# Applied Business

## Issues & Solutions

Applied Business: Issues & Solutions, 1(2022)
ISSN 2783-6967
https://doi.org/10.57005/ab.2022.1

## CONTENT

*Article*

# The Impact of Foreign Direct Investment in the High-Tech Sector Economy

Olga Nosova[1,2*]
[1] Department of Business, Vilnius Business College, Kalvarijų g. 129-401, LT-08221 Vilnius, Lithuania
[2] Department of Marketing, Management and Entrepreneurship, V. N. Karazin Kharkiv National University, Svobody Square, 4, Kharkiv, 61000 Ukraina
[*] Corresponding author, e-mail: *olgano59@gmail.com*

Received: 01 May 2022
Accepted: 11 June 2022
Online: 31 August 2022

**Abstract.** The article is aimed at studying the impact of FDI on the development of the high-tech sector of the economy. R&D effects are analyzed through the location of branches of foreign enterprises. The positive and negative technological spillover effects and FDI inflow consequences are considered. Foreign investment is assessed as a factor of economic development, modernization, income, and employment growth. The radical innovations' implementation by the innovative TNCs have been highlighted. The assessment of the success of the most innovative companies confirms the usage of artificial intelligence, platforms, and ecosystems.

**JEL:** M20, M21, M29.

**Keywords:** high-tech sector; innovations; transnational corporation; foreign direct investment.

**Citation:** Olga Nosova (2022) The Impact of Foreign Direct Investment in the High-Tech Sector Economy. – *Applied Business: Issues & Solutions* 1(2022)3–9 – ISSN 2783-6967. https://doi.org/10.57005/ab.2022.1.1

## Introduction

The modern development of the world economy is accompanied by processes of internationalization and transnationalization of capital, transfer of new technologies, digitalization, and increasing the use of skilled labor. Factors that stimulate the development of globalization include progress in the field of communications, cultural exchange and equalization of values, development of transport infrastructure, interstate competition for resources, and development of the global financial and stock market.

Foreign direct investment (FDI) is seen as an important tool for attracting foreign capital to the country, contributing to structural reforms and the development of scientific and technological capacity in the country. Direct investment is used by transnational corporations (TNCs) to accommodate knowledge-intensive production and apply modern technology. Companies maximize the benefits of attracting FDI under favorable investment policies.

Factors that negatively affect the inflow of foreign capital into the country include consumer uncertainty, exchange rate volatility, the impact of inflation expectations, and geopolitical instability in the world. The trend of TNC development, which is accompanied by increasing the economic potential of the company by increasing mergers and acquisitions, was positively assessed.

Recent studies have highlighted the considerable research efforts in developing an understanding of the subject of the investment motivation of FDI spillover effects, its impact on economic growth, competitive advantages inside the developed economies, and draw attention to short-term adjustment problems rather than to the long-term possibilities. Empirical studies of FDI spillover effects on domestic firms reflect various factors, conditions, and characteristics of the firm, at industry and national levels. The COVID-19 crisis negatively affected the movement of international capital. In the first half of 2020, global FDI flows fell by 50% compared with the last half of 2019, to $364 billion, because of the pandemic and the resulting supply disruptions, demand contractions, and pessimistic outlook of economic actors [1].

Although the COVID-19 pandemic has led to severe restrictions in developed countries, the economic turmoil is now entirely global. However, although developed countries use the full range of macroeconomic instruments to mitigate the effects, developing countries have a lack of money or fiscal capacity to reduce the effects of the pandemic. Countries' export earnings are declining and access to external sources of finance is declining, while the internal response to health threats will lead to a reduction in tax revenues, which are generally insufficient. The decline in commodity prices, in the market of iron ore, ferrous metals, and grain market is seen as a factor in reducing Ukrainian exports.

To rebuild the economy after COVID-19, the world will need a significant influx of investment resources. Governments need to be provided with favorable conditions for attracting and retaining productive investment and, more importantly, ensuring that the benefits of their use are maximized. To increase the inflow of foreign capital, countries need to reconsider approaches to creating attractive conditions for TNCs to locate new innovative branches, as well as to develop new strategies to attract foreign investors.

The purpose of the article is to study the FDI influence on the high-tech sector economy, R&D effects through the placement of branches of foreign enterprises, advanced technologies application, the spread of new forms of management organization, and increase of the welfare state.

The methodology of the research is the use of the Comparative Economics approach. It allows making comparisons of the main research methods and strategies of FDI attraction in stimulating innovation-science-intensive branches development, analysis of forms, strategies of TNCs activity, and foreign affiliates branches allocation.

## 1. Literature review

The study of the considerable amount of works devoted to the study of FDI impact on economic development in countries indicates the existence of a variety of theories and approaches. Technology transfer is an important aspect of the TNC's presence manifested through vertical connections. Technology transfer and dissemination and deployment are carried out in four interconnected channels: through vertical links with suppliers or buyers in recipient countries, through horizontal links with competing or complementary compa-

nies within the same industry, through migration of skilled workers, and the internationalization of R&D.

The study of the overall effect of technology transfer through FDI attraction, the influence of FDI inflow on the economic growth of the state, international investments, technologies in the conditions of a dynamic environment, innovations as a factor of competitiveness of TNCs are devoted to the significant number of scientific works.

Barell & Pain [2] investigate the role of FDI in the diffusion and assimilation of technologies and ideas across borders. They determine the consequences of FDI inflow on the host country's technical progress and export performances.

Kathuria [3] points out that domestic firms will not benefit from a foreign presence if it is measured by the share of sales at the same time, they get access to foreign capital reserves. Additional research shows that domestic firms that belong to the R&D subgroup have a positive impact and ensure the transfer of new technical knowledge.

The impact of foreign technology on growth depends on the degree of compliance of these technologies with local conditions, as well as on the basic technological level recipient country [4].

Assessment of foreign capital stocks helps increase the efficiency of domestic firms specializing in R&D. Firms that do not belong to the subgroup engaged in the creation of new knowledge do not benefit from the transfer of new technical information. The work of Acharya & Keller [5] is devoted to determining the effects of technology transfers in the US manufacturing sector from attracting FDI and imports of finished products. The study aimed to determine the impact of FDI and import growth on productivity growth in domestic firms receiving investment. The foreign presence can be measured as the share of employment in foreign affiliates compared with total employment in the industry. The study results indicate that FDI creates conditions for obtaining significant production benefits for domestic firms. External effects were assessed on a large scale compared with other studies. External effects accounted for 8–19% of the production growth of American firms in the period from 1987 to 1996. Effects differ and depend on the size and productivity of the firm. Acharya & Keller [5] hypothesized that a strong external effect is due to the influx of FDI into high-tech sectors. Small firms with low productivity receive a greater external effect of FDI compared with large firms with high productivity. The smaller external effect of imports of finished products compared with the effect of FDI can complement the positive feedback to sustain growth in the long run.

FDI has a more positive impact on labor productivity than domestic investment, the technological gap between local enterprises and foreign investors should be relatively small. Boghean and State [6] estimate the relationship between FDI and labor productivity in the European Union and confirm the existence of a strong connection between the volume of FDI outflows and productivity zones. The impact of FDI on the economies of host countries is mainly due to increased productivity through technology transfer, and management and marketing skills, which allows for long-term technological progress and economic growth. The authors argue that FDI depends on additional factors to exert a significant effect on growth, i.e. a significant level of domestic investment or export orientation is necessary.

Clark et al. [7] developed and substantiated an approach to determining the FDI effect, which is associated with positive technological spillovers, economic growth, and increasing income inequality. Floyd [8] used firm-level data from manufacturing industries for the period of 2000–2005 in Central and Eastern European countries and found that vertical effects tend to be higher and thus economically more important than horizontal ones. In many cases, spillovers are negative and thus the foreign presence might also have some adverse impact on local firms' productivity.

The study of the above-mentioned approaches allows identifying the factors and consequences of FDI inflow, through the placement of branches of foreign firms, the technology transfer effects, the development of competition of foreign affiliates and domestic firms, and increasing productivity by attracting FDI [9].

Empirical studies of FDI spillover effects on domestic firms across countries confirm the existence of direct and indirect effects and reflect different technologies, products, and characteristics at micro to macro levels. The reported results do not reproduce different effects of economic sectors, labor productivity, or undervalued labor costs per worker, and do not take into account the role of the shadow economy. Moreover, if internal and external effects act in the same direction, reducing labor costs per unit of output, they act as a factor stimulating the growth of productive efficiency, increasing output, increasing labor intensity, and improving a product's quality and competitiveness.

The external effect of technologies transfer by FDI inflow to the country occurs in horizontal or vertical directions. It depends on the nature of the investments that are invested in the country. Labor-intensive and market-oriented FDI creates a significant external effect for home firms compared to joint-stock firms. The research demonstrates the contradictory results of the external effect of technology transfer in developing countries and East European countries. Don Clarke argues that there is a foreign effect from encouraging FDI inflows policy. The effect is significant in industrially developed countries, which have a high share of high-tech sectors in GDP. The study of the FDI impact on the country's economic growth has found that the external effect of technology transfer is the main factor of long-term economic growth. The FDI effect increases income inequality, even though capital attraction contributes to the growth of prosperity in the country. FDI inflows are positively correlated with the external effect of technology transfer [7].

Diffusion of the acquired advanced technologies promotes their wider use in the process of exchanging ideas. The involvement of these technologies depends on the labor productivity of the labor force involved in the creation of R&D. GDP, the size of the country are significant factors in determining the level of welfare in the country [10].

Petrikova proposed a mechanism for determining the FDI influence on GDP [11]. The author has revealed the algorithm of FDI quantitative estimation and contribution to GDP growth and has estimated the macroeconomic indicators of foreign investments. Osano and Koine [12] confirm that trade competition is accelerating the process of transferring new technologies to local investors in the energy sector through the exchange of knowledge, and application of innovations in production, and R&D.

Radosevich identifies the institutional capabilities of the firm, branch, or economy to combine appropriate market and non-market incentives with the need to upgrade technological and additional prospects [13].

