantial advantages from technological improvements by utilizing their respective adaptability and resources. Micro- and medium-sized firms can overcome resource limits by focusing on strategic investment and using effective management methods. This is consistent with the requirement for strategic investments in technology, as proposed by Huamanchumo (2021), in order to promote a competitive edge. Venter and Grobbelaar (2023) emphasize that to succeed in a fiercely competitive climate, firms need to prioritize their ability to manage tech-nology effectively. To summarize, the efficient incorporation of technology into company operations is a complex undertaking that involves careful planning, strong leadership, and ongoing investment in both technological resources and human resources. Organizations can improve their competitiveness and maintain long-term sustainability in a digitalized market environment by taking a comprehensive ap-proach to technology management. This study offers unique insights into the dynamics of technology adoption and acts as a roadmap for future research and practical applications, ultimately contributing to the growth of the area of organizational management.

Conflict of interests

The author declares no any conflict of interests.

References

AMEJWAL, M., EL JAOUHARI, A., ARIF, J., FELLAKI, S. & JAWAB, F. (2022). Production Flow Management Based on Industry 4.0 Technologies. LOGISTIQUA2022. 40-46. ISSN 2166-7349. doi:10.1109/LOGISTIQUA55056.2022.9938064

ANGELOPOULOU, A., MYKONIATIS, K. & BOYAPATI, N. R. (2020). Industry 4.0: The use of simulation for human reliability assessment. ISM 2019. 42, 296-301. ISSN 2351-9789. doi: 10.1016/j.promfg.2020.02.094

ANGELOPOULOU, A., MYKONIATIS, K. & BOYAPATI, N. R. (2020). Industry 4.0: The use of simulation for human reliability assessment. International Conference on Industry 4.0 and Smart Manufacturing. 42, 296-301. ISSN 2351-9789. doi: 10.1016/j.promfg.2020.02.094

ASTUTI, P. D., DATRINI, L. K. & CHARIRI, A. (2023). Understanding the Antecedents and Consequences of Sustainable Competitive Advantage: Testing Intellectual Capital and Organizational Performance. Economies. 11(4). ISSN 2227-7099. doi:10.3390/economies11040120

DAI, Y. X., HOU, J. N. & LI, X. (2021). Industry policy, cross-region investment, and enterprise investment effi-ciency. Research in International Business and Finance. 56. ISSN 0275-5319. doi: 10.1016/j.ribaf.2020.101372

ELAFRI, N., TAPPERT, J., ROSE, B. & YASSINE, M. (2022). Lean 4.0: Synergies between Lean Management tools and Industry 4.0 technologies. IFAC Papersonline. 55(10), 2060-2066. ISSN 2405-8963. doi: 10.1016/j.ifacol.2022.10.011

FERNÁNDEZ-ALLES, M., HERNÁNDEZ-ROQUE, D., VILLANUEVA-FLORES, M. & DÍAZ-FERNANDEZ, M. (2022). The impact of human, social, and psychological capital on academic spin-off internationalization. Journal of International Entrepreneurship. 20(3), 433-473. ISSN 1570-7385. doi:10.1007/s10843-022-00311-4

FIGUEIRAS, P., LOURENCO, L., COSTA, R., GRACA, D., GARCIA, G. & JARDIM-GONCALVES, R. (2021). Big Data Provision for Digital Twins in Industry 4.0 Logistics Processes. IEEE METROIND4.0 & IOT. 516-521. doi:10.1109/METROIND4.0IOT51437.2021.9488507

FOCACCI, C. N. & PEREZ, C. (2022). The importance of education and training policies in supporting techno-logical revolutions: A comparative and historical analysis of UK, US, Germany, and Sweden (1830-1970)*. Technology in Society. 70, 251-274. ISSN 0160-791X. doi: 10.1016/j.techsoc.2022.102000

FOSFURI, A., GIARRATANA, M. S. & SEBREK, S. S. (2020). Resource partitioning and strategies in markets for technology. Strategic Organization. 18(2), 251-274. ISSN 1476-1270. doi:10.1177/1476127018791329

FUSTER, V. T. (2019). The Challenges of the Fourth Technological and Industrial Revolution in a European Pe-ripheral Society and Its Consequences in Political Governance and Legal Regulation. With Special Reference to the Valencian Community. Revista General De Derecho Administrativo. (50). ISSN 1696-9650. https://www-webofscience-com.ezproxy.techlib.cz/wos/woscc/full-record/WOS:000460613300014

GAJEK, A., FABIANO, B., LAURENT, A. & JENSEN, N. (2022). Process safety education of future employee 4.0 in Industry 4.0. Journal of Loss Prevention in the Process Industry. 75. ISSN 0950-4230. doi: 10.1016/j.jlp.2021.104691

GARCIA-LORO, F., PLAZA, P., QUINTANA, B., SAN CRISTOBAL, E., GIL, R., PEREZ, C., FERNANDEZ, M. & CASTRO, M. (2021). Laboratories 4.0: Laboratories for Emerging Demands under Industry 4.0 Paradigm. EDUCON. 909-915. ISSN 2165-9567. doi:10.1109/EDUCON46332.2021.9454095

GIUSTOZZI, F., SAUNIER, J. & ZANNI-MERK, C. (2018). Context Modeling for Industry 4.0: an Ontology-Based Proposal. KES-2018. 126, 675-684. ISSN 1877-0509. doi: 10.1016/j.procs.2018.08.001

GONG, L., ZHU, D. G., LIU, Y., XIE, J., MO, Z. C. & ZHU, M. R. (2021). Technology Network Construction and Analysis Method for Technology Trends Discovery. ICIEA 2021. 464-468. ISSN 2156-2318. doi:10.1109/ICIEA51954.2021.9516339

GOWRIPEDDI, V. V., BIJJAHALLI, M. C., JANARDHAN, N. & BHIMAVARAM, K. R. (2021). Role of Education 4.0 Technologies in Driving Industry 4.0. Cross Reality and Data Science in Engineering. 1231, 576-587. ISSN 2194-5357. doi:10.1007/978-3-030-52575-0_48

HERNANDEZ, J. G. V. & SALCEDO, M. T. A. (2020). The influence and benefits of technology as a strategy in organizations. Multidisciplinary Journal for Education Social and Technological Sciences. 7(1), 32-53. ISSN 2341-2593. doi:10.4995/muse.2020.10693

HUAMANCHUMO, P. A. G. (2021). Technological singularity and transhumanism. Teknokultura: Revista De Cultura Digital Y Movimientos Sociales. 18(2), 195-200. ISSN 1549-2230. doi:10.5209/TEKN.74056

JASKOVÁ, D. (2020). Labour Market Changes in the Regions of Slovakia in the Context of Industry 4.0. Vplyv Industry 4.0 Na Tvorbu Pracovnych Miest 2019. 183-191.

JIN, Y. G., DONG, N. Y., TIAN, G. L. & ZHANG, J. R. (2023). Wisdom of the masses: Employee education and corporate risk taking. Economic Modelling. 118. ISSN 0264-9993. doi: 10.1016/j.econmod.2022.106102

JOSEPHNG, P. S. (2023). Innovative Usage of Grid Solutions with a Technology Behavior Model in a Medium-Size Enterprise. Applied System Innovation. 5(1). ISSN 2571-5577. doi: 10.3390/asi6010011

KALENYUK, I., BOHUN, M. & DJAKONA, V. (2023). Investing in Intelligent Smart City Technologies. Baltic Journal of Economic Studies. 9(3), 41-48. ISSN 2256-0742. doi: 10.30525/2256-0742/2023-9-3-41-48

KIM, J., LEE, J., YANG, E. Y. & KANG, S. K. (2023). Technology forecasting from the perspective of integration of technologies: Drone technology. KSII Transactions on Internet and Information Systems. 17(1), 31-50. ISSN 1976-7277. doi: 10.3837/tiis.2023.01.003

KUCIAPSKI, M. (2019). How the Type of Job Position Influences

Technology Acceptance: A Study of Employees' Intention to Use Mobile Technologies for Knowledge Transfer. IEEE Access [online]. 7, 177397-177413. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2957205

LEE, N. & CLARKE, S. (2019). Do low-skilled workers gain from high-tech employment growth? High-technology multipliers, employment and wages in britain. Research policy 48(9). doi: 10.1016/j.respol.2019.05.012

LI, H. Z., AN, N. & LIU, J Y. (2021). Strategic planning, strategy map and management control: a case study. Nankai Business Review International. 12(3), 386-408. ISSN 2040-8749. doi: 10.1108/NBRI-10-2020-0054

LI, M. X. & HUANG, G. Q. (2021). Production-intralogistics synchronization of industry 4.0 flexible assembly lines under graduation intelligent manufacturing system. Internationa Journal of Production Economics. 241. ISSN 0925-5273. doi: 10.1016/j.ijpe.2021.108272

MASÁR, M. & HUDÁKOVÁ, M. (2020). The Impact of Human Factor on Projects of Implementation Concept In-dustry 4.0 in Manufacturing Enterprises. Vplyv Industry 4.0 Na Tvorbu Pracovnych Miest 2019. 268-276.

MESÁROVÁ, T., KORDOS, M. & SOKOL, J. (2019). The Impact of Industry 4.0 Initiative on Society. Vplyv Industry 4.0 Na Tvorbu Pracovnych Miest. 195-200.

MIN, H. (2022). Smart factory: a game changer or another fad in the era of fourth industrial revolution. Interna-tional Journal of Technology Management. 89(1-2), 26-45. doi: 10.1504/IJTM.2022.123006

NEWMAN, J., MINTROM, M. & O´NEIL, D. (2022). Digital technologies, artificial intelligence, and bureaucratic transformation. Futures. 136. ISSN 0016-3287. doi: 10.1016/j.futures.2021.102886

OZTURKLER, Z. (2021). Assessment of private Sector Employees in Terms of Education in Northern Cyprus. Propositos Y Representaciones. 9. ISSN 2307-7999. doi: 10.20511/pyr2021.v9nSPE1.854

PAPULOVÁ, Z., GAZOVÁ, A. & SUFLIARSKY, L. (2022). Implementation of Automation Technologies of Industry 4.0 in Automotive Manufacturing Companies. 3rd International Conference on Industry 4.0 and Smart Manufacturing. 200, 1488-1497. ISSN 1877-0509. doi: 10.1016/j.procs.2022.01.350

PERDANA, A., LEE, H. H., ARISANDI, D. & KOH, S. (2022). Accelerating data analytics adoption in small and mid-size enterprises: A Singapore context. Technology in Society. 69. ISSN 0160-791X. doi: 10.1016/j.techsoc.2022.101966

QUINN, W. (2019). Technological revolutions and speculative finance: evidence from the British Bicycle Mania. Cambridge Journal of Economics. 43(2), 271-294. ISSN 0309-166X. doi: 10.1093/cje/bey029

RAHMANI, A. M., EHSANI, A., MOHAMMADI, M., MOHAMMED, A. H., KARIM, S. T. & HOSSEINZADEH, M. (2022). A new model for analyzing the role of new ICT-based technologies on the success of employees' learning pro-grams. Kybernetes. 51(6), 2156-2171. ISSN 0368-492X. doi: 10.1108/K-02-2021-0164

REYES, S., LUIS, M., ORDONEZ, S. & BARBARA, A. (2020). Communication management for industries 4.0. Telematique. 19(1), 75-90. ISSN 1856-4194.

RICHNÁK, P. (2021). Key Concept of Industry 4.0 in Production Management in Slovak Wood Processing Man-agement. WOODEMA 2021.

ROSA, E. S., GODINA, R., RODRIGUES, E. M. G. & MATIAS, J. C. O. (2022). An Industry 4.0 Conceptual Model Proposal for Cable Harness Testing Equipment Industry. 3rd International Conference on Industry 4.0 and Smart Man-ufacturing. 200, 1392-1401. ISSN 1877-0509. doi: 10.1016/j.procs.2022.01.340

ROSALES, M. A., MAGSUMBOL, J. A. V., PALCONIT, M. G. B., CULABA, A. B., DADIOS & E. P. (2020). Artificial Intelligence: The Technology Adoption and Impact in the Philippines. HNICEM. ISSN 2475-7152.