TNCs play an important role in technology transfer. The economic activity of companies is characterized by high labor productivity indicators and significant expenses on research and development in comparison with national companies. Foreign firms use the external effect of technology transfer, applying different approaches. The transfer of new knowledge occurs as a result of the movement of the labor force, which has passed training in TNCs. The obtained new knowledge of workers can be considered as potential for future development of human capital in the country. Workers use the knowledge, skills, and experience gained in Western compa-

nies and in national companies. Local entrepreneurs imitate the production, management, and marketing of foreign branches. The acceleration of competition forces national companies to use resources and advanced technologies more effectively, which ensures that the appropriate effect of the use of skilled labor, as well as profit, is obtained. The growing competition is aggravated by the contradictions between local producers and international companies [14].

Local entrepreneurs import technologies from TNCs through the acquisition of production equipment, specialized capacities, and differentiated products. The external effect of the interaction of foreign affiliates and national companies at horizontal or vertical levels contributes to the increase of labor productivity in the country. Growing demand for intermediate products makes national companies take advantage of the economy [14:111].

Additional research shows that home firms, which belong to the sub-group dealing with NDCD positively influence and ensure the transfer of new technical knowledge. Access to foreign capital stocks contributes to the increase of efficiency of domestic firms specializing in NDRRs. Firms that do not belong to a sub-group that creates new knowledge do not get an external effect from the transfer of new technical knowledge.

Acharya and Keller is dedicated to identifying the impact of technology transfer in the US manufacturing sector from FDI and finished product imports. The study aims to determine the impact of FDI and import growth on labor productivity in home firms, where investments have been made. The authors put forward a hypothesis that a strong external effect takes place in connection with the influx of FDI into high-tech sectors. Small firms with low labor productivity receive a greater external effect of FDI than large firms with high labor productivity. The external effect of the import of finished products is much less compared to the effect of FDI [5].

The study of the considered approaches of influence of direct investments of TNC allows allocating as the main factors of stimulation of economic growth transfer of technologies using the placement of branches of foreign firms, development of competition with national companies. If the technologies are implemented in the format of creation of branches of foreign companies, it stimulates an increase in labor productivity, as well as the transfer of new methods of management, production skills, and business culture to national producers in the country.

## 2. Global strategies of TNCs in high-tech technologies

The activities of global TNCs are aimed at attracting investment, transferring new production and management technologies, stimulating GDP growth, and improving the balance of payments by increasing export revenues or reducing imports. These changes contribute to strengthening the country's economy and raising the living standards of the population. Under the conditions of the Covid-19 pandemic, according to UNCTAD's forecast, the impact of the 5000 largest TNCs is declining, with FDI falling from 30% to 40% in 2020-2021. The profit estimate for 2020 will decrease by an average of 30%, a maximum of 39%. The main affected sectors will be marked by a significant drop in production volumes: energy – 208% and automotive – by 47% [15].

The consequences of the pandemic are a halt to production, supply chain disruptions, and a reduction in capital investment, which can help to prolong the shock to global value chains, as well as to local suppliers and small businesses. Physical closure of businesses, manufacturing plants and construction caused delays in the implementation of global investment projects. In the first part of 2020, there was a 50-70% drop in mergers and acquisitions (M&A) [16].

An innovative policy of TNCs includes the development and promotion of R&D, expansion of technological links, organization of global value-added chains, use of incentives, and creation of industrial, technological, and scientific parks. The increased interest of scientists in this policy, which focuses on technological aspects, allows to significantly increase scientific potential due to technological external effects of FDI. Domestic firms use the results of new knowledge to create a scientific product. The coordination of policy in the field of FDI, combined with the use of research results, innovations, and regional political instruments, is considered a promising direction for the development of technologies. In the countries of the transition economy, FDI inflows are considered a financial source, for the transformation of the national innovation system following the requirements of the global knowledge system, as well as the dissemination of new knowledge. FDI in technology sector saw a 336% rise in Apr-Sep 2020 [17].

On average, the top 5000 multinational enterprises (MNEs), which account for a significant share of global FDI, have seen downward revisions of 2020 earnings estimates of 9% due to COVID-19. The hardest hit is the automotive industry (-44%), airlines (-42%), and energy and basic materials industries (-13%). Profits of MNEs based in emerging economies are more at risk than those of MNEs in developed countries; profit guidance for the latter has been revised downwards by 16% [18].

In the conditions of the worsening international situation, the number of "aggressive" and "hostile" acquisitions and, consequently, the reduction of "friendly". At the "aggressive" acquisitions, TNCs initially buy shares of a foreign company on the stock market and then enter interaction with the general shareholders' meeting. The result of "aggressive" absorption is a complete change in the heads of the consuming company. In the case of a "friendly" acquisition, the agreement is reached between the managers of TNC and shareholders of the consuming company for the purchase and sale of shares. Then shares of this company are exchanged for shares of TNC. Another option of "friendly" acquisition is acquisition in the form of transfer of the controlling stake of TNC to trust or trust management [5].

The participation of TNC branches in international operations to attract FDI through the creation of their foreign firms and joint ventures is aimed at increasing the control of the company. Examples of successful competitive global innovation companies are *Apple*, *Google*, *Samsung Group*, *Toyota*, and *BMW*.

Consider the top 50 of the most innovative Companies 2020 [19] published by Boston Consulting Group. It is an international company specializing in management consulting and is a leading business strategy consultant - see Table 1.

Among the leading companies, are five companies in the field of technology, car manufacturers – *Tesla*, *Toyota*, *Volvo*, and the electrical engineering company *Siemens*. As four factors contributing to the success of innovative companies, the specialists of the Boston Consulting Group have noted the ability of these companies to provide high speed of innovation development, improvement of processes efficiency in the field of research and development, maximum effective use of technological platforms, and systematic study of related markets.

The success of the top 50 - the most innovative companies is based on scientific research and new technologies, which are becoming increasingly important as factors of innovation development, because they promote the development of innovative products, such as those that give impetus to the creation of science-intensive industries of the economy. All ten top rating companies use AI (Artificial Intelligence), platforms, and ecosystems to allow themselves

Table 1. The largest innovation companies in 2021. Constructed on the data of Refs. [19-20]

| Rank | Corporation | Industry | Headquarter | Rank change over 2020 |
|---|---|---|---|---|
| 1 | *Apple* | Technology | USA | - |
| 2 | *Alphabet* | Technology | USA | - |
| 3 | *Amazon* | Consumer goods | USA | - |
| 4 | *Microsoft Corp.* | Technology | USA | - |
| 5 | *Tesla* | Automobile | USA | 6 |
| 6 | *Samsung* | Technology | South Korea | -1 |
| 7 | *IBM* | Technology | USA | 1 |
| 8 | *Huawey* | Telecommunications | China | -2 |
| 9 | *Sony* | Consumer goods | Japan | - |
| 10 | *Pfizer* | Pharmaceutical | USA | return |
| 11 | *Siemens* | Technology | Germany | 10 |
| 12 | *LG Electronics* | Electronics | South Korea | 6 |
| 13 | *Facebook* | Technology | USA | -3 |
| 14 | *Alibaba Group* | Consumer goods | China | -7 |
| 15 | *Oracle* | Technology | USA | 10 |

and others to search for new products, services, and ways of working. There was little change in the ranks of top-ten innovators. *Apple* and *Google* parent *Alphabet* retain the top two spots. But in addition to 33 holdovers from last year (whose continued presence shows the enduring qualities of serial innovators), the 2021 list contains 12 companies that have returned to the top 50 after an absence of at least one year, and 5 firms that are new to the rankings [19].

The leading companies in 2020 included companies from the USA, Europe, Asia, China, and India. Among the leading companies - Chinese Internet company *Tencent* and the American corporation is a manufacturer of personal computers *Dell*, as well as the British-Dutch oil and gas company *Royal Dutch Shell*. It should be noted that the position of *Facebook* has deteriorated. The Japanese company *Sony* returned to the rating this year. Most companies consider AI as well as strong innovations to be positive in business development. Nine out of ten respondents in the current top 50 survey believe that their companies invest in AI, and more than 30% expect investment, point to AI, which will have the greatest impact on its industry over the next three to five years.

Platforms are technologies and technological services that provide the basis for the development of other business processes. Numerous industrial goods companies, including *Siemens* (21) and *Boeing* (11), have created a significant business platform for predictable service to complement their traditional engineering and production facilities. *Amazon*, *Microsoft*, and *IBM*, among others, offer a full range of software and services from their cloud platforms.

Analysis of data of top-100 companies investing in R&Ds shows the growth of total volume of global investments. In total, the Top 1000 Companies spent at least a combined US $858 billion on R&D in 2018, reflecting R&D spending increases in all regions and nearly all industries - see Fig. 1.

According to Strategy Stanley Black & Decker [23], their list of the Top 1000 companies accounted for approximately 40% of all R&D spending worldwide if we extrapolate that total R&D spending worldwide is in the region of $2 trillion annually. By spending money on R&D, the innovation company develops new technologies and research to create new products and services. R&D allows companies like *Amazon* to outcompete and work in the future. The largest share of R&D spending worldwide is on computing and electronics, and the USA is the leading country in R&D spending worldwide. Five large TNCs with the highest costs of R&D include mainly technology companies [21] - see Table 2 .

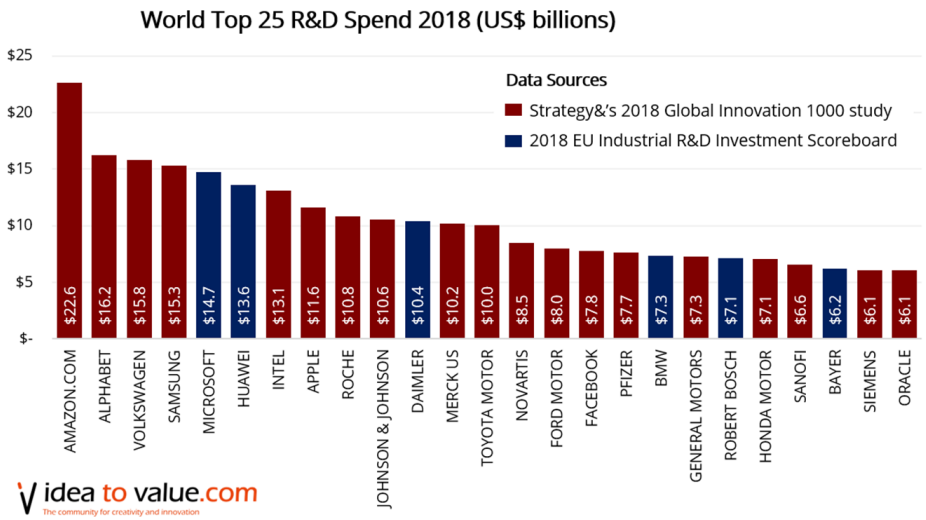



Fig. 1. Top 1000 companies that spend the most on R&D. Adapted according to Ref. [22].