RUSSO, M. (2020). 39% of Firm Leaders Expect That They Will Invest in New Technology This Year. Architect. 109(6), 51-51. ISSN 1935-7001.

SCEPANOVIC, S. (2019). The Fourth Industrial Revolution and Education. MECO. 704-707. ISSN 2377-5475.

SU, C. W., YUAN, X., UMAR, M. & LOBONT, O. R. (2022). Does technological innovation bring destruction or creation to the labor market? Technology in Society. 68. ISSN 0160-791X. doi: 10.1016/j.techsoc.2022.101905

SULEIMAN, Z., SHAIKHOLA, S., DIKHANBAYEVA, D., SHEBAB, E. & TURKYILMAZ, A. (2022). Industry 4.0: Clustering of concepts and characteristics. PRODUCTION & MANUFACTURING. 9(1). doi: 10.1080/23311916.2022.2034264

TANER, Z. T. & PARLAK BICER, Z. Ö. (2021). Effect of Industry 4.0 on Project Management Factors. Journal of Polytechnic-Politeknik Dergisi. 24(4), 1461-1472. ISSN 1302-0900. doi: 10.2339/politeknik.741566

VAN ZEEBROECK, N., KRETSCHMER, T. & BUGHIN, J. (2023). Digital "is" Strategy: The Role of Digital Tech-nology Adoption in Strategy Renewal. IEEE Transactions on Engineering Management. 70(9), 3183-3197. ISSN 0018-9391. doi: 10.1109/TEM.2021.3079347

VENTER, S. & GROBBELAAR, S. (2023). A Technology Management Capabilities Framework for Technology Platforms. IEEE Transactions on Engineering Management. 70(7), 2558-2573. ISSN 0018-9391. doi: 10.1109/TEM.2022.3172720

WAMSGANSS, A., BRORING, S., SALOMO, S. & SICK, N. (2023). Technology strategies in converging technology systems: Evidence from printed electronics. Journal of Product Innovation Management. ISSN 0737-6782. doi: 10.1111/jpim.12693

ZEBA, G., DABIC, M., CICAK, M., DAIM, T. & YALCIN, H. (2021). Technology mining: Artificial intelligence in manufacturing. Technological Forecasting and Social Change. 171. ISSN 0040-1625. doi: 10.1016/j.techfore.2021.120971

ADVANCED COMPUTER ARCHITECTURE OPTIMIZATION FOR MACHINE LEARNING/DEEP LEARNING

Shefqet Meda [1], *Ervin Domazet* [2]

[1] Department of Electronics & Telecommunications Engineering, Faculty of Engineering, Canadian Institute of Technology, Albania, shefqet.meda@cit.edu.al, ORCID:0000-0003-3654-6606

[1] Department of Computer Engineering, Faculty of Engineering, International Balkan University, Skopje, North Macedonia, meda.shefqet@ibu.edu.mk, ORCID:0000-0003-3654-6606

[2] Department of Computer Engineering, Faculty of Engineering, International Balkan University, Skopje, North Macedonia, ervin.domazet@ibu.edu.mk, ORCID:0000-0001-7743-469X

Abstract

The recent progress in Machine Learning (Géron, 2022) and particularly Deep Learning (Goodfellow, 2016) models exposed the limitations of traditional computer architectures. Modern algorithms demonstrate highly increased computational demands and data requirements that most existing architectures cannot handle efficiently. These demands result in training speed, inference latency, and power consumption bottlenecks, which is why advanced methods of computer architecture optimization are required to enable the development of ML/DL-dedicated efficient hardware platforms (Engineers, 2019). The optimization of computer architecture for applications of ML/DL becomes critical, due to the tremendous demand for efficient execution of complex computations by Neural Networks (Goodfellow, 2016). This paper reviewed the numerous approaches and methods utilized to optimize computer architecture for ML/DL workloads. The following sections contain substantial discussion concerning the hardware-level optimizations, enhancements of traditional software frameworks and their unique versions, and innovative explorations of architectures. In particular, we discussed hardware including specialized accelerators, which can improve the performance and efficiency of a computation system using various techniques, specifically describing accelerators like CPUs (multicore) (Hennessy, 2017), GPUs (Hwu, 2015) and TPUs (Contributors, 2017), parallelism in multicore architectures, data movement in hardware systems, especially techniques such as caching and sparsity, compression, and quantization, other special techniques and configurations, such as using specialized data formats, and measurement sparsity. Moreover, this paper provided a comprehensive analysis of current trends in software frameworks, Data Movement optimization strategies (A.Bienz, 2021), sparsity, quantization and compression methods, using ML for architecture exploration, and, DVFS (Hennessy, 2017),, which provides strategies for maximizing hardware utilization and power consumption during training, machine learning, dynamic voltage, and frequency scaling, runtime systems. Finally, the paper discussed research opportunity directions and the possibilities of computer architecture optimization influence in various industrial and academic areas of ML/DL technologies.

The objective of implementing these optimization techniques is to largely minimize the current gap between the computational needs of ML/DL algorithms and the current hardware's capability. This will lead to significant improvements in training times, enable real-time inference for various applications, and ultimately unlock the full potential of cutting-edge machine learning algorithms.

Keywords: Computer Architecture Optimization, Machine Learning, Deep Learning, Parallelism, Sparsity, Data Movement Optimization, Quantization, Compression, Software Framework Optimization, DVFS, TPU, CPU, GPU, TensorFlow, Pytorch.

1. INTRODUCTION

Machine Learning/Deep Learning are among the rapidly growing and most transformative fields of science. Image recognition, natural language processing, health care, autonomous vehicles, and several other fields of applications have seen a massive progress in recent years. This progress is driven by the development of increasingly complex models capable of tackling ever-more challenging tasks. As the scale and complexity of ML/DL applications continue to grow, there is an increasing demand for efficient and scalable computing infrastructure to support these workloads. However, these advancements are often bottlenecked by the limitations of traditional computer architectures. These architectures, designed for general-purpose computing, struggle to efficiently handle the massive computational demands of modern ML/DL algorithms. The inefficiency manifests in several ways:

*Corresponding author:

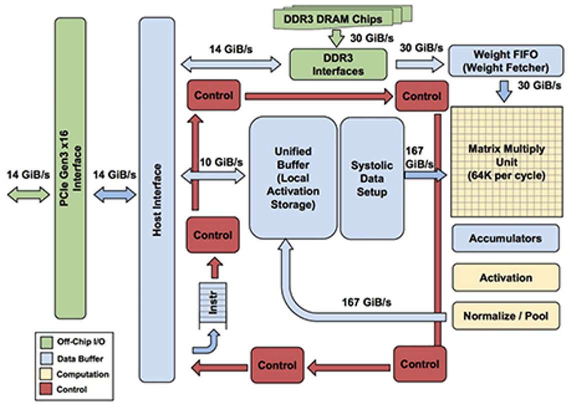
- **Training takes too long.** Entire teams are spending days or weeks training models to find better hyperparameters and experiments in fine print.
- **High Inference Latency** (Hennessy, 2017),. Deploying these models on resource-constrained devices, like mobile phones and embedded systems, is often impractical due to the high computational requirements for real-time inference.
- **Significant Power Consumption.** Large-scale training of complex models incurs substantial energy costs, raising concerns about sustainability and cost-effectiveness. To solve these challenges, we shall focus on advanced computer architecture optimization for ML and DL. This research theme encompasses specialized hardware designs and software techniques that target ML/DL applications' unique characteristics. By optimizing the underlying hardware and software infrastructure, we consider the following objectives-to be achieved:

    - Accelerate and optimize training speeds. This involves reducing the time required to train complex ML/DL models, which speeds up the development process and allows for more iterative experimentation.

    - We will also aim to develop real-time inference capabilities.
    This will allow us to deploy ML /DL models to inferences close to the edge devices through faster, less efficient hardware, thereby enabling a greater range of AI applications on - devices.

    - Finally, we will optimize the power efficiency of ML/DL models.
    This last objective seeks to reduce power consumption needed for training and inference, which will make it more feasible and cost-effective to deploy ML/DL models.

This paper discusses the techniques of advanced computer architecture optimization for ML/DL. We will discuss the specific challenges faced with traditional architectures, analyse relevant hardware and software optimization strategies, and explore the potential benefits of these techniques. We will focus on techniques and strategies that enhance performance, energy efficiency, and scalability. The optimization of computer architecture encompasses a wide range of topics, except above mentioned, including hardware design, software frameworks, and system-level optimizations, all of which are critical for achieving optimal performance in ML/DL workloads. Specialized hardware accelerators (Sze, Chen, Yang, & Emer, 2017) such as CPUs (multicore), GPU, TPU, and custom ASICs have shown more efficiency on ML/DL computation. These accelerators are specifically designed to accelerate the matrix and tensor operations that are prevalent in ML/DL applications further neural network computations, enabling faster model training and inference. They have made it feasible for model training to be faster.



TensorFlow processing unit architecture. Source: Google (Google, 2017)

In addition to hardware optimizations, software frameworks and runtime systems also contribute significantly to optimizing computer architecture for ML/DL applications. While frameworks such as TensorFlow (TensorFlow, 2024), PyTorch (Migacz, 2024), and Apache MXNet (apache.org, 2024) offer higher-level abstractions and API to build, train, and deploy neural network models (Sze, Chen, Yang, & Emer, 2017), enabling faster model training and inference.
Runtime systems are also optimized to run ML/DL workloads on diverse hardware platforms. Moreover, machine learning techniques for architecture exploration have shown promise in optimizing computer architecture for ML/DL applications. Specifically, the techniques using supervised learning, reinforcement learning, and Bayesian optimization (Tony Pourmohamad, 2021) have proved very effective in exploring the design space and discovering the optimal configuration to optimize performance and power efficiency. In conclusion, advanced optimization of computer architecture for ML/DL is a multidisciplinary field that mixes expertise in hardware, software, and machine learning. Through state-of-the-art in computer architecture optimization, researchers and engineers can unlock incredible possibilities for ML/DL technology that results in groundbreaking AI advancements and applications across a range of fields. This paper has tried to provide a clear picture of this exciting and rapidly growing field and its recent progress and future work.

2. HARDWARE OPTIMIZATION:

ACCELERATORS – CPU, GPU, TPU

Hardware optimization is a technique used to enhance the performance and efficiency of computing systems (Larkin Ridgway Scott, 2021). It involves increasing performance of accelerators like GPUs, TPUs and FPGA, considering designing to handle heavy workloads more efficiently. Traditional

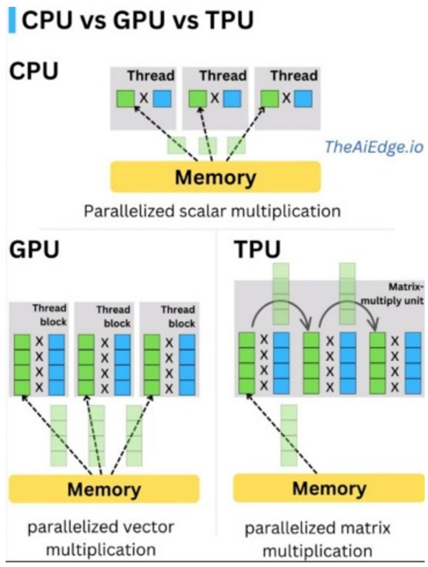
ones have their limitations include: CPUs not being efficient in support for upsized vectorized instructions such as AVX, along with efficient memory access to handle large datasets and model parameters. Limitations in parallelism is slowing down performance (Hennessy, 2017), in some of the highly parallel tasks like Matrix multiplications. Most of them are not fully optimized for some ML/DL operations and the high-power consumption among others.
In the below Pseudocode 1, an example of how a TPU might function to achieve matrix multiplication is presented.

```
// Pseudocode just for example
void matrixMultiply(float* X, float* Y, float* Z) {
   // Load matrices X and Y into on-chip memory
  MXE_Load(X, Y);

   // Perform matrix multiplication using MXE hardware
  MXE_Multiply();

  // Store result matrix Z in on-chip memory
  MXE_Store(Z);
}
```

Pseudocode 1: *TPU pseudocode for matrix multiplication.*

There are many strategies and efforts to optimize hardware. After we have carefully studied them, we have grouped of the most optimal ones that can be considered applied not specifically to solve specific problems and tasks but to cope with the processing, analysis, and output of significant amounts of data for applications of Artificial Intelligence (ML/DL):
- Parallelism (Hennessy, 2017),. This involves utilizing the parallel processing capabilities of the hardware architecture. It enables multiple processing units within the core, CPU to compute multiple independent instructions; other techniques include SIMD and distributed computing.
- Vectorization (Hwu, 2015). It considers vector processing units such as SSE (Streaming SIMD Extensions) and AVX in CPUs and GPUs to perform operations on more than one data simultaneously. It sends one operation to the core simultaneously sent to multiple data elements.
- Cache Optimization (Hendrik Borghorst, 2019). Maximizing cache utilization to reduce memory access latency and improve data locality. Techniques include cache blocking, prefetching, and data alignment.
- Instruction-Level Optimizations. Optimizing code to make the best use of instruction pipelines, branch prediction, and instruction scheduling.
- Memory Optimization (Joo-Young Kim, 2022). This reduces memory consumption and accesses, to reduce stalls and improve overall performance.

- Data Layout Optimization. Organizing data structures and memory layouts to optimize for cache coherence, spatial locality, and access patterns.
- Algorithmic Optimization. Choosing or designing algorithms that are well-suited for the underlying hardware architecture, minimizing unnecessary computations and memory accesses.
- Power Efficiency Build. It is especially critical to mobile devices and some limited energy environments, particularly when predicting systems in real-time.
- Profiling and Benchmarking. This process allows optimizing critical codes by observing or finding the bottlenecks of performance.
- Hardware-Specific Optimization. Specific features such as specific to CPU, GPU, or TPU architectures, specialized instruction sets, tensor cores, or memory hierarchies.
- Furthermore, the below flowchart explains hardware specialization. While traditional CPUs serve as effective general-purpose chips, TPUs or AI accelerators have dedicated characteristics based on the specific requirements for DL



CPU, GPU, or TPU architectures. Source: eitc.org (eitc.org, 2023)

- **Compiler Optimizations.** Making use of compiler optimizations to automatically transform code for better performance, including loop unrolling, function inlining, and auto-vectorization.
- **FPGA and ASIC Optimization:** For specialized hardware like FPGAs and ASICs, optimization involves hardware design at a lower level, focusing on logic utilization, routing efficiency, and clock frequency (Wijtvliet, 2019).
We can recommend another approach like

algorithms specifically for FPGA execution, considering the parallel and pipelined nature of FPGAs. An example is conclusions of a group of researchers for accelerating the learning process of CRBM (Restricted Boltzmann Machine) with FPGAs and OpenCL, and conducting an extensive scalability study for different model sizes and system configurations (Zoran Jakšić, 2020).

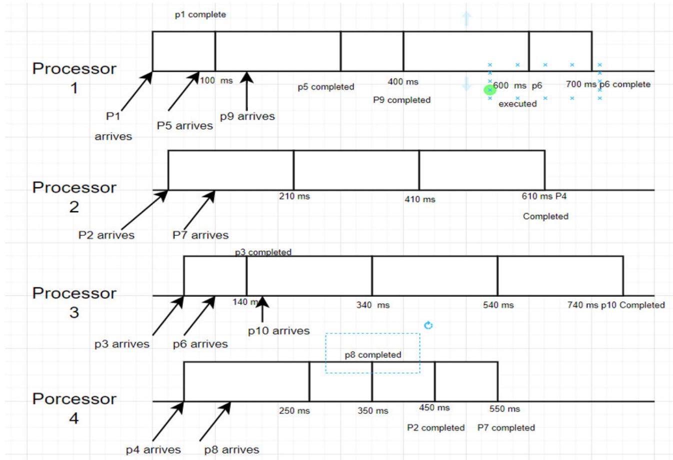
- **Quantization and Pruning:** Additional optimizations like for neural network accelerators such as TPUs include Quantization (Jacob, 2018) and pruning accelerate computation by reducing the precision of weights and activations of weights and removing redundant connections of the neural network. Give developers exercises to apply these enhancements to hardware accelerators across all computing platforms.

3. PARALLELISM - MULTICORE ARCHITECTURES

Utilizing parallelism is a fundamental consideration of computer architecture optimization for ML/DL applications, especially when it comes to multicore architectures. Enabling computers to utilize parallelism is critical for expediting the compute-intensive nature of ML/DL workloads and realizing the desired performance and efficiency levels (Larkin Ridgway Scott, 2021). There are several restrictions connected with traditional computers when it comes to parallelism: they were not designed to capitalize on parallelism in the context of ML/DL workloads by parallelizing matrix operations, convolutions, and other operations conducted by neural networks across various CPU cores.
Such multicore architectures (Hennessy, 2017), enable the following benefits, over traditional one's, for ML/DL on CPUs:
- faster training and inference,
- added scalability,
- parallel processing power.

Multicore CPUs put more cores on the same chip, essentially allowing them to run more instructions at once. This leads to better performance for ML/DL tasks that can be subdivided into separate computations.

- **Core specialization:** Modern cores are rich in dedicated units, such as vector processing units. These types of units can run the same operation on multiple data elements simultaneously. This is very beneficial for ML/DL workloads that involve a lot of vector or matrix computations.
- **Multiple cores that can run in parallel:** Multiple cores on a CPU can operate at the same time. When subtasks of certain instructions can be run at the same time, multicore CPUs will show an excellent parallel processing speedup. For instance, ML/DL heavily rely on linear algebra and matrix multiplication, which can be trivially partitioned per sub-step. To implement CPU, all the processor splits the workload per step and runs multiple cores to speed up the whole calculation.
- **Thread-level parallelism:** Modern CPUs support additional execution threads along cores, so-called hyper-threading or simultaneous multi-threading Further enhance parallelism because it keeps the cores busy even while waiting for data or cheesing on another thread. Below is an example of the application of quantization according to Round Robin in the multicore architecture and a part of the corresponding code. If the outputs for two different algorithms are compared, as we did, the superiority of quantization and multicore in execution is clearly observed.



Multicore architecture running threads and code

```
int main() {
  int n, quantum;
  cout << "Enter The Number of Processes: ";
  cin >> n;
  cout << "Enter The Time Quantum: ";
  cin >> quantum;

  Process processes[n];
  cout << "Enter process details:\n";
  for (int i = 0; i < n; i++) {
    cout << "Process " << i + 1 << ":\n";
    processes[i].id = i + 1;
    cout << "   Arrival Time: ";
    cin >> processes[i].arrivalTime;
    cout << "   Burst Time: ";
    cin >> processes[i].burstTime;
  }

  calculateTimes(processes, n, quantum);
  calculateTurnaroundTime(processes, n);
  calculateWaitingTime(processes, n);

  cout << "\nRound Robin Scheduling Results:\n";
  printTable(processes, n);

  return 0;
}
```

- **Software optimization for Multicore Exploitation:** Be able to save the workload of multiple cores properly, the framework has to be appropriately designed. Software frameworks like TensorFlow or PyTorch use python Threading to subdivide the load of data on the CPUs. They chop small slivers of data, and the computation is about a model into smaller clones of the data and add them to separate cores (sample above).

- Libraries such as OpenMP or Intel Threading Building Blocks (Ruud Van Der Pas, 2017): Provide tools for programmers to write code that takes advantage of multiple cores.

• **Data and Model Parallelism:** Both data and model parallelism techniques are enabled by multi-core architectures. Splits the training data into smaller batches and distributes them across multiple cores. Each core processes its assigned data independently, accelerating computations, especially for large datasets.

- Model parallelism divides the machine learning or deep learning model across cores. For extremely large and difficult models which would not fit on a single core, this is necessary. However, it must be done with caution, since the various cores must collaborate to ensure the model changes are coherent.

- Simple code in Python:

```
# Pseudocode for data parallel training with
model parallelism on two GPUs
def train_step(data_batch):
# Split data batch across GPUs
data_batch_gpu1, data_batch_gpu2 = split_
data_batch(data_batch)
# Forward pass on GPU 1 (first half of the
model)
with tf.device("/gpu:0"):
activations1 = model_part1(data_batch_gpu1)
# Send activations from GPU 1 to GPU 2
activations1 = send_data(activations1, "/
gpu:1")
# Forward pass on GPU 2 (second half of the
model)
with tf.device("/gpu:1"):
output = model_part2(activations1)
# Backpropagation and gradient accumulation
(simplified)
return output
# Training loop
for epoch in range(num_epochs):
for data_batch in training_data:
output = train_step(data_batch)
# ... update model weights
```

Pseudocode 2: *For data parallel training with model parallelism on two GPUs*

- Software Optimization Frameworks like TensorFlow and PyTorch are optimized in deployed on multicore architectures. Automatically divide data and models across cores for parallel processing. cores must communicate and collaborate on model changes.

3.1. Challenges and Considerations

While it is easy to understand the benefits of processing multiple tasks simultaneously, a variety of factors determine how effective such parallel execution can be. While cores can simultaneously perform in parallel, communication between them may be the bottleneck if they have to share data and any updates. Techniques such as gradient accumulation and efficient communication protocols are critical. Additionally, optimized memory access patterns are crucial to success.
Cores must be allowed to quickly and efficiently access whichever data they require from memory, and if this is not allowed or limited by memory bandwidth, the numbers of parallel operations cannot be optimized for faster performance. Finally, not all parts of an ML/DL application can be perfectly parallelized. Amdahl's law determines the maximum speedup due to parallelism. Since there will always be serial sections of the code that cannot be divided into parallel parts, the maximum speedup can only go so far. Memory bandwidth influences cannot be avoided; rather than intensified, these drawbacks are magnified.

Conclusion: so we can conclude that, conventional CPU architectures are vial in achieving the parallelism in ML/DL applications. When combined with several cores and specialty hardware and software representations, they lead to quicker trainand and recast times. Optimal algorithms and communication strategies will be key to ensure the best potential from these machines as core counts proceed to expand.

We suggest that the use of high-level languages, or otherwise modern programming languages and open source platforms would be another opportunity to enable researchers and developers to provide optimal solutions for ML/DL applications and increasing performance.

4. DATA MOVEMENT OPTIMIZATION - DATA PREFETCHING, MEMORY HIERARCHY, CACHING

Traditional computer architectures can have limitations in data movement that can impact the performance and efficiency of Machine Learning (ML) and Deep Learning (DL) workloads. Here's are listed some aspects how these limitations affect ML/DL specifically:

Data Prefetching. In traditional architectures, software prefetching is used, where the processor asks for the data from memory explicitly before its need. This can introduce latency if the request isn't anticipated correctly or if there's concurrency for memory bandwidth. In ML/DL, algorithms often exhibit predictable access patterns, repeatedly accessing specific data elements. Inefficient prefetching can lead to situations where the needed data isn't readily available when the processing unit requires it, causing delays.

Memory Hierarchy Bottlenecks (Joo-Young Kim, 2022). Traditional architectures utilize a hierarchy of memories with varying speeds and capacities (e.g., registers, cache, main memory, disk). Accessing data from slower levels in the hierarchy takes longer. ML/DL algorithms often work with large datasets. If the frequently accessed data resides in slower memory, retrieving it for each operation can significantly impact performance.

Another limitation is - Storage I/O Bottlenecks. Traditional architectures use mechanical hard disk drives or even slower to use this data for storing load from memory, and the storage I/O bottleneck can occur in ML/DL systems. The problem with storage I/O bottlenecks is that systems require huge data and input/outing this big data makes the system slow. There are slow read/write speeds and high latency in storage devices that can prevent ML/DL storage from efficient work, especially when loading data and preprocessing it. Therefore, these bring us to the following limitations that bottleneck ML/DL workloads:

Increased Latency and Reduced Throughput.
Delays caused by inefficient prefetching, accessing slower memory levels, and cache misses all contribute to longer execution times for training and inference. Limited data movement capabilities can restrict the amount of data processed per unit time, hindering overall performance.
What we can do to optimize and avoid these limitations in performing with ML/DL apps includes:

Data Prefetching:
- Software Prefetching: by predicting and prefetching data into the cache before accessed by the computation to reduce latency, the computation can be overlapped with data movement.
- Hardware Prefetching: Exploit hardware mechanisms of modern processors to automatically prefetch data into cache according access patterns observed during runtime.
- Spatial and Temporal Prefetching: Prefetch data blocks that are spatially or temporally related to the currently accessed data, such as prefetching neighboring elements in arrays or prefetching data expected to be accessed in future iterations.