Table 2. Top 100 R&D Spenders. Constructed on the data of Refs. [21,24-27]

| Rank | Company Name | Country | 2018 Revenue (US$ billions) | Max 2018 R&D spend (US$ billions) | R&D intensity 2018 % | Data Source Used |
|---|---|---|---|---|---|---|
| 1 | *Amazon.Com* | USA | 177.87 | 22.62 | 12.72 | [25] |
| 2 | *Alphabet* | USA | 110.86 | 16.23 | 14.64 | [25] |
| 3 | *Volkswagen* | Germany | 277.00 | 15.77 | 5.69 | [25] |
| 4 | *Samsung* | South Korea | 224.27 | 15.31 | 6.83 | [25] |
| 5 | *Microsoft* | USA | 110.42 | 14.74 | 13.34 | [26] |
| 6 | *Huawei* | China | 92.75 | 13.60 | 14.66 | [26] |
| 7 | *Intel* | USA | 62.76 | 13.10 | 20.87 | [25] |
| 8 | *Apple* | USA | 229.23 | 11.58 | 5.05 | [25] |
| 9 | *Roche* | Switzerland | 57.20 | 10.80 | 18.89 | [25] |
| 10 | *Johnson & Johnson* | USA | 76.45 | 10.55 | 13.81 | [25] |
| 11 | *Daimler* | Germany | 197.20 | 10.40 | 5.27 | [26] |
| 12 | *Merck USA* | USA | 40.12 | 10.21 | 25.44 | [25] |
| 13 | *Toyota Motor* | Japan | 259.85 | 10.02 | 3.86 | [25] |
| 14 | *Novartis* | Switzerland | 50.14 | 8.51 | 16.97 | [25] |
| 15 | *Ford Motor* | USA | 156.78 | 8.00 | 5.10 | [25] |

According to the analysis by Stanley Black and Decker [23], R&D Spending overall in their list of the Top 1000 spenders on R&D has continued to rise over the past decade, even following the recession. The top innovators were *AAC Technologies*, *Adidas*, *Altria*, *Amadeus IT*, *Ammer Sports*, *Ametek*, and *Apple* companies.

The most attractive sectors of financing of top-20 companies include automotive, computing and electronics, Software and Internet, telecommunications, chemicals and energy, industries, and health care. Statistics show that investments in R&D health care are the largest and make up $61,7 billion and grow rapidly. The automotive industry ranks second among the investment sectors with a value of $43,7 billion. Investments in computers and electronics make up $37,9 billion, software makes up $30,5 billion [28].

The main trends in the field of innovations that influence the TNC's competitiveness are the large openness of companies within the framework of firm cooperation in the field of science and technology, as well as the transfer of a significant part of research in the branch of foreign countries. Increasingly, the acquisition practices of the largest TNCs that create an innovative product or technology are becoming more common. Parent companies make a decisive contribution to the provision of innovative TNCs. U.S. parent TNCs are the first in the world to provide investment in R&D, which accounts for 52% of all revenues of 20 companies, followed by Germany and Switzerland [28].

The rate of production process speed is becoming an important competitive advantage of TNCs. To accelerate the production process and increase pressure on competitors, international companies reduce bureaucratic restrictions and procedures, create a step-by-step functional working group, restructure and organize production, and introduce new technologies. The technological competitive advantages of TNCs are largely due to the formation of close long-term partnerships with suppliers. Optimization of supply chain management, in particular efficient logistics organization and after-sales service, plays an important role in ensuring the competitiveness of the production system of TNCs [29].

Global TNCs are widely used in cross-border transactions in international mergers and acquisitions (M&A). The knowledge and technology transfer affect leading companies and provide changes in a competitive position in global markets. M&A is one of the indicators of the intensity of the world economy's progress. Due to M&A, as well as the implementation of the corporate strategy, market concentration and economic efficiency rise. Immediate FDI impact on a specific company-investment object is expressed in the new knowledge absorption in the process of joint work organization, cost reduction, and new forms of activities within a company formation.

In 2019, the financial industry became the leader – it accounted for up to 35% of all international agreements.

The global M&A market activity continued to grow. In the conditions of globalization and internationalization M&As become the main source of obtaining competitive advantages for corporations regarding the possibility of rapid formation of investment portfolios due to the attraction of local assets of different countries, reception of new sources of raw materials, integration of stages of the production process, development of new markets, application of new markets.

With the help of M&A, global TNCs can obtain the most efficient and effective access to markets, maintain greater stability, and attract financial resources on better terms.

M&A is primarily cross-border, aimed at gaining competitive advantage by combining financial, scientific, technical and labor resources of firms-participants and obtaining on this basis synergistic effect; achievement of economy effect at the expense of scale of production; access to new markets; diversification of production; use of new assets, especially intangible (such as know-how, know-how, trademarks, organizational knowledge).

The ability of TNCs to use a flexible mechanism of investment activity, organizational methods, and means for global economy change is reflected in the merger of national capital into the global asset network. They have developed a well-developed marketing system and modern advertising methods that allow them to manipulate consumers' tastes and preferences all over the world.

Among the factors that contributed to the high rating of these companies, the following should be highlighted: a) flexible policy in the conditions of economic crisis; b) active production location on the territory of other countries; c) large scale activity and information technologies application.

Most TNCs apply a comprehensive integration strategy, within which firms transform their geographically distributed branches and fragmented production systems into production and distribution networks that are carried out globally or regionally on the global market. By the number of transactions, M&A is leading the technology sector, especially the software sector. In total, 10 thousand transactions were concluded in this sector with a total volume of 808 billion dollars. In 2018, the second place was taken by the sphere of health care with 2.9 thousand transactions and a volume of $580 billion [18]. According to China's strategic needs, three Gorges' cash acquisitions amounted to $3.6 billion, 84% of shares of Peruvian

electric company *Luz del Sur*, the acquisition of Beijing Auto 5% of shares of German car manufacturer *Daimler* and *Jiangsu Shagang Steel* Group, Acquisition of Global Switch Holdings makes up the amount of $2.2 billion located in London Global Switch Holdings in 2019 [30].

Instability in the global financial markets and the current challenges of the global pandemic COVID-19 create uncertainty, and high risks of international agreements that affect the behavior of major TNCs and force them to apply an adaptive strategy, adapting to the new international environment.

In a transformed economy, the prerequisites for economic growth can be achieved by increasing the development and use of science-intensive technologies and increasing their production efficiency because of the application of the advantages of direct investment by TNCs. Separation of priority investment sectors depending on their level of development, and importance for the needs of the national economy. Regional need in FDI will facilitate the development of international business inflow in the innovation sector.

## Conclusions

The study of a variety of modern approaches to foreign direct investment effects on TNCs allowed us to define the main factors stimulating economic growth. The technologies' transfer by the location of foreign affiliates, competition development with local firms, labor productivity increase FDI inflow. TNC's comprehensive integration strategy, within which firms transform their geographically distributed branches and fragmented dedicated production systems into production and distribution networks, global-integrated or regional in the world, is based on the advantages of its application.

The high-tech sector's development of the most innovative TNCs intensifies technologies' creation and efficiency increases because of the usage of international capital movement advantages and new innovative branch's location. R&D investment is the main factor in radical innovation development. The basis for the success of innovative companies in the world has been highlighted. The role of the usage of scientific-intensive technologies and the enhancement of their efficiency of TNCs production by application of the advantages of global value-added chains have been determined. The key factors for achieving TNC's technological leadership by using global production systems are well-grounded.

## Abbreviations

| | | |
|---|---|---|
| AI | - | Artificial Intelligence |
| BGC | - | Boston Consulting Group |
| FDI | - | Foreign Direct Investment |
| M&A | - | Merges and Acquisions |
| MNEs | - | Multinational Enterprises |
| OECD | - | Organisation for Economic Cooperation and Development |
| R&D | - | Research and Development |
| TNC | - | Transnational Corporation |
| UNCTAD | - | United Nations Conference on Trade and Development |