Memory Hierarchy Optimization:
- Cache Awareness (Hendrik Borghorst, 2019): design algorithms and data structures that can maximize cache usage and take advantage of spatial and temporal locality.
- Cache blocking: Partition data into blocks that fit into the cache, enabling reuse of cache lines and reducing the frequency of cache misses.
- Cache Replacement Policies: Use cache replacement policies (e.g., LRU, LFU, Optimal) that prioritize keeping frequently accessed data in the cache to minimize cache thrashing.
- Cache Compression: compress the stored data in the caches to increase the capacity of the caches and lessen the impact of cache pollution.

Caching (Hendrik Borghorst, 2019):
- Data Reuse: identify and leverage the opportunities to reuse intermediate results and computation to minimize data movement across memory levels.
- Result Caching: considering that computations in DNN are iterative, the results can be stored in memory to prevent repetitive computations and lower latency.
- Model Caching: the trained models and parameters can be cached in memory to reduce the load on the main memory and enhance the performance. The example illustrated below is an application of Fuzzy Logic in cache optimization by applying dynamic membership. As can be seen from the result, we have an execution time almost twice as short.

RESULT AND PERFORMANCE EVALUATION

Process ID	Arrival Time (ATI)	Burst Time (BTI)	Static Priority (PTI)	Dynamic Priority (DPI)	Deadline (D)
P1	0	18	1	0.136	53
P2	0	2	3	0.894	18
P3	0	1	2	0.95	13
P4	0	4	6	0.79	35
P5	0	3	5	0.84	22
P6	0	12	1	0.917	15
P7	0	13	7	0.58?	53

Comparison Table

Algorithm	Average Waiting Time	Average Turnaround Time	Average Response Time
Priority Algorithm	23.86	31.43	23.86
IFCS	14.86	22.43	14.85
AFCS	14.71	22.43	14.57

Fuzzy Logic in cache optimization with dynamic membership

Memory access patterns.
- Sequential Access Optimization: Arrange data structures and access patterns to optimize for sequential memory access, minimizing the overhead of random memory accesses.
- Data Layout Optimization: Organize data in memory to improve cache locality and alignment

with the underlying hardware architecture, such as using row-major or column-major layouts for matrices depending on access patterns.
• Strided Access Optimization: Minimize the stride (distance between consecutive elements) in memory access patterns to improve cache utilization and reduce memory access latency.

Data Compression and Encoding:
- Lossless Compression: Compress data before storing it in memory or transferring it between levels of the memory hierarchy to reduce storage requirements and bandwidth usage.
- Lossy Compression: Apply the lossy compression technique to sacrifice precision for better memory bandwidth utilization through varied sparsity methods for ML/DL such as quantization and low-precision data representation.

Advantages of Data Movement Optimization:
- Decrease Training and Inference Time: By minimizing wasted time waiting for data movement, these techniques can significantly accelerate the overall training and inference processes in ML/DL applications.
- Improved Resource Utilization: Efficient data movement allows for better utilization of processing power by reducing idle time due to data access delays.

5. SPARSITY: SOME SPECIAL TECHNIQUES, LIKE USING SPECIALIZED DATA FORMATS AND HARDWARE UNITS.

- Sparsity (Brandon Reagen, 2017) is essential in the computer architecture of Machine Learning and Deep Learning applications because it refers the presence of a large number of zeros in the data being processed. Some of the most relevant limitations from several issues that the traditional computer has when performing ML/DL apps are as follows:
- Sparsity Handling: Generally, sparsity is not directly addressed in traditional computer architecture design because most algorithms and hardware units are designed to work with the dense data format efficiently without specific optimizations for sparse data.
- Data Formats: Classic computer architectures support standard numerical data format, like float-point or integer, and do not directly support sparse data representation.

Hardware units: Sparse data is leveraged for many high-performance applications because the units in classic computer architecture, including CPUs and GPUs, are optimized for dense computations; they are not equipped with specialized units optimally designed for sparse data. Sparse matrix-matrix multiplication and other sparse operations are performed on these general-purpose hardware units, providing suboptimal performance due to the sparse data. Specialized techniques yield significant performance improvements because of the sparsity of the data.
The following are some of the aspects of computer optimization for ML/DL applications that use sparsity:

Sparse data formats: Compressed sparse formats. These arise through the use of compressed formats such as Compressed Sparse Row (CSR) or Compressed Sparse Column (CSC) to represent sparse matrices efficiently. They only store the non-zero elements, with their indices, yielding reduced storage requirements and better cache locality.

Sparse tensor representation: The data stored in these structures will be sparse, so the storage pattern must account for this. Design specialized data structures to optimize the storage and access patterns because sparse data structures such as graphs and embeddings are seen frequently in ML/DL models

Sparse matrix operations: Sparse matrix multiplication. It can be performed by developing optimized algorithms and units for sparse matrix multiplication, as this is fundamental to ML/DL applications. Memory access reduction and computational complexity can be achieved by explicitly identifying how to achieve modular multiplication based on sparsity.
Sparse tensor operations. We can extend hardware accelerators to support the computation of sparse tensors as the representations and operations involved in many ML/DL applications are seen as tensors.

Sparsity-aware hardware acceleration: Sparse tensor cores. Design specialized hardware units capable of efficiently handling sparse tensor operations, exploiting sparsity to reduce computational and memory requirements. These units can perform computations only on non-zero elements, improving energy efficiency and performance for sparse ML/DL models. Sparsity-Aware Memory Systems in developing memory architectures and controllers optimized for handling sparse data access patterns, with mechanisms for efficient storage and retrieval of non-contiguous data elements in sparse tensors.

Quantization and Pruning: Quantize models or model activations for sparse representation and reduce the precision for zero and near-zero values while maintaining the accuracy of non-zero elements. Sparse quantization can considerably reduce memory bandwidth and reduce memory storage overheads required to support large-scale ML/DL models. Remove redundant or insignificant weights from neural network models to obtain sparse weight matrices. Sparse weight matrices can be further sparsified using special storage formats and hardware units designed specifically for sparse data representation. Moreover, sparse weight matrices will reduce the memory overhead and the computational complexity of their processing during inference.

Dynamic Sparsity: Develop algorithms and hardware mechanisms that can detect sparsity and use the acquired knowledge durably at runtime. Dynamic sparsity algorithms and methods may detect variation in sparsity and adjust the computational approach based on changes in input data, sparsity models, and other factors. Integrating specialized data formats, hardware units and using sparsity – aware algorithms in computer systems can significantly accelerate ML/DL applications, improve their energy efficiency, and enhance scalability.

6. QUANTIZATION AND COMPRESSION.

The increasing size and complexity of machine learning based algorithms and models create two main challenges: it may not be possible to deploy an ML/DL model on the resource – constrained devices, and training such big models may be slow. The number of bits in which the data are represented in the model is reduced using quantization. In a traditional setting, high precision formats like 32-bit floating point numbers are used to represent the weights and activations of a model. Techniques considered in quantization include being able to represent these values in lower precision formats such as 8-bit integers with minimal effects on the model's accuracy. The model's memory footprint is significantly reduced, easing the storage on devices with fewer memories. During training, it reduces the storage knowledge footprint. Compression techniques are considered to work in tandem with quantization as quantization is focused on reduction of the redundant bits within an encoding while compression thrives from a reduction in information in the representation. To remove redundancy information within the representation, a for example may be removed. Examples of compression approaches include pruning, where unimportant figures are removed from the model or knowledge distillation, where a shallower model is trained on the end weights to mimic the complex model. The limitations of the traditional computers that affect quantization and compression (Jacob, 2018) in ML/DL applications are: Compression and quantization are unfamiliar concepts in classic computer architecture. In classic computer architecture, compression algorithms aim at storage and transmission of the data. The data can be files, as in Huffman compression, or progression, as seen in LZ compression techniques. Sounds and images are quantized to reduce the size before being transmitted since this approach savesbits. Data quantization and compression are techniques used in many computer ways in storage, transmission, and retrieval. These two techniques have not been adequately studied since accuracy in representation is less of concern. The critical issue is reduced bit storage while assured of minimal to no information or precision loss. A crucial area that further needs to be researched is promoting zero or one firing without compromising the model since all that does is alters the model behaviour. Identifying this optimal precision or compression level is a challenge if the problem area or the algorithms do not dictate the amount of information retained in a unit.

Techniques in ML/DL apps.

- quantization and compression algorithms to be used to reduce the precision of model parameters and activations thereby decreasing memory usage and computational complexity

- while preserving model accuracy techniques like weight pruning, quantization, and low-rank factorization are commonly employed to compress neural network models for deployment on resource-constrained devices or for efficient distributed training.

Benefits

- Reduced memory footprint. Lower precision data representations require less storage space, enabling the training of larger models on limited hardware resources
- Faster training and inference. Reduced precision leads to faster computations during the training and inference stages
- Lower power consumption. Lower power consumption due to smaller data footprints and potentially less complex computations makes these techniques attractive for mobile and embedded devices

Challenges

- Finding the Right Balance: Determining the optimal level of quantization or compression requires careful consideration, as too much reduction in precision can significantly impact model accuracy.
- Framework and Hardware Support: Not all frameworks and hardware platforms offer the same level of support for various quantization and compression techniques.
- Potential Need for Algorithm Adjustments: Depending on the technique used, algorithms within the neural network might need adjustments to maintain accuracy with lower precision data.

7. SOFTWARE FRAMEWORK OPTIMIZATION

Software framework optimization is a critical aspect of computer architecture optimization for machine learning (ML) and deep learning (DL) applications, facilitating efficient execution of ML/DL workloads on diverse hardware platforms while maximizing performance and scalability. ML/DL frameworks provide high-level abstractions and APIs for developing, training, and deploying neural network models, abstracting away hardware-specific details and enabling portability across different architectures. (Abadi, 2016). This kind of optimization serves to support the efficient execution of machine learning and deep learning workloads on any hardware platform and to maximize ML/DL task performance and out-of-the-box scalability. ML/DL frameworks are high-level abstractions and APIs that let users create, train and deploy neural network models without worrying about underlying hardware-specific details. However, before listing what might be optimized to enhance the performance of traditional computers, let's outline what such computers imply prevents such optimizations:
- Software frameworks. In classic computer architecture, software frameworks are developed for any general-purpose computing work that may apply to many different applications. They include an operating system, a programming language and libraries.
- Limited Hardware acceleration (Vinh Nguyen, 2020). Hardware acceleration to leverage optimizations for explicit instruction sets and memory-specific optimizations is mostly not used by this framework.
- Programmer Responsibility. More of the burden of optimization code is on the program, using techniques like manual vectorization or optimization of cache sometimes. Lower the frameworks offer much lower to no leverage of hardware-specific optimizations for specific instruction set and memory hierarchy.
- Lower-Level Control. The frameworks provide greater control over memory and instruction schedule but require more effort from the programmer to optimize.

Now let's see what can be optimized under given constraints.

- Optimizing ML/DL frameworks to leverage hardware-specific features and optimizations, such as specialized instructions, memory hierarchy, and parallelism capabilities. Frameworks like TensorFlow, PyTorch, Keras, and Apache MXNet incorporate backend optimizations to exploit hardware acceleration, enabling efficient execution of neural network computations on CPUs, GPUs, and specialized accelerators like Tensor Processing Units (TPUs). These optimizations include autotuning techniques, kernel fusion, and memory layout optimizations tailored to specific hardware architectures, ensuring optimal performance across a wide range of hardware platforms. They offer High-Level Abstractions, Automatic Optimization, and Focus on Parallelism
- Focus on minimization of overhead and improvement of efficiency in data processing and model inference pipelines. This is possible due to for example model quantization, tensor slicing, or even asynchronous execution which allows for efficient memory usage and pipelining of calculations. In both cases, latency is minimized and throughput is maximized for ML/DL workflows.
- Distributed training and inference are also considered to require minimal overhead with, for example, parameter servers or communication primitives, which allow distributed computing to be scalable and fault-tolerant. This extends to runtime systems and libraries supporting ML/DL applications and offering low-level abstractions and optimizations. Libraries such as cuDNN, or math Kernel Library MKL-DNN known for optimized implementations of common neural network operations. They utilize hardware-specific optimizations and parallelism for GPUs, and CPUs that can be used during ML and DL computational tasks to improve the efficiency and performance of neural networks.