## References


1. OECD (2020).
2. Barell R., Pain N. (1997) Foreign Direct Investment, Technological Change, and Economic Growth Within Europe. – *The Economic Journal* Vol. 107, No. 445.
3. Kathuria V. (2000) Productivity spillovers from technology transfer to Indian manufacturing firms. – *Journal of International Development* April. No 2(3) – P.343-369.
4. Foreign Direct Investment for Development. Maximising Benefits, minimizing costs (2002). OECD. URL: https://www.oecd-ilibrary.org/docserver/9789264199286-en.pdf?expires=1655060530&id=id &accname=ocid49016910 &checksum=166DDA7200D0AD78DDA772EDE8C8245C
5. Acharya P., Keller W. (2007). Technology transfer through imports. NBER working paper series. Working paper 13086. URL: URL: http://www.nber.org/papers/w13086
6. Boghean C., State M. (2015) The Relation between FDI and Labour Productivity in the European Union Countries. – *Procedia Economics and Finance* 32, 278-285. – URL: https://doi.org/10/1016/S2212-5671(15)01392-1
7. Don Clark P., Hihgfill J., De O. Campino J., Rehman, S. S. (2011) FDI, technology spillovers, growth, and income inequality: a selective survey. – *Global economy journal* Vol. 11. Issue 2. Article 1.
8. Floyd D. Foreign Direct Investment in Poland: Is Low-Cost Labour Really the Sole Determinant? – *Economic Issues* 1996. vol. 1, Part 2, September, 69.
9. Nosova O. (2016) Effect of Attracting FDI to Economic Growth of TNCs. – *Bulletin of the Research Center of Corporate Law*, Management and Venture Capital of Syktyvkar State University. No. 4: 13-23.
10. Eaton J., Kortum S. (2001) Technology, trade, and growth: a unified framework. – *European Economic Review* Vol. 45. No. 4-6. P. 754.
11. Petrikova E.P. (2009). Foreign direct investment and economic growth. – *Statistical issues: scientific and informational journal* No 9. 14-21.
12. Osano H.M., Koine P., W. (2016) Role of Foreign Direct Investment on Technology Transfer and Economic Growth in Kenya: a Case of the Energy Sector. – *Journal of Innovation and Entrepreneurship* No. 5(1). November. Retrieved from DOI: 10.1186/s13731-016-0059-3
13. Radosevic S. (1995) Science and technology capabilities in economies in transition: effects and prospects. – *Economics of Transition* Vol. 3, No 4, 472.
14. Nosova O., Lypov V. (2021) Transforming Competitiveness by Introducing Digital Platforms. – *The Journal of World Economy: Transformations & Transitions* No. 3.
15. UNCTAD (2020) Investment Trends Monitor. Impact of the Coronavirus outbreak on global FDI. United Nations, New York, and Geneva, accessed March 2020. URL: https://unctad.org/system/files/information document/diae_gitm34_coronavirus_8march2020.pdf.
16. UNCTAD (2020) World Investment Report: International Production Beyond the Pandemic, United Nations, New York, and Geneva, accessed June 2020. URL: https://unctad.org/system/files/official-document/wir2020_en.pdf.
17. Economic Survey (2020). URL: https://economictimes.indiatimes.com/tech/technology/fdi-in-technology-sector-saw-a-336-rise-in-apr-sep-2020-economic-survey/ articleshow/80586966.cms?utm_source=contentofinterest&utm_medium=text&utm_campaign=cppst
18. UNCTAD (2020) Trade and development report 2020 from global pandemic to prosperity for all: avoiding another lost decade. United Nations, New York and Geneva, accessed June 2020. URL:
19. 50 Most Innovative Companies (2021) The serial innovation imperative. https://www.visualcapitalist.com/ranked-the-most-innovative-companies-in-2021.
20. Overcoming the Innovation Readiness Gap. Most Innovative Companies 2021 (2021). BGC. April. URL: https://www.bcg.com/publications/2021/most-innovative-companies-overview

21. The Global Innovation 1000 Study (2018) Investigating trends at the world's largest corporate R&D spenders. URL:https://www.strategyand.pwc.com/gx/en/insights/innovation1000.html
22. Top 1000 companies that spend the most on R&D , https://www.ideatovalue.com/inno/nickskillicorn/2019/08/top-1000-companies-that-spend-the-most-on-research-development-charts-and-analysis
23. Stanley Black and Decker Company Profile (2021) Fortune 500. – https://fortune.com/company/stanley-black-decker/fortune500
24. Fortune Global 500. – URL: https://fortune.com/global500/
25. Strategy& 2018 Top 1000. – URL:https://www.strategyand.pwc.com/gx/en/insights/innovation 1000/2018-global-innovation-1000-fact-pack.pdf
26. The 2018 EU Industrial R&D Investment Scoreboard. – URL: https://iri.jrc.ec.europa.eu/scoreboard/2018-eu-industrial-rd-investment-scoreboard
27. 2018 Industrial R&D Scoreboard: EU companies increase research investment amidst a global technological race. (2018). European Commission. 17 December. https://ec.europa.eu/info/news/2018-industrial-rd-scoreboard-eu-companies-increase-research-investment-amidst-global-
28. Ang C. (2020) Ranked The 50 Most Innovative Companies. – URL: https://www.visualcapitalist.com/top-50-most-innovative-companies-2020
29. Nosova O. (2021) Effects of Foreign Direct Investment on Labour Productivity. In the book: Productivity of Contemporary Economics. Theory and Evidence. – Eds. O.M. Moskalenko, A.S. Filipenko, Y.K. Zaitsev. Cambridge Scholar Publishing. – Chapter 7.
30. 2020 Global M&A Outlook (2020) Published by J.P. Morgan's M&A team. – URL: https://www.jpmorgan.com/content/dam/jpm/cib/complex/content/investment-banking/2020-global-ma-outlook/pdf-0.pdf

*Article*

# Competitiveness of the Shared Economy Model for Sustainable Management of Logistics Systems

Keisha LaRaine Ingram*
Department of Foreign Languages, Vilnius Business College, Kalvarijų g. 129-401, LT-08221 Vilnius, Lithuania
* Corresponding author, E-mail: *keisha.ingram@kolegija.lt*

Received: 26 May 2022
Accepted: 11 June 2022
Online: 31 August 2022

**Abstract.** This paper analyses the shared economy model to sustainably manage the tangible and intangible resources of logistics systems. A critical review of literature on the current supply chain management policy and technological platform in current use that supports it was done to suggest a new conceptual framework for logistics processes' sustainable management on a shared economy platform. This was examined according to how the concepts of new technologies influence logistics and the role of sustainable management platforms of the shared economy in enabling greater improved logistics processes. The inductive methodology approach was applied using multi-criteria analysis interpretive research method. The impact of the shared business model on each stakeholder and beneficiary varies according to how resources are consumed and its adoption according to the core business models requirements of each. Current scientific literature does not identify the impact this phenomenon has on companies in different sectors, as there is a lack of detailed analysis and evidence to fill this gap, particularly as the Internet of Things (IoT) monetize digital assets autonomously through the Economy of Things (EoT) marketplaces. From the analysis conducted, the findings provide a concept of the prototype framework required for the shared economy in the e-logistics' ecosystems rather than traditional ones, modelled using multi-criteria analysis interpretive methods as a strategic resource within the shared economy of supply chain management systems.

**JEL:** Q01.

**Keywords:** crowdsourcing; sustainable management; sharing economy; digital assets.

**Citation:** Keisha LaRaine Ingram (2022) Competitiveness of the Shared Economy Model for Sustainable Management of Logistics Systems. – *Applied Business: Issues & Solutions* 1(2022)10–16 – ISSN 2783-6967.
https://doi.org/10.57005/ab.2022.1.2

## Introduction

The logistics sector is a main economic driver for Lithuania. A local shared logistics infrastructure should enable continuous operations without any costly significant disruptions during crisis situations [1-9]. Services of e-logistics became central for promulgating the digitalization of the logistics sector in the shared economy. However, there are some risks that require evaluation. The providing framework developed could integrate smart services and technologies as well as enable the elimination of risks [10-15].

In addition, the framework would promulgate management activities for such logistic processes corresponding to the directives of the 2030 Sustainable Agenda [16]. Crowdsourcing capabilities of the shared economy ensures liabilities face by secondary stakeholders are environmentally assessed during the early phases of implementation. Therefore, the main contribution is to provide justification of widening the realm of innovation smart technologies through the shared model. This would increase more resources and introduction of artificial intelligent autonomous systems into the present and future logistics ecosystems of Lithuania.

The aim of the paper is to propose conceptual sustainable processes that is supported by a shared economy model to improve logistics systems and increase the competitiveness of the logistics sector. As services of the logistics sector are inclined towards digitalized, improved technological processes, several problems identified currently impacts Lithuania's logistics sector.

1. The compromised, rippling consequences of the COVID-19 pandemic globally has led to dampening effects on the resilient EU regional economy and, nationally to Lithuania's very own. The logistics sector is one of the main economic drivers for Lithuania and with its enviable geographical positioning, its local shared logistics infrastructure should enable continuous operations without any costly significant disruptions during crisis situations like the pandemic. This trend continues accelerating, however valuable research on the shared economy during normal and crisis events is needed to understand the strategic benefits of smart technologies implementation for sustainably managing e-logistics within a shared economy framework.

2. The scalability of the current model could increase the nation's current market share and potentially increase to a projected margin capacity of 20% by 2025. Moreover, in line with the 2030 Sustainable Agenda, the shared economy platform is environmentally critical for reducing carbon imprint, utilization of warehouse spaces and offsetting climate change. Research on the management implications on the eco-efficiency and usefulness of the shared economy is required to understand the reverberating economic impacts and value capture of tangible and intangible logistics systems.

3. The theoretical categorization of the shared economy according to its economic, environmental, and social benefits substantiates the practical achievement of the 2030 Sustainable Agenda implementation. However, this implementation should be clear and cohesive with the common collaborative envisaged goals set for all stakeholders involved.

## 1. Conceptual Framework of the Shared Economy in Logistics Systems

Stakeholders of third party (3PL) and fourth-party logistics (4PL) logistics sector are unaware of innovative platforms that facilitate long-term competitive advantages. By reducing environmental

pollution, road corridor improvements as well as extending the design life of urban roadway, the management of the e-logistics shared model can minimize environment impacts, waste, and increase its scalability and use by all stakeholders. With the advent of the Internet of Things (IoT), the sharing of digital assets autonomously through marketplaces developed from the Economy of Things (EoT), has enabled ubiquitous commercialization of digital assets, immediate liquefaction of the physical assets indexed to be search and traded as online commodities [17-25]. The shared economy has ratified and authenticated the concept of crowdsourcing and partaking of digital assets through digitalized platforms.

In addition to this, sustainability has gained significance in Lithuania in recent years, with the current Lithuanian logistics sector morphing towards a sustainable, shared economy platform. A strategy for implementing and applying smart technologies for sustainable management of the processes in e-logistics should be grounded on an improved behavioural approach of all stakeholders of the sector. This entails inclining towards responsible economic, environmental, and corporate standards at all levels of the micro and macroeconomy [26-41]. The shared economy model would initiate ubiquitous productivity tools, software and digital technologies developed from the Economy of Things (EoT). The myriad of possibilities offered without the necessity of ownership is one of the appealing factors of the shared model. Moreover, transformation of logistics processes to resilient sustainable, management systems enable greater scalability of shared transport and warehousing systems models fostered by strategic eco-sustainable imperatives and green logistics. Baryšienė et al [42] points out that 'green logistics', a term synonymous to sustainability has gained increasing significance in Lithuania. Green logistics in Lithuania addresses only the environmental aspects of logistics and perceived as the main way to retain and increase business competitiveness in that sector [43,42].