8. DYNAMIC VOLTAGE AND FREQUENCY SCALING (DVFS)

DVFS is a power management technique that adjusts the voltage and frequency of a processor dynamically based on workload requirements to optimize energy efficiency while maintaining performance.
Let's see limitation of Dynamic Voltage and Frequency Scaling (DVFS) in traditional computer architecture before optimization.
- DVFS is commonly implemented at the operating system level or through hardware features provided by the CPU. The operating system dynamically adjusts the CPU voltage and frequency based on workload characteristics, system power constraints, and thermal considerations. Needs to reduce power consumption during periods of low workload while maintaining adequate performance during periods of high workload.
- DVFS in classic computer architecture is used in a variety of applications, including general-purpose computing tasks, web browsing, office productivity, and multimedia playback. We have to been focused on improving energy efficiency and prolonging battery life in mobile devices, laptops, and servers for better performance.
Considerations for Optimization in case of ML/DL Applications:
- In architectures optimized for ML/DL applications,

DVFS may be implemented at multiple levels of the hardware and software stack, including specialized hardware accelerators (e.g., GPUs, TPUs) and ML/DL frameworks. DVFS techniques are used to dynamically adjust the voltage and frequency of processing units (e.g., CPU cores, GPU cores) based on the computational demands of ML/DL workloads.

- Optimization Goals: DVFS in ML/DL-optimized architectures aims to maximize the performance and efficiency of neural network computations by dynamically adjusting the voltage and frequency of hardware accelerators and processing units. The goal is to balance computational throughput with energy consumption, ensuring optimal performance for ML/DL workloads while minimizing power consumption.
- Applications: DVFS in ML/DL-optimized architectures is used to accelerate model training and inference tasks in neural network frameworks such as TensorFlow, PyTorch, and MXNet. Optimization efforts focus on improving the efficiency of ML/DL workloads on specialized hardware accelerators (e.g., GPUs, TPUs) and maximizing the throughput of neural network computations.
- Some frameworks might integrate with DVFS controls to dynamically adjust processing power based on the specific workload characteristics of the ML/DL task.

Challenges when we go for optimization of DVFS:

- Thermal Constraints: ML/DL workloads can be computationally intensive, leading to heat generation. DVFS can be used to manage thermal throttling by dynamically adjusting voltage and frequency to stay within safe operating temperatures.
- Impact on Accuracy: Very aggressive DVFS settings might introduce slight rounding errors in some ML/DL models. Careful calibration is needed to balance power savings with acceptable accuracy loss.

9. UTILIZING MACHINE LEARNING FOR DESIGN.

The ever-growing complexity of Machine Learning (ML) and Deep Learning (DL) workloads demands continual innovation in computer architecture design. Traditionally, this exploration process has relied on manual design and simulation, which can be time-consuming and computationally expensive. Advanced computer architecture optimization leverages machine learning (ML) for design space exploration, accelerating the discovery of efficient hardware for ML/DL tasks. The design space for ML/DL architectures is vast, encompassing numerous parameters like memory organization, processing unit configurations, and interconnect structures. Manually evaluating all potential configurations is impractical. Simulating the performance of different architectures for various workloads is computationally expensive, hindering the exploration process.

ML techniques offer a powerful approach to address these challenges. By employing techniques like reinforcement learning or evolutionary algorithms, we can automate the exploration of the architecture design space.

These algorithms can:

-Efficiently Search the Design Space: ML models can be trained on historical data of architecture performance and workload characteristics. This allows them to identify promising design configurations and prioritize them for further evaluation through simulation or hardware prototyping.

-Learn from Experience: As the exploration progresses, the ML model can learn from the performance of evaluated architectures. This allows it to refine its search strategy and focus on more promising design directions over time.

-Benefits of ML-driven Exploration: Faster Design Exploration: ML can significantly accelerate the discovery of efficient architectures compared to traditional methods. This allows for quicker adaptation to the evolving demands of ML/DL algorithms.

-Improved Design Quality: By exploring a wider range of configurations, ML can potentially identify more efficient architectures compared to what might be discovered through manual exploration.

-Data-Driven Design Decisions: The exploration process becomes data-driven, with decisions based on performance evaluations and learned relationships between architecture parameters and workload characteristics.

However, utilizing Machine Learning for Design Space Exploration also presents challenges:

-Data Quality: The effectiveness of the exploration process heavily relies on the quality and quantity of data used to train the machine learning models. High-fidelity simulations or real-world performance data are crucial for accurate exploration.

-Interpretability: Understanding the rationale behind the proposed architecture by the machine learning model can be challenging. This limits the ability to refine the exploration process or incorporate domain knowledge from human architects.

10. CASE STUDIES WHERE THESE OPTIMIZATION TECHNIQUES HAVE BEEN SUCCESSFULLY APPLIED

- Google's BERT Model Optimization
BERT model – popular state-of-the-art natural language processing model. The goal was to decrease memory size and make inference faster. Model size decreased by 60%, inference made 1.5x faster. Techniques used: sparsity handling to work with sparse tensor representations more effectively. quantization and compression techniques to reduce precision and memory usage.
- NVIDIA's GPU Optimization for Deep Learning (Vinh Nguyen, 2020)
Techniques employed: Hardware optimization: Accelerators – GPUs, Software Framework optimization: Using CUDA – Compute Unified Device Architecture.
NVIDIA improved its GPU architecture and CUDA software framework to be used effectively in deep learning. It resulted in reduction in training and inference time for deep learning models, more efficient use of the GPU, hence savings. It used accelerated matrix multiplication to exploit the matrix operations that power deep learning. Optimization of the CUDA software framework to provide edge execution of neural network operations.
- Qualcomm's Snapdragon AI Engine used Hardware Optimization: AI accelerators (Hexagon DSP, Adreno GPU). Enabled on-device AI capabilities for tasks such as image recognition, natural language processing, and object detection. Improved power efficiency and reduced latency for AI tasks on mobile devices. Applied Leveraged specialized AI accelerators (Hexagon DSP, Adreno GPU) for efficient execution of neural network computations. Optimized memory access and caching to reduce latency and improve performance.
- Tesla's Autopilot Hardware 3. Used Hardware Optimization - Custom hardware (Tesla FSD Chip). It enabled real-time processing of sensor data and AI-based decision-making for autonomous driving tasks.

11. CONCLUSION AND FUTURE CONSIDERATIONS

In conclusion, computer architecture optimization for machine learning and deep learning applications is essential for realizing the full potential of ML/DL workloads by enabling the execution of computation and neural network computations. As discussed in this paper, there are multiple ways that we optimize computer architecture to meet the needs of ML/DL workloads. These include hardware-level optimizations, software framework enhancements, and other novel architectures for exploration. Hardware optimizations, including the use of GPUs, TPUs, and custom ASICs, and other accelerators, have been widely employed. Researchers have also developed multicore architectures to take advantage of parallelism and have also optimized data movement. Software frameworks have also grown and improved dramatically, enabling the scaling and execution of ML and DL models. Furthermore, the application of machine learning to explore architecture is speeding up the design process, allowing researchers to explore the design space much more quickly and, with the help of few compute resources, identify optimal hardware configurations. With the use of supervised learning, reinforcement learning, and Bayesian optimization (Tony Pourmohamad, 2021), learning-based methods enable researchers to effectively probe the variety of architectural parameters and results, leading to the creation of high-performance, energy-efficient computing systems suited to ML/DL workloads. As machine learning and deep learning applications thrive, and novel technologies and computational strategie s are introduced into sophisticated and varied hardware devices, the field of computer architecture optimization for ML/DL has ample room for growth. In practice, new technologies may include greater exploration of novel hardware architectures and optimized algorithms for ML/DL workloads, as well as incorporating hardware and software optimizations to ensure end-to-end efficiency across heterogeneous hardware. To summarize, computer architecture optimization is critical for the successful rise of machine learning and deep learning applications, allowing an efficient manner to execute ML workloads and DL across a broad array of hardware devices and use settings. By extending existing limits of knowledge in computer architecture optimization (Hennessy, 2017), researchers and engineers play an essential role in addressing the key computational constraints of the future of ML and DL applications and promoting AI's influence in advancement across different fields.
This work has established a solid foundation for progressing in advanced computer architecture optimization for ML/DL research. Hence, here, some promising pathways for future exploration have been highlighted (Gao, 2019):
- Co-optimizing software and hardware for even more effective optimizations.
- Exploring emerging technologies, such as neuromorphic computing, for hardware acceleration potential.
- Developing domain-specific hardware architectures designed for families of ML/DL algorithms' requirements.
- Improving Machine Learning for Design Space Exploration to generate more precise, interpretable outcomes.
- High-level languages, or otherwise modern programming languages (PATTERSON, 2019).
- Security and privacy considerations: Architectures

that incorporate hardware-based security measures to protect against the data privacy and model integrity threats facing modern models.

- Domain-specific architectures: Customizing hardware architectures to the needs of the specialized application domain such as image processing, NLP, recommender systems.

By pursuing these and related directions for further advancements in computer architecture optimization, we can enable the next stage of progress in ML/DL models and maximize their impact across the numerous fields.

References

A.Bienz, L. N. (2021). Modeling Data Movement Performance on Heterogeneous Architectures. *IEEE High Performance Extreme Computing Conference (HPEC)* (pp. 1-7). Waltham, MA, USA: Institute of Electrical and Electronics Engineers Inc.

Abadi, M. B. (2016). TensorFlow: A System for Large-Scale Machine Learning. *12th USENIX Symposium on Operating Systems Design and Implementation* , 265–283.

apache.org. (2024). *APACHE MXNET:A FLEXIBLE AND EFFICIENT LIBRARY FOR DEEP LEARNING*. Retrieved from https://mxnet.apache.org/versions/1.9.1/

Brandon Reagen, R. A.-Y. (2017). Deep Learning for Computer Architects. In P. U. Margaret Martonosi, *Synthesis Lectures on Computer Architecture*. Springer Nature Switzerland.

Contributors. (2017, June 26). *In-Datacenter Performance Analysis of a Tensor Processing Unit*. Retrieved from https://arxiv.org/pdf/1704.04760

eitc.org. (2023, January). *cpu-vs-gpu-vs-tpu*. Retrieved from cpu-vs-gpu-vs-tpu: http://www.eitc.org/research-opportunities/photos1/cpu-vs-gpu-vs-tpu_012023a/image_view_fullscreen

Engineers, I. o. (2019). 25th IEEE International Symposium on High Performance Computer Architecture. *IEEE International Symposium on High Performance Computer Architecture*, p. 734 .

Gao, L. W. (2019). An Overview of Machine Learning in Computer Architecture. *Journal of Computer Science and Technology*, 709–731.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition*. O'Reilly Media.

Goodfellow, I. B. (2016). *Deep Learning*. Cambridge, Massachusetts, London: MIT.

Google. (2017, May 12). *An in-depth look at Google's first Tensor Processing Unit (TPU)*. Retrieved from https://cloud.google.com/blog/u/1/products/ai-machine-learning/an-in-depth-look-at-googles-first-tensor-processing-unit-tpu

Hendrik Borghorst, O. S. (2019). CyPhOS – A Component-Based Cache-Aware Multi-core Operating System. *Architecture of Computing Systems – ARCS 2019* (pp. 171–182). Springer, Cham.

Hennessy, J. L. (2017). *Computer Architecture-A quantitave Aproach* . Morgan Kaufman.

Hwu, W.-m. W. (2015). GPU Computing Gems. *Emerald Edition*.

Jacob, B. K. (2018). Quantization and training of neural networks for efficient integer-arithmetic-only inference. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2704-2713.