The sharing economy has garnered attention by policymakers and scientists, due to the potential benefits granted to transform greater economic, technological, and environmental benefits of underutilized assets in P2P service platforms enabled by the Internet. The sharing economy is its action-based state which was initially demarcated by Belk [44], then applied as an economic concept by Zervas et al [45] and Hu et al [46] beyond rent seeking of good, services and resources at renting goods at reduced costs through a conventional supplier. Therefore, the viability of the shared economy concept as a sustainable competitive model through EoT in the logistics sector would be the associated reduced transactional costs typically associated with traditional purchasing or rental activities through the P2P platform, as a form of crowd-based capitalism. This capitalism, as contended by Hu et al [46] provides seamless transference of ownership made possible by on-demand access facilitated through internet connectivity which results in collaborative consumption of services and ubiquitous conversion of digital intangible resources to physical consumable goods. Though Hu et al [46] contends the sharing economy transcends into a triadic framework of agents consisting of service providers, customers, service enablers monetizing and converting underutilized assets in the shared economy model. Therefore, an understanding of its dynamic configuration and development, which disruptively changes according to the requirements (customers), markets (service enablers), technologies (service providers) and its structure of interactive commercial activities (an intent to share) is required to assess its attractiveness configured in a sustainable business model to increase competitiveness in the logistics sector. Hu et al [46] argues within this context, that mechanisms determining value capture, value creation, technological, economic as well as environmental drivers determining the level of sustainability is typically designated during the development of the framework initially before ascertaining competitiveness.

Despite this, the current situation reveals that the sector is unaware of the concept green or sustainable logistics or even a shared economy built on an EoT platform due to minimal information on that sphere [42]. Moreover, research has shown that strategies for implementing such initiatives should be grounded on the changing the behavioural needs of stakeholders economic, environmental, and corporate responsibility in the adoption of sustainable, innovative technological platforms. The shared economy could provide this for both traditional and electronic logistics systems in supply chain management. This potentially revolutionizes its traditional business model through disruptive sustainable innovations platforms offered by EoT, Big Data Analytics, Cloud Computing technologies' productivity tools [47-51] for the logistics sector in a shared, scalable model. The possibilities offered without the necessity of their ownership emphasize the resilience of the model for sustainable management of logistics processes for both traditional and electronic logistics. The shared economy model accelerates and broadens common use of intelligent artificial systems beyond the scope of 3PL and 4PL stakeholders for greater consumption with minimal environmental impact and waste.

Fig. 1 represents conceptual framework of the shared economy model.

## 2. Research Methodology

An analytical comparative review of related, limited research works in the field of logistics management in Lithuania was done to evaluate the proposed model of shared economy framework integrated in traditional and e-logistics ecosystems. The review focused on e-logistics which is a recent developing area of research in Lithuania. The sustainability of this model was evaluated through multi-criteria decision-based methods applied to demonstrate its feasibility as a strategic resource in e-logistics ecosystems' processes. A critical review of the literature on the current technological platform was done to support and identify its accompanying challenges particularly during crisis events such as Covid-19. The review and analysis done was according to how the concepts of new technologies impact the competitiveness of logistics network systems in the context of the sharing economy.

Researchers Batarlienė and Meleniakas [54], Dabbous and Tarhini [55], Apte and Davis [56] note that factors such as the level of application of technology have a significant impact on the growth of the sharing economy. Moreover, businesses in the logistics sector have throughout the years created valuable IoT data [57-58,49,59,48,47], stored and hosted on cloud-based servers processed by conventional analytical computer algorithms and applications, creating the EoT. The management of logistics processes is strongly focused on data, processing, and transmission. However, the use of smart technologies is not only for the realization of information collection, systemization, storage, and monitoring needs, but also for the rapid selection and transmission of targeted information to the end-user. Due to the full volume of information generated it enables automatic alerts or messages to both creators and users. The use of smart technologies for the management of logistics processes can accelerate digitalization, promote resource efficiency, and reduce environmental pollution as evidenced by Bučaitė-Vilkė and Tereškinas [60], Moldabekova [61], Akkad and Bányai [62], Davidavičienė et al [63], Wei et al [64], Junge [65], Leichteris et al [66] and Viederytė [67].

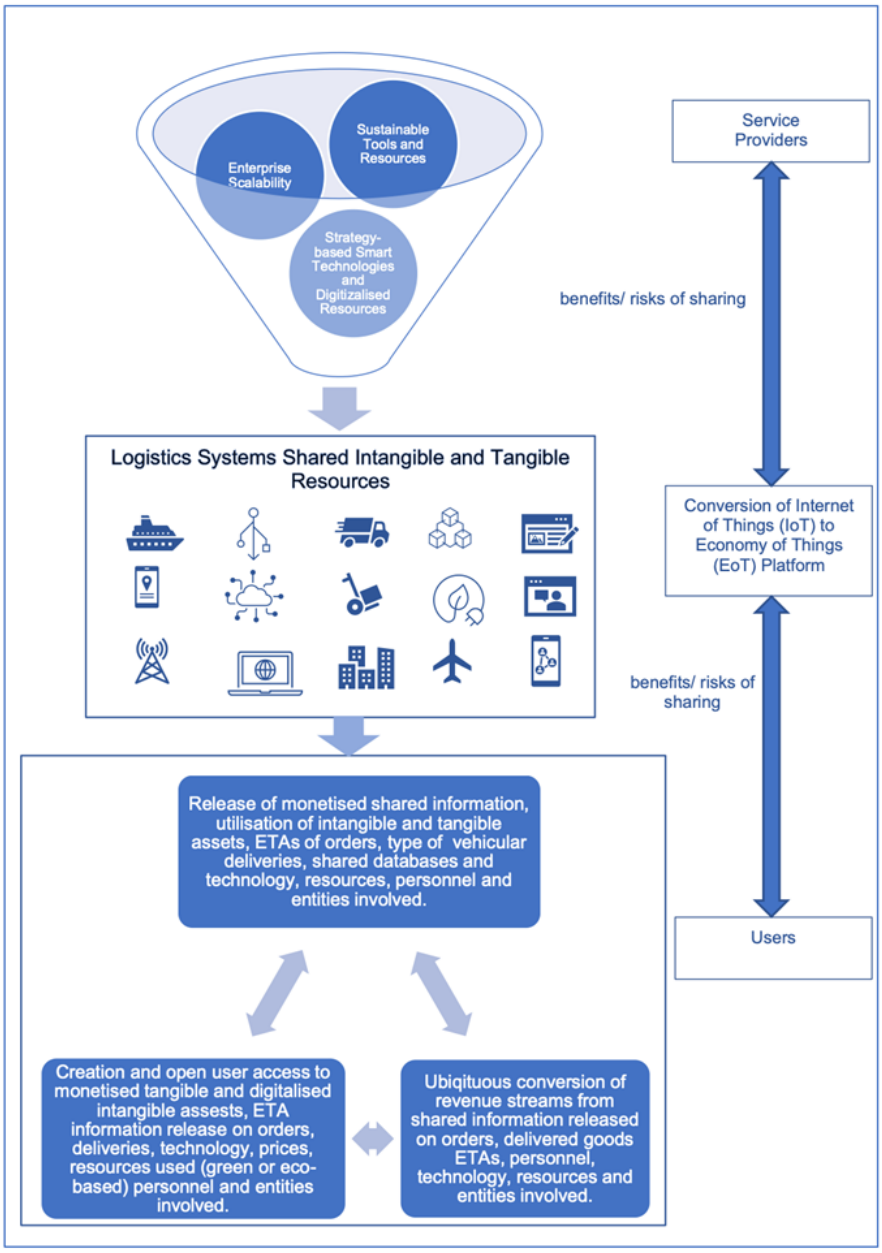



Fig. 1. Conceptual Framework of the Shared Economy Model. Developed by the Author according to Refs. [42,47-49,52-53].

Sustainability issues are seen as the future objectives of the European Union. These objectives are mentioned in the European Commission programme [68], which focus on saving materials, efficient energy consumption and reducing waste. The program set earmarked the priority areas of the EC for greater cost-effective and efficient use of resources within the economic activities of each EU Member State logistics sector that would increase its competitiveness and broaden common revenue streams from that sector, sustainably [69-77,62,78-80,42].

Regarding the competitiveness of Lithuania's logistics sector, Vienažindienė et al. [81] as well as further research by Bučaitė-Vilkė and Tereškinas [60], Katinienė et al [82] and Navickas et al [43] lack focus on key areas that contribute to the sector's competitiveness in Lithuania particularly in commercial enterprises services offered both private and public sector entities. Though the current analysis is focused on suggestions on how the shared model analyzed in a multi-dimensional approach could improve the competitiveness of the sector, potential complex unresolved challenges associated with the sustainability could rebuff the smart technology aspect of the model as contended by Bhawsar and Chattoopadhyay [83], Vyas et al. [84], Drobyazko et al. [85], Falciola et al [86] and Rahulina [53].

## 3. Results and Findings

For the shared economy model recommended for 3PL/4PL systems in Lithuania, achieved competitiveness are categorized according to 'the intent to share', 'ecological' and 'economic risks/benefits' for each stakeholder, which correspondingly are: a) intent to share; b) ecological eenefits; c) economic benefits - see Fig. 2.

### 3.1. Intent to Share

At the macro-regional level, the contribution of each EU Member State's logistics sector is evident for the common EU economy revenue streams as well as the job-creation capabilities created by the sector. The sector employs approximately eleven million (11 million) persons forming up to five percent (5the logistics sector is expected to grow by more than 50and vital to increasing its competitiveness under a shared economy model. The beneficiaries and stakeholders set to gain from this model, primarily are: a) enterprises; b) consumers; c) government. The benefits for each vary according to how resources are consumed within the model's ecosystem and core business models. Nonetheless, a hybrid situation should be adopted due to the networked structure of the proposed ecosystem and its impact on the type of resources shared: intangible and tangible.

Pertaining to ownership transference from shared activities for intangible and tangible resources that form part the transactional aspects of the model, this would vary according to the economic value interest of the goods and service being shared. Furthermore, it can be of several forms: non-ownership, renting, accessing, and temporary transference of ownership which dictates the type of framework system for the shared economy model recommended for Lithuania's competitiveness.

### 3.2. Ecological Benefits

Several ecological benefits could be described as presented below:

i) introduction of leaner, sustainable innovative technologies for management of e-logistics and conventional processes;
ii) ecological and sustainable use of resources;
iii) greater consumption of resources and reduced carbon imprint;
iv) establishment of a community consisting of all stakeholders that are driven by common goals.

### 3.3. Economic Benefits

Several economic benefits must be described as presented below.

1. Greater costs savings and eco-efficiency of resources.
2. Reduced time and maintenance costs of private and public infrastructure as well as conservation of energy utilized for contract execution of each transaction.
3. An interactive database at the governmental level where new and existing actors can update the goods or services offered and newer actors can register. This would transform the logistics industry from competitive to sharing which is an attractive feature of regional competitiveness.

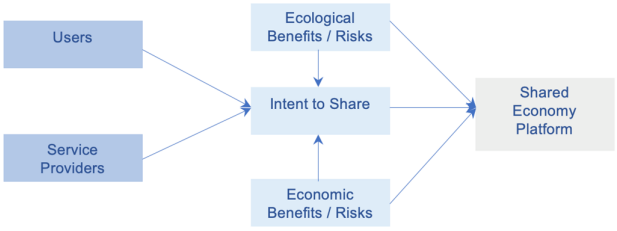



Fig. 2. Key Dimensional Drivers of the Shared Economy. Developed by the Author.

4. Generation of a new labour employment model. This new labour model would be termed as micro-entrepreneurships resulting from sustainable activities executed in the shared economy model.
5. Increased anti-theft solutions and greater security of capital infrastructure and technological resources.

## 4. Discussions

The sharing of resources, technology and cooperation is not a new phenomenon in Lithuania's logistics sector. One the most important strategic and operational success factors of this phenomena are leaner, sustainable systems that are crucial for increasing the attractiveness of the sector. The analysis done have supplement knowledge to the field of sustainability and supply chain management systems supported by innovative smart technologies. Therefore, improved technological processes sustained by the shared economy model could increase the attractiveness and competitiveness for Lithuania in the following ways.

1. Where logistics enterprises and business entities are more inclined towards those strategies that enable the efficient 'sharing' tangible and intangible assets in e-logistics processes, there will be more potential to focus less on the 'ownership of things' and more towards 'sharing of things' perspectives. This integration would be more inclined to consistent sustainable economic stability during normative and crisis events for Lithuania. Moreover, it would enable identification within the shared economy ecosystem the most important strategic factors for success and future requirements for sustainable management of the model. Moreover, decision-making processes concerning improving present physical infrastructure ecologically to achieve 'greener' logistics increase business competitiveness in Lithuania.

2. By analysing the peripheral interconnection between entities and stakeholders that share resources in the current model, partnerships forge according to common economic benefits increases Lithuania's competitive advantage regionally. This strategic orientation enables more eco-scalable operations in the sector, particularly with respect to the identification of roles that each stakeholder and entities would have in the shared model ecosystem. Through greater consolidation of capital infrastructure and resources at the national level, government subsidies can be better utilized to monitor tangible and intangible resources of the proposed shared model economically.

3. Due to the interconnectedness resulting from common shared activities, improved performance will enable the sector to increase its competitive advantage. The shared model supported by cloud computing hosting platforms, EoT and IoT intangible assets, Big Data analytics will foster greater use of physical and digital assets more meaningfully. The sharing of common resources increases greater utilization of available resources flexibility, lowered infrastructural and operational costs. This leads to quality in the services offered and enable better process management and planning at the policy level for the logistics sector of Lithuania.

**Research Limitations.** The main limitation of this paper is that more research is needed to explore which policies and practices would sustainably mitigate waste and greater resource management in the industry, through data warehousing capable IoT, EoT and other cloud hosting platform technologies, both for fourth-party logistics (4PL) and third-party logistics (3PL) framework systems in Lithuania. Moreover, it is recommended that further research on the key areas within the shared economy framework, that directs the designation and use of the green logistics strategy by stakeholders

of privately and publicly owned commercial enterprise services is needed to assess whether commercial competitiveness is sustained or depleted. This promulgates whether the shared economy model is attractive for the logistics sector.

## Conclusions

1. A decade ago, a new phenomenon called the *shared economy* have steadily grown. The shared economy offers new opportunities for consumers and entrepreneurs as well as the environment through the efficient use of resources. It can be argued that the shared economy is a new business model that has a specific competitive advantage over traditional models used in the logistics sector.

2. The logistics sector is one of the main economic drivers for Lithuania and with it an enviable geographical positioning. The proposed shared economy model is central for promulgating the digitalization of the entire logistics sector within realms of sustainability, thereby garnering both economic relevance and environmental practicality for Lithuania according to the goals of the 2030 Sustainable Agenda. Furthermore, the study indicates that the crowdsourcing capabilities garnered by the shared economy ensures perceived liabilities faced by primary and secondary stakeholders are assessed during the early phases of implementation. This enhanced sustainable management system, supported by a shared economy platform promulgates further crowdsourcing, renting and monetizing of digitalized assets, infrastructure, and resources in both tangible and intangible logistics systems.

3. An initial understanding of the strategic use of the 'green logistics' to sustainably manage traditional and e-logistics systems should extend beyond the environmental scope to retain and increase commercial competitiveness as well as value capture for all stakeholders. Nonetheless, a hybrid situation is suggested due to the networked structure of the shared economy ecosystem. When extended to *Industry 4.0* policy projections earmarked for Lithuania, it becomes even more integral in achieving value capture as well as the creation of new labour market models in sustainable framework for logistics systems.

## Abbreviations

| | | |
|---|---|---|
| EoT | - | Economy of Things |
| GDP | - | Gross Domestic Product |
| IoT | - | Internet of Things |
| P2P | - | Peer-to-Peer |
| 3PL | - | Third-Party Logistics |
| 4PL | - | Fourth-Party Logistics |

## References


1. Burinskienė A. (2021) Digitization Of Socio-Economic Processes in Term of the Pandemic. – Actual issues of modern development of socio-economic systems in terms of the COVID-19 pandemic, 258.
2. Liu W., Zhang J., Wei S., Wang D. (2021) Factors influencing organisational efficiency in a smart-logistics ecological chain under e-commerce platform leadership – *International Journal of Logistics Research and Applications* 24(4), 364-391.
3. Azhari B., Taufik T. (2021) Strategy Formulation of Smart Logistics Development in A National Logistics Company. – In Conference: The 3rd International Conference on Management of Technology, Innovation, and Project (MOTIP 03).
4. Woschank M., Zsifkovits H. (2021) Smart logistics–conceptualization and empirical evidence. – *CMUJ. Nat. Sci* 20(2), e2021030.
5. D'Amico G., Szopik-Depczyńska K., Dembińska I., Ioppolo G. (2021) Smart and sustainable logistics of Port cities: A framework for comprehending enabling factors, domains and goals. – *Sustainable Cities and Society* 69, 102801.
6. Göçmen E. (2021) Smart Airport: Evaluation of Performance Standards and Technologies for a Smart Logistics Zone. – *Transportation Research Record* 2675(7), 480-490.
7. Busso D., Berardino D. D., Tonelli A. (2021) The Debate on IC, Smart Technologies, and Digitalization. –In Intellectual Capital, Smart Technologies and Digitalization Springer, Cham.: pp. 1-20.
8. Yan X., Liu W., Lim M. K., Lin Y., Wei W. (2022) Exploring the factors to promote circular supply chain implementation in the smart logistics ecological chain. – *Industrial Marketing Management* 101: 57-70.
9. Li W., Wang Y. (2021, Dynamic evaluation of logistics enterprise competitiveness based on machine learning and improved neural network. – *Journal of Ambient Intelligence and Humanized Computing* 1-15.
10. Issaoui Y., Khiat A., Haricha K., Bahnasse A., Ouajji H. (2022, An advanced System to enhance and optimize Delivery Operations in a Smart Logistics Environment. – *IEEE Access*.
11. Aydınocak E. U. (2022, Internet of Things (IoT) in Marketing Logistics. – In: Logistics 4.0 and Future of Supply Chains (pp. 153-169). Springer, Singapore.
12. Bukhtoyarova M. V. (2022, The Need to Introduce Digital Technologies in Logistics Companies. – In: Proceedings of the International Scientific Conference "Smart Nations: Global Trends in The Digital Economy" – Springer, Cham.: pp. 395-401.
13. Morelli G., Magazzino C., Gurrieri A. R., Pozzi C., Mele M. (2022, Designing Smart Energy Systems in an Industry 4.0 Paradigm towards Sustainable Environment. – *Sustainability* 14(6), 3315.
14. Yuan H., Yang B. (2022, System Dynamics Approach for Evaluating the Interconnection Performance of Cross-Border Transport Infrastructure. – *Journal of Management in Engineering* 38(3), 04022008.
15. Ji Y. (2022, Exploratory Research on The Relationship Between Digital Service Supply Chain Capability and Supply Chain Performance. – *The EUrASEANs: Journal on Global Socio-Economic Dynamics* (1 (32)), 7-20.
16. 2030 Sustainable Agenda (2020)
17. Azcoitia S. A., Laoutaris N. (2022, A Survey of Data Marketplaces and Their Business Models. – *arXiv preprint* arXiv:2201.04561.
18. Da Silva F., Núñez Reyes G. (2022, The era of platforms and the development of data marketplaces in a free competition environment.
19. Salih K. O. M., Rashid T. A., Radovanovic D., Bacanin N. (2022, A Comprehensive Survey on the Internet of Things with the Industrial Marketplace. – *Sensors* 22(3), 730.
20. Staab P., Pietrón D., Hofmann F. (2022, Sustainable Digital Market Design: A Data-Based Approach to the Circular Economy.
21. Manganelli A., Nicita A. (2022, Digital Capitalism and the New Economy (ies). – In: Regulating Digital Markets (pp. 35-54). – Palgrave Macmillan, Cham.
22. Abbas A. E., Agahari W., van de Ven M., Zuiderwijk A., de Reuver M. (2021 Business Data Sharing through Data Marketplaces: A Systematic Literature Review. – *Journal of Theoretical and Applied Electronic Commerce Research* 16(7), 3321-3339.
23. Kumar A., Sharma S., Singh A., Alwadain A., Choi B. J., Manual-Brenosa J., Goyal N. (2021) Revolutionary Strategies Analysis and Proposed System for Future Infrastructure in Internet of Things. – *Sustainability* 14(1), 71.