Joo-Young Kim, B. K.-H. (2022). *Processing-in-Memory for AI: From Circuits to Systems*. Springer Nature.

Larkin Ridgway Scott, T. C. (2021). *Scientific Parallel Computing*. Princeton University.

Migacz, S. (2024). *Performance Tuning Guide*. Retrieved from https://pytorch.org/tutorials/recipes/recipes/tuning_guide.html

PATTERSON, J. L. (2019). A New Golden for Computer Architecture. *Communications of the ACM*, 48-60.

Ruud Van Der Pas, E. S. (2017). *Using OpenMP-The Next Step: Affinity, Accelerators, Tasking, and SIMD (Scientific and Engineering Computation)*. MIT Press.

Sze, V., Chen, Y.-H., Yang, T.-J., & Emer, J. S. (2017, November 20). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proceedings of the IEEE*, pp. 2295 - 2329.

TensorFlow. (2024). *Theoretical and advanced machine learning with TensorFlow* . Retrieved from https://www.tensorflow.org/resources/learn-ml/theoretical-and-advanced-machine-learning

Tony Pourmohamad, H. K. (2021). *Bayesian Optimization with Application to Computer Experiments*. Springer.

Vinh Nguyen, T. G. (2020, June 18). *Optimizing the Deep Learning Recommendation Model on NVIDIA GPUs*. Retrieved from https://developer.nvidia.com/blog/optimizing-dlrm-on-nvidia-gpus/

Wijtvliet, M. W. (2019). Accelerating Machine Learning Workloads with OpenCL on FPGAs. *IEEE International Conference on Cluster Computing*.

Zoran Jakšić, N. C. (2020). A highly parameterizable framework for Conditional Restricted Boltzmann Machine based workloads accelerated with FPGAs and OpenCL. *Elseiver*, 201-211.

Glossary

Term	Definition
Machine Learning (ML)	A subfield of artificial intelligence (AI) that uses statistical techniques to give computers the ability to learn from data without being explicitly programmed.
Deep Learning (DL)	A subset of machine learning where artificial neural networks, algorithms inspired by the human brain, learn from large amounts of data.
Computer Architecture Optimization	Techniques aimed at enhancing computer systems to better accommodate the computational demands and data requirements of machine learning and deep learning algorithms.
CPUs (Central Processing Unit)	The primary component of a computer that performs most of the processing inside a computer.
GPUs (Graphical Processing Unit)	A specialized electronic circuit designed to accelerate the creation of images in a frame buffer intended for output to a display.
TPUs (Tensor Processing Unit)	An application-specific integrated circuit (ASIC) developed by Google specifically for accelerating machine learning workloads.
Parallelism in Multicore Architectures	Utilizing multiple cores within a CPU to execute multiple processes simultaneously, thereby increasing computational efficiency.
Data Movement Optimization	Techniques aimed at improving the movement of data within a computer system, including caching, addressing sparsity, compression, and quantization.
Sparsity	A property of large data sets in which only a small percentage of the data is non-zero.
Compression	Techniques used to reduce the size of data, which can lead to improved storage and processing efficiency.
Quantization	The process of reducing the precision of a numerical representation, which can lead to reduced memory and computational requirements.
Software Framework Optimization	Enhancements made to software frameworks such as TensorFlow, PyTorch, and Apache MXNet to improve their performance and efficiency in handling machine learning and deep learning workloads
DVFS (Dynamic Voltage and Frequency Scaling)	A technique used in computer systems to optimize power consumption by adjusting the voltage and frequency of the CPU according to the current workload.
Runtime Systems	Systems optimized to run machine learning and deep learning workloads on various hardware platforms.

Supervised Learning	A type of machine learning where the algorithm learns from labeled data, making predictions and decisions based on that data.
Reinforcement Learning	A type of machine learning where an agent learns to make decisions by trial and error, receiving feedback in the form of rewards or penalties.
Bayesian Optimization	A method for optimizing objective functions that are expensive to evaluate, by building a probabilistic model of the objective function it to select the most promising points to evaluate.
ASICs (Application-Specific Integrated Circuits)	Customized microchips designed for a particular application, such as machine learning or deep learning.
Edge Devices	Devices that perform data processing at or near the source of data generation, rather than relying on a centralized data-processing warehouse or cloud storage.
API (Application Programming Interface)	A set of protocols, tools, and definitions that allow different software applications to communicate with each other.
Hardware Optimization	Techniques used to enhance the performance and efficiency of computing systems.
Accelerators	Specialized hardware components designed to handle heavy workloads more efficiently than traditional CPUs.
FPGA (Field-Programmable Gate Arrays)	Customizable integrated circuits that can be reprogrammed to implement different hardware functionalities after manufacturing.
Vectorization	Utilizing vector processing units such as SSE (Streaming SIMD Extensions) and AVX (Advanced Vector Extensions) in CPUs and GPUs to perform operations on more than one data simultaneously.
Multicore Architectures	Architectures that enable multiple processing cores on the same chip, allowing for parallel processing and improved performance for ML/DL tasks.
Data Prefetching	A technique used to improve memory access latency by predicting and loading data into cache before it is actually needed.
AVX (Advanced Vector Extensions)	Is a set of instructions that extend the capabilities of x86 processors from Intel and AMD. It allows them to perform operations on multiple pieces of data simultaneously, significantly improving performance for specific workloads.

REVOLUTIONIZING ENTERPRISE RESOURCE PLANNING: A TRANSFORMATION JOURNEY IN THE ALBANIAN TELECOMMUNICATION INDUSTRY

Eralda Caushi [1], Klejda Caushi [2], Malvina Halilaj [3]

[1] Department Of Software Engineering, Faculty of Engineering, Canadian Institute of Technology , Albania, eralda.caushi@cit.edu.al, ORCID: 0009-0004-8667-5320

[2] Department of Economics and Business, Faculty of Economics, Political Sciences, Catholic University Our Lady of Good Councel , Tirana, Albania, kcaushi@unizkm.al, ORCID: 0009-0003-4163-3322

[3] Department of Computer Science, Faculty of Natural Sciences, University of Tirana , Albania, malvina.halilaj@fshn.edu.al, ORCID: 0009-0001-1981-128X

Abstract

This paper explores the critical aspects of ERP (Enterprise Resource Planning) transformation in the Albanian telecommunications industry. The focus is on SAP tools that include asset modules, the platform for managing ERP stock and inventory, and a new frontline platform for selling device and accessories for all the channels by supporting both mobile and fixed devices. The telecom industry has evolved significantly, and the need for efficient operations, better integration, enhanced security, and improved user experiences has become dominant. Cloudification of the SAP solution running in GCP is part of the revolution. The technical implementation is strongly related to the advantages of an ERP transformation within the telecommunications industry. These benefits include enhanced efficiency and productivity, seamless integration, and impeccable data accuracy, bolstered security and compliance, and an overall enhancement in user experiences. Furthermore, this implementation also navigates the challenges and critical considerations essential in this transformative journey. Re-usability is a key component since the successful implementation in a single market for a single local market, can be replicated in the other markets as well by justifying the high implementation cost of 3 million Euros and the involvement of 32 local stakeholder and 45 global stakeholders.

Keywords: ERP, cloudification, SAP, GCP, telecommunications.

I. INTRODUCTION

This paper explores the critical aspects of ERP (Enterprise Resource Planning) transformation within the telecommunications industry, based on the experience of the authors that have led the implantation in Vodafone Albania and on joint research activities and support. The technical implementation in the context of an ERP transformation within a telecommunications company offers a multitude of advantages, with SAP in the cloud playing a pivotal role. These benefits include heightened efficiency and productivity, seamless integration leading to enhanced data accuracy, fortified security measures ensuring regulatory compliance, and an overall enhancement of user experiences. SAP in the cloud not only brings about these improvements but also addresses the complexities and factors that are integral to this transformative process.

Cloudification is another important aspect of the implementation with huge contributions to streamlining operations and supporting the organization's agility and scalability. The rationale for this transformation is driven by the need to streamline complex operations, reduce manual tasks, optimize resource allocation, and enhance customer experiences. However, it comes with challenges, including the complexities of integrating ERP systems with extensive network infrastructures, navigating strict regulatory requirements, addressing data security concerns, and managing employee resistance to change. Despite these challenges, ERP transformation presents significant opportunities, including streamlined operations, data-driven decision-making, innovation, and improved customer satisfaction, making it a vital step for telecom companies looking to thrive in a dynamic and highly competitive industry.

*Corresponding author:

SAP (Systems, Applications, and Products) as a Leading Tool for the ERP Transformation in Telecommunications plays a pivotal role in ERP (Enterprise Resource Planning) transformation within the telecommunications industry, providing a comprehensive suite of tools and solutions that facilitate various aspects of the transformation process. Two crucial components of SAP's offerings, the asset modules and inventory management platforms, are instrumental in enhancing efficiency, improving inventory management, and ultimately delivering better sales and customer service experiences. Here, we will elaborate on their roles and provide specific examples and case studies to highlight the benefits of leveraging SAP's solutions. SAP's asset modules are essential for telecom companies as they help manage and optimize their extensive network infrastructure. By integrating the asset modules into their ERP system, telecom companies gain real-time visibility into their infrastructure, enabling them to optimize maintenance, to proper resource allocation, to monitor the performance, to optimized stock levels and improve the order fulfillment. (SAP SE Corporate, 2023).

Telecoms can schedule preventive maintenance more effectively, reducing downtime and ensuring the network's reliability. There is the possibility to allocate resources based on asset health, location, and priority, minimizing operational costs.

Real-time data allows telecoms to monitor asset performance and proactively address issues before they impact services. Below a print screen on the monitoring interface.

Sl. No	Interface ID	Middleware	Successful Messages	Failure Messages	Total Messages
1	MSISDN_SIMDeliveryConfirmation	GCP - APIGEE	0	0	0
2	EquipmentDetailsStandard	GCP - APIGEE	31	0	31
3	POSSales	GCP - APIGEE	5289	0	5289
4	CustomerReturn	GCP - APIGEE	0	0	0
5	CustomerMaster	GCP - APIGEE	38978	32775	71753
6	POSProducts	GCP - APIGEE	3353	0	3353
7	POSPriceQuery	GCP - APIGEE	4500	0	4500
8	ArticleStock	GCP - APIGEE	4495	5	4500
9	ReservationDocument	GCP - APIGEE	4	0	4
10	BillingDocumentDetails	GCP - APIGEE	114	0	114
11	BillOfMaterial	GCP - APIGEE	147	0	147
12	InspectionLogs	GCP - APIGEE	0	0	0
13	JournalEntry	GCP - APIGEE	1	0	1

PFB comments for interface report shared today:

- Customer Master API: Errors are due to incorrect payload. Issues already highlighted to CRM team. Fix will be implemented from CRM side
- Article Stock : Errors are due to incorrect plant value in payload

No Action required from SAP Side to fix these errors

Figure 1. *Interface Monitoring Report for Vodafone*

SAP's inventory management platforms cater to businesses of different sizes and helps manage and streamline fulfillment and product storage. (SAP, Inc, 2023)

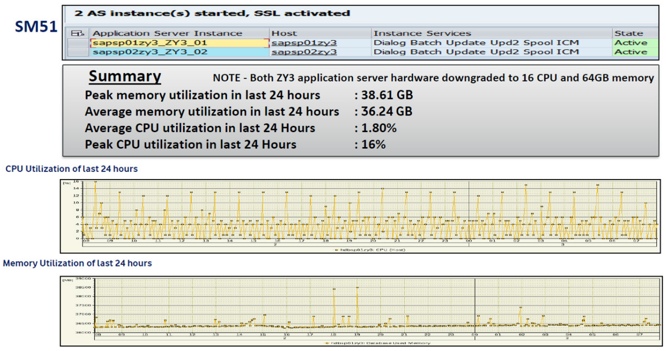


Figure 2. *ZY3-HDB and ABAP Monitoring Report for Vodafone*

It enables telecom companies to streamline their supply chain and inventory control processes. These platforms offer robust features for tracking, managing, and optimizing inventory levels, leading to more efficient operations. Telecom companies can benefit in the following ways: (SAP, Inc, 2023). By using demand forecasting and real-time inventory tracking, companies can maintain optimal stock levels, reducing carrying costs and minimizing stockouts.