24. Serrano W. (2021) Validation and Verification of Data Marketplaces. – In: IFIP International Conference on Artificial Intelligence Applications and Innovations. – Springer, Cham.: pp. 780-792
25. Vadlamudi S. (2021) The economics of internet of things: An information market system. – *Asian Business Review* 11(1), 35-40.
26. Barykin S. Y., Kapustina I. V., Kalinina O. V., Dubolazov V. A., Esquivel C. A. N., Alyarovna N. E., Sharapaev P. (2021) The Sharing Economy and Digital Logistics in Retail Chains: Opportunities and Threats. – *Academy of Strategic Management Journal* 20, 1-14.
27. Strulak-Wójcikiewicz R., Wagner N. (2021) Exploring opportunities of using the sharing economy in sustainable urban freight transport. – *Sustainable Cities and Society* 68, 102778.
28. Rześny-Cieplińska J., Szmelter-Jarosz A. (2021) Stakeholders' analysis of environmental sustainability in urban logistics: a case study of Tricity, Poland. Energies, 14(5), 1274.
29. Lazar S., Klimecka-Tatar D., Obrecht M. (2021) Sustainability orientation and focus in logistics and supply chains. – *Sustainability* 13(6), 3280.
30. Koshta N., Patra S., Singh S. P. (2022) Sharing economic responsibility: Assessing end user's willingness to support E-waste reverse logistics for circular economy. – *Journal of Cleaner Production* 332, 130057.
31. Ali S. S., Kaur R., Khan S. (2022) Evaluating sustainability initiatives in warehouse for measuring sustainability performance: an emerging economy perspective. – *Annals of Operations Research* 1-40.
32. Savin G., Katochkov V., Grishina V. (2022) Digital Logistics as Institution for Improving Coordination of Flow Processes. – In: Transport and Logistics System of the Smart City. – *Gene (Mutation)* 1(2.0), 3-0.
33. Li D., Huang Y., Sun H., Zhi B. (2022) Achieving sustainability in sharing-based product service system: A contingency perspective. – *Journal of Cleaner Production* 332, 129997.
34. Šepel'ová L., Calhoun J. R., Straffhauser-Linzatti M. (2022) Sustainability Drives of the Sharing Economy. – In: Developments in Information Knowledge Management for Business Applications, Springer, Cham.: pp. 139-154
35. Vorozheykina T. M., Larina L. B., Erokhin S. G., Vertiy M. V., Donskova L. I. (2022) Sharing Economy as a New Form of Consumer Cooperation in the Digital Age. – In: Cooperation and Sustainable Development (pp. 141-148). Springer, Cham.
36. Remondino M., Zanin A. (2022) Logistics and Agri-Food: Digitization to Increase Competitive Advantage and Sustainability. Literature Review and the Case of Italy. – *Sustainability* 14(2), 787.
37. Simonovits B., Balázs B. (2022) From Uberisation to Commoning: Experiences, Challenges, and Potential Pathways of the Sharing Economy in Food Supply Chains in Europe. – *The Sharing Economy in Europe* 137.
38. Yang H., Lee H. (2022) How does the perceived physical risk of COVID-19 affect sharing economy services? – *Current Issues in Tourism* 1-17.
39. Rosário A., Dias J. C. (2022) Sustainability and the Digital transition: a literature review.
40. Rojanakit P., de Oliveira R. T., Dulleck U. (2022) The sharing economy: A critical review and research agenda. – *Journal of Business Research* 139, 1317-1334.
41. Zhou Z., Wan X. (2022) Does the sharing economy technology disrupt incumbents? Exploring the influences of mobile digital freight matching platforms on road freight logistics firms. – *Production and Operations Management* 31(1) 117-137.
42. Baryšienė J., et al. (2015) Analysis of the current logistics and transport challenges in the context of the changing environment. – *Transport* 30. 233-241. 10.3846/16484142.2015.1046403.
43. Navickas V., Petrokė I., Bačiulienė V. (2021) Development of creative energy potential of managers in the aspect of increasing business competitiveness. – In: 14th international scientific-methodical-practical conference on sustainable regional development: Economics, management, law and technology opportunities 2021 Klaipeda, Lithuania, 1-2 October 2021: abstract book (pp. 36-37). Hungarian University of Agriculture and Life Science.
44. Belk R. (2007) Why not share rather than own? – *Ann. Am. Acad. Political Soc. Sci.* 611, 126–140.
45. Zervas G.; Proserpio D.; Byers J. (2015) A First Look at Online reputation on Airbnb, Where Every Stay Is above.
46. Hu H., Huang T., Cheng Y., Lu H. (2019) The evolution of sustainable business model innovation: Evidence from a sharing economy platform in China. – *Sustainability* 11(15), 4207.
47. Miller S. R. (2016) First Principles for Regulating the Sharing Economy. – *Harvard Journal on Legislation* 52.
48. Cusumano M. A. (2015) How Traditional Firms Must Compete in the Sharing Economy. – *Communications of the ACM* 58, 1, 32-24, http://dx.doi.org/10.1145/2688487.
49. Wieteska G., and Ocicka B. (2017) Sharing economy in logistics and supply chain management. – *LogForum* 13, 183-193, 10.17270/J.LOG.2017.2.6.
50. Pagoropoulos A., Pigosso D. C., McAloone T. C. (2017) The emergent role of digital technologies in the Circular Economy: A review. – *Procedia CIRP* 64, 19-24.
51. Trushkina N., Dzwigol H., Serhieieva O., Yu S. (2020) Development of the Logistics 4.0 Concept in the Digital Economy. – *Ekonomicheskij Vestnik Donbassa (in rus.)* (4 (62)), 85-96.
52. Prause G. (2015) Sustainable business models and structures for Industry 4.0. – *Journal of Security – Sustainability Issues* 5(2).
53. Rahulina E. (2022) Enhancing the use of sustainable transport in Logistics: the case of transportation companies in Lithuania –Master's thesis.
54. Batarlienė N., Meleniakas M. (2021) Claims Solutions Using a Blockchain System in International Logistics. – *Sustainability* 13(7), 3710.
55. Dabbous A., Tarhini A. (2019) Assessing the impact of knowledge and perceived economic benefits on sustainable consumption through the sharing economy: A sociotechnical approach. Technological Forecasting and Social Change. 149. 119775. 10.1016/j.techfore.2019.119775.
56. Apte U. M., Davis M. M. (2019) Sharing economy services: Business model generation. – *California Management Review* 61(2), 104-131.
57. Wasielewska-Marszałkowska I. (2021) Identification of Innovativeness Provided by Logistics Service Providers. – *European Research Studies* 24, 40-52.
58. Trenz M., Frey A. and Veit D. (2018) Disentangling the facets of sharing: A categorization of what we know and don't know about the Sharing Economy. – *Internet Research* Vol. 28 No. 4, pp. 888-925. https://doi.org/10.1108/IntR-11-2017-0441
59. Grybaitė V., and Stankeviciene J. (2016) Motives for participation in the sharing economy - Evidence from Lithuania. – *Ekonomia i Zarzadzanie* 8.10.1515/emj-2016-0028.
60. Bučaitė-Vilkė J., Tereškinas A. (2021) Key lessons from the seven COHSMO countries: Lithuania. Cohesion on the ground: perspectives and experiences/editors Maja Neergaard and Rikke Skovgaard Nielsen – Aalborg: Aalborg University, 2021.
61. Moldabekova A., Philipp R., Reimers H. E., Alikozhayev B. (2021) Digital technologies for improving logistics performance of countries. – *Transport and Telecommunication Journal* 22(2), 207-216.
62. Akkad M. Z., Bányai T. (2020) November, Applying Sustainable Logistics in Industry 4.0 Era. – In: Vehicle and Automotive Engineering, Springer, Singapore.: pp. 222-234.
63. Davidavičienė V., Raudeliūnienė J., Viršilaitė R. (2019) User experience evaluation and creativity stimulation with augmented reality mobile applications. – *Creativity Studies* 12(1), 34–48. https://doi.org/10.3846/cs.2019.3576
64. Wei F., Alias C., Noche B. (2019) Applications of digital technologies in sustainable logistics and supply chain management. – In: Innovative Logistics Services and Sustainable Lifestyles, Springer, Cham.: pp. 235-263