Efficient inventory management ensures that customer orders are fulfilled promptly, enhancing the overall customer experience. Cost reduction is a benefit and specifically reduced carrying costs and waste translate into significant cost savings.

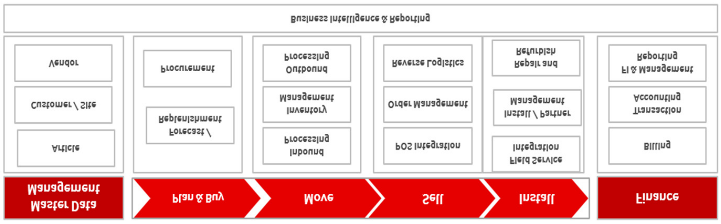


Figure 3. *Impacted Areas: Financial, Procurement and Sales*

Case Study: Verizon, another leading telecommunications company, integrated SAP's inventory management platform into its ERP system. This allowed Verizon to achieve a 15% reduction in carrying costs and a 20% improvement in order fulfillment rates, resulting in increased customer satisfaction and higher profitability.

II. CLOUDIFICATION OF SAP SOLUTIONS

Cloudification means more than just 'moving towards the cloud'. It also refers to the immense business opportunity that follows when you transform to a new business model, enabled by cloud computing and other technologies (Gyseling, 2023) s. In our case, the Cloudification, the process of moving SAP solutions to the cloud, holds significant importance in the telecommunications industry, particularly when considering its integration with ERP systems. In this context, cloud deployment, such as on the Google Cloud Platform (GCP), offers substantial benefits that enhance the agility and scalability of SAP solutions, ultimately boosting the efficiency and competitiveness of telecommunications companies. In general, the typical problems that cloud computing promises to solve are as follows: concerns about the total cost of ownership of information technology services, the scalability / elasticity of the IT solution, the rigidity of investment timing, and agility. (Bernus, 2016). Cloud deployment of SAP solutions, such as on GCP, provides telecommunications companies with the agility to adapt to rapidly changing market conditions. In this highly dynamic industry, telecoms need the ability to scale resources up or down as needed. This flexibility is invaluable for staying competitive in an environment where customer demands, and technology trends evolve rapidly. The scalability of SAP in the cloud is crucial for telecommunications companies with fluctuating workloads. They often experience spikes in network traffic, especially during special events or product launches. Running SAP in the cloud allows for easy scalability, ensuring that systems can handle increased loads without the need for substantial upfront investments in hardware. This dynamic scalability is essential for ensuring a seamless customer experience and preventing service disruptions during peak usage periods.

Traditional on-premises SAP deployments require substantial investments in data centers, servers, and networking equipment. Cloud deployment, on the other hand, significantly reduces these infrastructure costs. Telecommunications companies can shift from a capital-intensive CAPEX model to an operational expenditure (OpEx) model, paying only for the resources they consume.

Accessibility is important in the telecommunications sector, where employees often need to access critical data and applications remotely. Cloud-based SAP systems ensure that employees can work from anywhere, improving collaboration, responsiveness, and decision-making. This accessibility also supports business continuity, as data and applications are available even in the face of unforeseen disruptions, such as natural disasters or pandemics. Below we showcase the APIGEE Traffic Monitoring in Cloud for Vodafone:

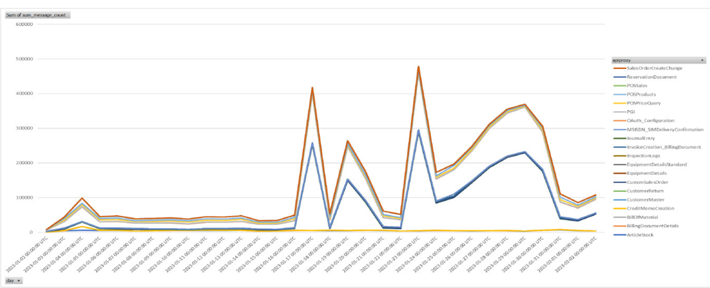


Figure 4. *Apigee Traffic Cumulative Count Report as on 3/2/2023*

Case Study: T-Mobile USA, one of the largest telecom operators in the United States, embraced cloudification by deploying SAP solutions on the Google Cloud Platform. This move allowed T-Mobile to scale its infrastructure to meet surges in customer demand during product launches and promotional events. A 20% improvement in operational efficiency and significant reduction in infrastructure costs.

III. TECHNICAL IMPLEMENTATION

There are a lot of technical challenges that are faced during every implementation, especially due to the complex infrastructures that exist in telecommunication companies. Addressing these challenges is crucial for the successful deployment and operation of ERP systems in this context. Telecommunications companies operate in a wide and elaborated network infrastructures to provide services to customers.

Integrating SAP-based ERP systems with these networks can be very challenging. ERP systems need to interact seamlessly with customer relationship management systems, billing systems and more. The below figure shows an example of an architectural view implemented in the telecommunication.

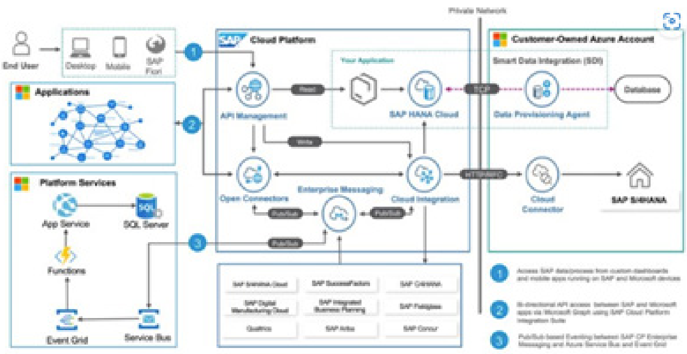


Figure 5. *Local Market architecture reference*

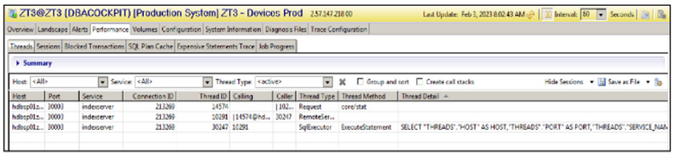
SAP Cloud Platform

Data accuracy is important in the telecommunications sector, where billing, service provisioning, and network management depend on precise and up-to-date information. Integrating ERP systems with existing data sources and ensuring the consistency and accuracy of data can be technically challenging.

Data synchronization, data cleansing, and data validation processes are necessary to prevent discrepancies, billing errors, and service interruptions. Telecom companies handle sensitive customer data and proprietary network information, making data security a top priority. The integration of SAP-based ERP systems introduces new potential vulnerabilities. Ensuring data security during the exchange of information between different systems and within the ERP system itself is a complex task.

Robust encryption, access controls, and regular security audits are vital to protect against data breaches and unauthorized access. (Saharia, Koch, & Robert Tucker, 2008)

Threads:



Blocked Transaction:

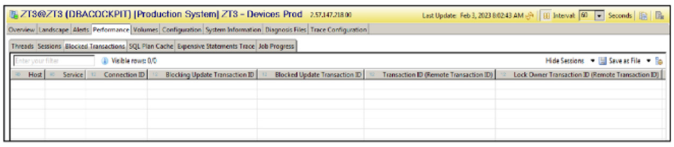


Figure 6. *Threads and Blocked Transactions Report for Vodafone*

Also, telecom companies typically have a variety of legacy systems, including network management tools, CRM systems, and billing platforms. Ensuring the compatibility of the new ERP system with these existing systems is a critical technical challenge. It often involves developing custom interfaces or middleware to facilitate data exchange and process synchronization. Compatibility issues can lead to operational inefficiencies and data discrepancies if not addressed adequately.

Data migration from legacy systems to the new ERP platform is a technical challenge. It involves extracting, transforming, and loading large volumes of data while maintaining data integrity and consistency. This process requires accurate planning and testing to ensure a smooth transition without data loss or corruption. Failure in data migration can lead to operational disruptions and costly data recovery efforts.

IV. OVERCOMING IMPLEMENTATION CHALLENGES

To overcome the technical implementation challenges in deploying SAP-based ERP systems in the telecommunications industry, telecom companies can employ a combination of strategies and best practices.

Here are some insights into how to mitigate complexities, ensure data security, and effectively manage data migration, along with real-world examples of successful implementation strategies. (Slevin & Pinto, 1986)

First, there is the need to comprehensive analysis of your network infrastructure. Identify key data sources, processes, and systems that need integration with the ERP. This analysis will help the implementation teams to understand the scope of the integration and prioritize critical components.

There is the need to implement an integration middleware. Using integration middleware, such as API gateways, to facilitate communication between ERP and network systems. This middleware simplifies the interaction between various applications and ensures data consistency. Implementing data validation and cleansing processes can help ensure data accuracy during integration. Develop data quality rules and automated validation scripts to detect and correct inconsistencies. Establishing data governance policies and practices to maintain data consistency across systems. Assigning data stewards responsible for data quality and accuracy and enforce data standards and protocols, can be an added value to the successful implementation.

Encryption and Access Controls are crucial, so teams need to pay particular attention to the implementation of a robust encryption protocols to protect data in transit and at rest. Utilize as well, access controls and role-based permissions to limit access to sensitive information. Encrypting all sensitive customer data is suggested.

The best-in-class guidelines are to conduct regular security audits and vulnerability assessments to identify and mitigate potential security risks. There is always a strong suggestion to take ahead security measures which are up to date with industry standards and best practices. Regular penetration testing and security audits were conducted to detect vulnerabilities and ensure data remained secure.

Create a detailed data mapping and transformation plan as well that outlines how data will be migrated from legacy systems to the new ERP. Use ETL (Extract, Transform, Load) tools to automate data migration. Extensive testing of data migration processes is crucial. Test data integrity and accuracy before, during, and after migration. Establish rollback procedures in case of unexpected issues.

Overcoming technical challenges in implementing SAP-based ERP systems in the telecommunications industry requires a strategic and systematic approach. These strategies can contribute to a smoother ERP transformation process, ultimately leading to improved operational efficiency and customer satisfaction.

V. HEIGHTENED OPERATIONAL EFFICIENCY AND PRODUCTIVITY

Telecom companies have realized significant efficiency gains through ERP transformation, particularly when utilizing SAP solutions deployed in the cloud. These improvements are evident in key performance indicators (KPIs).

Some important KPIs based the very first weeks after the go live:

- o 35,000 API calls per day on average
- o 823 transactions customer faced on average per day.

The above KPIs can be translate in reduction in manual tasks with streamlined processes and automation, telecoms reduce the need for manual data entry and processing, resulting in a decrease in errors and delays. Faster decision-making especially when cloud-based, provide real-time access to data and analytics. Below a report that shows the 30 days KPIs

Daily report as of 03.02.2023					
Business area	Business Transactions	Daily(02/02/2023)		Cumulative (from 11th December)	
		# of transaction	Quantity	# of transaction	Quantity
Customer faced	Sales Orders	33	6281	1,186	87,423
	Sales Goods Issues	121	5192	6,659	89,464
	Dealer SO Reversals	0	0	1	1
	Advanced Return Management	0	0	19	35
	Sale Invoices for sale from Dealer	77	1014	4,582	28,034
	Reversal Invoices for sales	0	0	24	166
	Sale Invoices for sale from Own Shops	104	135	6,158	7,877
	Reversal Invoices for sales	1	1	80	117
Warehouse Activity	Purchase Orders	0	0	12	19,228
	Good Receipts	0	0	139	393,341
	Stock Transfer Orders (WH -> Dealers)	13	6685	142	51,880
	Stock Transfer Orders (WH -> Own shops)	0	0	10	20
	Storage Location to Storage Location (Installers' stock up)	12	229	492	20,527

	Category	Value (YTD)
Financial	Revenue Posting / Payment Order / Sales Order(Credit)	688,595,113 ALL
	Purchase Orders	135,249,885 ALL
	Good Receipts	169,963,921 ALL

Figure 7. *Key Business Process Execution Report for Vodafone*

Resource optimization is one of the key benefits. The ability to scale resources up or down in the cloud optimizes cost management, reducing operational overhead and ensuring that resources are used efficiently.

The implementation of ERP solutions, especially in the cloud, significantly enhances data accuracy and security and operational monitoring.

ERP systems incorporate data validation rules and automated checks, minimizing data entry errors. The Interface monitoring is a great support to minimize and act on the flows that show inconsistencies.

Interface ID	Message type	Sender System (S4 Retail-ZT3)			Receiver System (S4- ZE1)		
		Success	Failure	Pending	Success	Failure	Pending
AL-FIN-INT-008 (G/L Interface from Alex to EVO)	ACC_DOCUMENT	386	0	0	367	19	0
AL-SCM-INT-005 (GR from Alex to EVO)	WMMBXY	28	0	0	27	1	0
AL-SCM-INT-006 (Asset intf from Alex to EVO)	MBGMCR	260	0	0	253	7	0

Interface ID	Message type	Sender System (S4 -ZE1)			Receiver System (S4 retail- ZT3)		
		Success	Failure	Pending	Success	Failure	Pending
AL-FIN-INT-009 (vendor invoice from EVO to Alex)	Z_MM_INVOICE	0	0	0	0	0	0
AL-SCM-INT-007 (PO from EVO to Alex create)	ORDERS	0	0	0	0	0	0
AL-SCM-INT-008 (PO from EVO to Alex change)	ORDERS	0	0	0	0	0	0
AL-FIN-INT-012 (Asset info from EVO to Alex)	Z_ASSET_MSGTYP	1	0	0	1	0	0

Interface ID	Message type	POS	Receiver System (S4 retail- ZT3)		
			Success	Failure	Pending
CBM-SCM-INT-816 (POS information to ALEX)	WPUBON		191	39	0

Interface ID	Message type	Sender System (MDG-ZM4)			Receiver System (S4 retail- ZT3)		
		Success	Failure	Pending	Success	Failure	Pending
AL-MDG-INT-001/CBM-MDG-INT-045 (Article master)	ARTMAS	12	0	0	11	1	0

Figure 8. *ZT3 Module Interface Report for Vodafone*

Cumulative Count	WK50	WK51	WK52	WK01	WK02	WK03	WK04	WK05
Successful IDocs	8157	12545	14533	44466	4993	11175	9196	6863
Failed IDocs	0	6	7	5	4	13	10	65
To be Dispatched	0	0	0	0	0	0	0	0
Total Count	8157	12551	14540	44411	5365	11188	9206	6928

Figure 9. *Failed & Successful IDoc Report for Vodafone*

Cloud platforms provide robust encryption, securing data at rest and in transit. As with any form of data encryption, cloud encryption renders the information indecipherable and therefore useless without the encryption keys. This applies even if the data is lost, stolen, or shared with an unauthorized user. (Puzas, 2022). Access control is the selective restriction of access for an individual or entity to a physical location or computer systems, networks, files, and data. (Sangfor Technologies, 2023) Role-based access controls limit data access to authorized personnel only.

Case Study: Orange, a global telecommunications provider, implemented SAP ERP in the cloud, leading to a 90% reduction in data entry errors. The company also experienced zero data breaches since implementing advanced security measures.

Case Study: Vodafone Group, a global telecommunications company, ensured data security during its SAP ERP implementation by encrypting all sensitive customer data. Regular penetration testing and security audits were conducted to detect vulnerabilities and ensure data remained secure.

VI. CONCLUSIONS

ERP transformation in the telecommunications industry is a critical step towards achieving operational efficiency, meeting customer expectations, and staying competitive in a rapidly evolving market. SAP tools, asset modules, mobile and fixed device support, CRM integrations, and cybersecurity principles all play a crucial role in this transformation. This paper aims to provide insights and guidance for telecom companies considering or undergoing ERP transformation, based on the successful implementation of the erp transformation leaded by the main author. ERP transformation, especially with SAP solutions in the cloud, has a positive impact on user experiences and customer service. Cloud-based SAP systems seamlessly integrate with various customer-facing systems, improving the accuracy and timeliness of customer interactions. Real-time data and analytics provide insights that enable personalized customer interactions and better service. Enhanced ERP capabilities often include self-service portals, empowering customers to manage their services efficiently.

In summary, SAP's asset modules and inventory management platforms are instrumental in the ERP transformation within the telecommunications industry. They enable telecom companies to optimize their network infrastructure, streamline inventory management, and enhance their overall operational efficiency. Leveraging SAP's solutions empowers telecom companies to stay competitive in a rapidly evolving industry while delivering better services to their customers. The significance of cloudification in the telecommunications industry cannot be overstated. Deploying SAP solutions in the cloud, such as on the Google Cloud Platform, empowers telecom companies with enhanced agility, scalability, reduced infrastructure costs, and improved accessibility. These advantages are particularly critical in a sector where rapid responses to changing customer needs, cost efficiency, and accessibility are vital for success. By leveraging the cloud implementations, the telecommunications companies have position themselves for sustained growth and competitive edge in an ever-evolving industry. The technical challenges faced by telecommunications companies during the implementation of SAP-based ERP systems are complex. They incorporate the integration with complex network infrastructures, ensuring data accuracy, implementing robust data security measures, managing data migration, and addressing compatibility with existing systems. Overcoming these challenges is essential to unlock the benefits of ERP systems and to streamline operations in this data-driven and highly competitive industry. Careful planning, thorough testing, and continuous monitoring are key elements in successfully navigating these challenges.

To overcome the challenges in the implementing SAP-based ERP systems in the telecommunications industry requires a strategic and systematic approach. Telecom companies can mitigate complexities by starting with a thorough analysis, ensuring data accuracy through validation and governance, enhancing data security through encryption and access controls, and managing data migration with careful planning and testing. Real-world case studies demonstrate how these strategies can contribute to a smoother ERP transformation process, ultimately leading to improved operational efficiency and customer satisfaction.

In conclusion, the tangible benefits of ERP transformation in the telecommunications sector are substantial, and they are further amplified when SAP solutions are deployed in the cloud. Telecom companies experience heightened operational efficiency, improved data accuracy and security, and enhanced user experiences, ultimately leading to improved customer service and satisfaction. Real-world examples, such as those from Vodafone, Orange, T&T demonstrate how these transformations have translated into concrete improvements in KPIs and customer interactions, making ERP transformation a valuable investment in this competitive industry.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the colleagues in Vodafone Albania, Vodafone Group and VOIS India for their dedication and their hard work for achieving every single milestone till the successful journey. Their expertise has greatly enriched this work. We would also like to express our heartfelt appreciation to our co-authors, Malvina Halilaj, who served as a technical lead, and Klejda Caushi, a diligent researcher. Their tireless collaboration across all facets of this endeavor, from research to implementation, has been indispensable. Klejda' s technical ability and research skills have played a pivotal role in achieving the objectives of this project. Their contributions have not only strengthened the quality of our work but have also propelled us towards success in our ERP transformation endeavors within the telecommunications industry. We are immensely grateful for their invaluable contributions and unwavering support throughout this journey.

REFERENCES

[1] Ali, M. M. (2017). ERP System Implementation in Large Enterprises-A Systematic Literature Review. Journal of Enterprise Information Management.
[2] Anaya, L. (2019.). To what extent is it viable to apply benefits management approach for ERP systems? Procedia Computer Science 164, 33–38.
[3] Anaya, L. A., & Dag Håkon Olsen. (September 2014). Implementing ERP in a Challenging Environment: The Case of a Palestinian Telecom Company.
[4] AT&T Global field marketing. (2012, May). Business Continuity. AT& newsletter issue. Retrieved from https://www.corp.att.com/wp-content/assets/content/worldwide/11/White_Paper_Business_Continuity.pdf
[5] Bahssas, D. M. (2015). Enterprise Resource Planning (ERP) Systems: Design,Trends and Deployment. The International Technology.
[6] Benedict Bender, C. B. (2021). Future ERP Systems: A Research Agenda. 23rd International Conference on Enterprise Information Systems (ICEIS 2021) - Volume 2,, pages 776-783.
[7] Bernus, O. N. (2016). Business Cloudification: An Enterprise Architecture Perspective. Proceedings of the 19th International Conference , Vol. 3, pp.353-360.
[8] Chen, C. (2001). Planning for ERP Systems: Analysis and Future Trend. Business Process Management Journal, 374-386.
[9] Eko Wahyu Tyas Darmaningrat, F. A. (2019). Communication Management Plan of ERP Implementation. The Fifth Information Systems International Conference .

[10] Fadi Taher Qutaishat1, S. A. (2012). The Effect of ERP Successful Implementation on Employees' Productivity, Service Quality and Innovation: An Empirical Study in Telecommunication Sector in Jordan. International Journal of Business and Management .

[11] Gyseling, P. (2023). Digital Vision for Cloudification. Retrieved from ATOS: https://atos.net/en-be/belgium/digital-vision-cloudification/cloudification-enables-businesses-to-become-truly-digital-and-agile

[12] Huang, S.-M. I.-C.-H.-T. (2004). Assessing Risk in ERP Projects: Identify and Prioritize The Factors. Industrial Management & Data Systems, 681-688.

[13] LEYH, C. (2011). hy do Companies Implement ERP Systems? - The Goals and Reasons behind ERP Implementation Projects. In R. P. P.-M. Léger, Readings on Enterprise Resource Planning (pp. Chapter 02, pp. 19-35.). Montreal: ERPsim Lab, HEC Montreal.

[14] LEYH, C. (2011). Tailoring of ERP Systems. In R. P. P.-M. Léger, Readings on Enterprise Resource Planning (pp. Chapter 06, pp.78-90). Montreal: ERPsim Lab, HEC Montreal.

[15] Maguire, S., Udechukwu Ojiako, & Al Said. (2010, February). ERP implementation in Omantel: a case study. Industrial Management & Data Systems.

[16] Mathiassen, L., & Carsten , S. (2020). Managing New Tools.

[17] Matrix Web Technologies. (2022). Matrix ERP System Documentation. Retrieved from Matrix ERP system : www.matrixwebtech.com

[18] McKay, J. P. (2014). Rethinking Communication in IT Project Management. 47th Hawaii International. Hawaii.

[19] Microsoft . (2023). What is access control? Retrieved from Microsoft: https://www.microsoft.com/en/security/business/security-101/what-is-access-control#:~:text=Access%20control%20is%20an%20essential,control%20policies%20protect%20digital%20spaces.

[20] Puzas, D. (2022). Cloud Encryption. Retrieved from crowdstrike: https://www.crowdstrike.com/cybersecurity-101/cloud-security/cloud-encryption/

[21] Saharia, A., Koch, B., & Robert Tucker. (2008). ERP SYSTEMS AND INTERNAL AUDIT. Issues in Information Systems VOL IX, No. 2, 578-586.

[22] Sangfor Technologies. (2023, October 17). What is Access Control and How Does It Work? Retrieved from Sangfor - Cyber Security: https://www.sangfor.com/glossary/cybersecurity/what-is-access-control

[23] SAP SE Corporate. (2023, May 11). Enterprise Resource Planning (ERP). Retrieved from SAP SE: https://www.sap.com/products/erp.html#benefits

[24] SAP, Inc. (2022, 12). What are the Best Practices for ERP Implementation? Retrieved from SAP: https://www.sap.com/sea/products/erp/what-is-erp/erp-implementation-best-practices.html

[25] SAP, Inc. (2023). SAP Inventory Management. Retrieved from Selected Hub: https://www.selecthub.com/p/inventory-management-software/sap-inventory-management/#:~:text=SAP%20Inventory%20Management%20caters%20to,throughout%20an%20entire%20supply%20chain.

[26] SAP, Inc. (2023, May 11). SAP Inventory Management. Retrieved from SAP SE: https://help.sap.com/docs/SAP_Inventory_Manager

[27] Slevin, D. P., & Pinto, J. K. (1986, September). The Project Implementation Profile: New Tool For Project Managers.

[28] Zulch. (2014)). Leadership Communication in Project Management. Procedia - Social and Behavioral Sciences, 172 – 181.

crj.cit.edu.al

CIT REVIEW JOURNAL

C R J

PUBLISHER
Canadian Institute of Technology
Tirana, Albania