65. Junge A. L. (2019) Digital transformation technologies as an enabler for sustainable logistics and supply chain processes–an exploratory framework. – *Brazilian Journal of Operations & Production Management* 16(3), 462-472.
66. Leichteris E., et al. (2018) Lithuanian Industry Digitalization Roadmap 2019-2030 Annex I Analytical Report (Pre-Final Version). – Vilnius, 2018-05-25 (revised 2018-11-20). – [Interactive] [Accessed on 27-12-2020] retrieved from https://inovacijos.lt/media/industry%20digitalization%20roadmap%20ANNEX%20I%20pre-final%20draft.pdf.
67. Viederyte R. (2021) Organizational and Process Innovations in International Logistics Companies: The Relevance and Expected Benefits. – *Regional Formation and Development Studies* 20(3), 134-146. doi:10.15181/rfds.v20i3.1350.
68. European Commission (2021) Sustainable development in the European Union — Monitoring report on progress towards the SDGs in an EU context — 2021 edition. European Union. (2021). [Interactive] [Accessed on 15-11-2021] https://ec.europa.eu/eurostat/documents/3217494/12878705/KS-03-21-096-EN-N.pdf/8f9812e6-1aaa-7823-928f-03d8dd74df4f?t=1623741433852
69. Qian X., Fang S. C., Yin M., Huang M., Li X. (2021) Selecting green third party logistics providers for a loss-averse fourth party logistics provider in a multiattribute reverse auction. – *Information Sciences* 548, 357-377.
70. Qian X., Yin M., Chan F. T., Zhang J., Huang M. (2021) Sustainable–responsive winner determination for transportation service procurement auctions under accidental disruptions. – *Journal of Cleaner Production* 320, 128833.
71. Kiggell L. S., Kilbourn P. J., Heyns G. J. (2021) Sustainable cost reductions obtained by using a fourth party logistics provider. – *Journal of Transport and Supply Chain Management* 15, 13.
72. Koliński A., Werner-Lewandowska K. (2021) Determinants of the Use of Logistic Labels by 3PL and 4PL Operators-Results of Studies in Poland. – *European Research Studies* 24(2B), 871-881.
73. Dzwigol H., Trushkina N., Kvilinskyi O. S., Kvilinskyi O. S. (2021) Green logistics as a sustainable development concept of logistics systems in a circular economy – Doctoral dissertation, Proceedings of the 37th International Business Information Management Association (IBIMA)).
74. Singhdong P., Suthiwartnarueput K., Pornchaiwiseskul P. (2021) Factors Influencing Digital Transformation of Logistics Service Providers: A Case Study in Thailand. – *The Journal of Asian Finance, Economics and Business* 8(5), 241-251.
75. Cohen M.J. (2020) Does the COVID-19 Outbreak Mark the Onset of a Sustainable Consumption Transition? – *Sustain. Sci. Pract. Policy* doi:10.1080/15487733.2020.1740472.
76. Ivanov D. (2020) Predicting the Impacts of Epidemic Outbreaks on Global Supply Chains: A Simulation-Based Analysis on the Coronavirus Outbreak (COVID- 19/SARS-Cov-2). – *Case. Transp. Res. Part E Logist. Transp. Rev.* 136, doi: 10.1016/j.tre.2020.101922.
77. Kumar A., et. al. (2020) COVID-19 Impact on Sustainable Production and Operations Management. – *Sustainable Operations and Computers* Volume 1, Pages 1-7, ISSN 2666-4127. Interactive. [Access on 06/01/ 2021] (https://doi.org/10.1016/j.susoc.2020.06.001).
78. Hosseini S. et al. (2019) Review of Quantitative Methods for Supply Chain Resilience Analysis. – *Transp. Res. Part E Logist. Transp. Rev.* 125, 285–307, doi: 10.1016/j.tre.2019.03.001.
79. Kayikci Y. (2018) Sustainability impact of digitization in logistics. – *Procedia manufacturing* 21, 782-789.
80. Rossi et. al. (2015) The logistics service providers in eco-efficiency innovation. – *Supply Chain Management: An International Journal* 18/6 (2013) 583–603. Emerald Group Publishing Limited [ISSN 1359-8546] [DOI 10.1108/SCM-02-2012-0053].
81. Vienažindienė M., Tamulienė V., Zaleckienė J. (2021) Green Logistics Practices Seeking Development of Sustainability: Evidence from Lithuanian Transportation and Logistics Companies. – *Energies* 14(22), 7500.
82. Katinienė A., Jezerskė Ž., Vaičiūtė K. (2021) Research on competencies of logistics specialists in transport organisations. – *Journal of Business Economics and Management* 22(5), 1308-1322.
83. Bhawsar P., Chattopadhyay U. (2018) Evaluation of industry cluster competitiveness: a quantitative approach. – *Benchmarking: An International Journal* 25. 00-00. 10.1108/BIJ-02-2017-0022.
84. Vyas N., Beije A., and Krishnamachari B. (2019) Blockchain and the supply chain - concepts, strategies and practical applications. – New York, USA: Kogan Page Publishers.
85. Drobyazko S., Barwińska-Małajowicz A., Ślusarczyk B., Zavidna L., Danylovych-Kropyvnytska M. (2019) Innovative Entrepreneurship Models in the Management System of Enterprise Competitiveness. – *Journal of Entrepreneurship Education* 22(4).
86. Falciola J., Jansen M., Rollo V. (2020) Defining firm competitiveness: A multidimensional framework. – *World Development* 129. 104857. 10.1016/j.worlddev.2019.104857.

*Article*

# The Role of Communication for Strategic Management in 21st Century

Audrius Dabrovolskas*
Department of Communication and International Relations, Vilnius Business College,
Kalvarijų g. 129-401, LT-08221 Vilnius, Lithuania
* Corresponding author, email: *Audrius.Dabrovolskas@kolegija.lt*

Received: 01 June 2022
Accepted: 12 July 2022
Online: 31 August 2022

**Abstract.** 21st century is heavily influenced by rapid development of digital technologies and the volumes of information. This brings challenges to organizations striving to acquire better position in the global marketplaces. One of the challenges is how to bring innovations in this fiercely competitive global environment that would lead companies to achieve their strategic objectives. The article implies that all employees must be involved in maintaining internal and external communication, which has to be strategically oriented. It means that communication planning and performance should be integrated in the companya lations business strategy and executed in all levels of management since it affects the image of an organization.

**JEL:** M000.

**Keywords:** internal and external communication; strategic communication; strategic management; organizations.

**Citation:** Audrius Dabrovolskas (2022) The Role of Communication for Strategic Management in 21st Century. – *Applied Business: Issues & Solutions* 1(2022)17–21 – ISSN 2783-6967. https://doi.org/10.57005/ab.2022.1.3

## Introduction

In 21st century both organizations and users are facing challenges brought by intense globalization, information, and communication technologies (ICTs). More and more information is produced worldwide and distributed through different media channels and newly emerging platforms, which according to Manuel Castells, created "Networking Society" [1] and brought "The Information Age" and informational capitalism [2]. From the user perspective, this situation brings issues related with how to deal with such amount of information and maintain critical balance while using sources from internet. On the other hand, organizations must invent new ways of attracting user attention to sell their services and products, therefore, management would achieve its objectives and provide substantial results. According to Marin Buble, "the changes in the external environment have led to a number of changes in the internal environment, especially in the organizational structure"[3:12] also design [4] and in the communication context it is reflected in customer-driven marketing environment and "relationship-based marketing", which show the changes of marketing and communication.

This article aims to discuss the role of communication for 21st century strategic management and argue that in 21st century communication must be not only implemented in all levels of management, but also strategically oriented internally and externally to effectively maintain the image of innovative organization.

## 1. The Context

Business environment in the 21st century is constantly affected by the development of digital technologies and wider processes that are shaped by economy. According to Robert M. Grant, business strategy in the 21st century is affected in three main ways: how companies are flexible and responsive, how they develop strategies for competing in a broader market space and what are the responsibilities of corporation and role of management [5:444]. Digitally driven knowledge revolution raised the questions about new directions in strategic thinking and uniqueness of every company that must create a unique combination of strategic variables and management practices [5:449]. This also is needed to gain competitive advantage in the marketplace, because focusing on strategy fundamentals does not necessarily lead to simple strategies and in many industries increased competition requires that established players access new sources of profitability [5]. It is relevant to raise a question how to create this new source of competitive advantage?

Since the success of any organization depends on "the effective and efficient use of communication by its workforce" [6:7], it is paramount to emphasize that without communication it would not be possible to plan, lead, organize and control the resources of organizations [7]. Consequently, communication in 21st century has faced variety of transitions influenced by rapid globalization and technological innovations that shapes how organizations plan their communication and are led, organized, and controlled by their management.

According to Kitchen and Burgmann, the multiplication of media, demassification of consumer markets, and the value of the Internet in today society are just three of the areas in which technological innovation has impacted communication [8:1]. A global marketplace became more transitory through the internet and that may have led to a customer-driven and focused marketing environment. It also means that "demassification of markets and outside-in approach most customers make their purchase decisions based on information of perceived value and not based only on the information a company chooses to present to them "[8:4].

On the other hand, organizations started to pay more attention for communication measurability rather than its planning. Moreover, it was noticed by M. Reid, S. Luxton and F. Mavondo that "in response to concerns about the impact of hostile marketing environments on brand equity and increased management expectations related to marketing performance and accountability, many organizations considered how to improve the management and integration of their marketing communication programs using integrated marketing communication (IMC) "[9:110].

Therefore, it is vital to notify that in this age of increased and fierce competition in the global marketplaces focusing only on external communication and its monitoring and measurability might

not be enough for the companies to achieve better position in the market or bring new innovations. According to Drucker, core competencies are different for every organization, but every organization—not just businesses—needs one core competence: innovation [10:119]. One of the new ways of communication management is connected not only to communication strategy and strategic communication, but additional tools such as IMC as well. Implementation of IMC in organizations needs to be taken into consideration, because it requires the involvement of the whole organization and its agents from the chief executive downward and it needs consideration from the highest corporate strategic level down to the day-to-day implementation of individual tactical activity [9:11]. Consequently, constant learning, training and adaptability to changes should be enhanced in organizational communication context.

## 2. Communication in Organizational Context

There are different descriptions of communication as a process. According to Onyesom and Onyesom, communication is "a process of generating, transmitting, receiving and interpreting messages in interpersonal, group, public and mass audience contexts through written and verbal formats" [6:8]. Another description stresses that communication is a process of information, ideas, and opinions exchange within, between or among individuals, groups, organizations, nations in a social context [11].

The latter concept leads to organizational context where more specific group of people is involved in maintaining communication. That specific group is usually employees of the organization and even though in early 2000s organizational communication was seen as internal communication which involves the transactions between individuals and groups of various levels and in different areas of specialization and those transactions which are intended to design and redesign organizations [12], later, it has been observed that most companies focus more on producing communication plans for external use, like the development of marketing strategies and often "internal communication is not addressed in the same way" [13:21].

It is also vital to mention that in the corporate setting, the communication activities are usually targeted to the company's internal and external audiences, employing various methods of communication appropriate for each, therefore, companies form specialized departments to manage their communication [13:21]. If we are taking into consideration organization's strategic objectives and how they are realized including consequences, communication "needs to strike a balance of both positive and negative consequences based on objective measures" [14:95]. It means that certain alignment and coordination between business units responsible for strategy execution is needed.

Communication becomes crucially important during changes within organization and its management and according to Voehl and Harrington, "you begin at the top, create an enrolment plan, shift people's paradigms toward alignment, and have the leadership team communicate and demonstrate what it will take for everyone (themselves included) collectively and personally to succeed" [14:95].

Moreover, there are at least 10 managerial roles [15:7] where the importance of communication for leadership and management can be recognized. There is information processing, interpersonal and decision-making roles. Looking more detailed into these roles you can find the roles of spokesperson, liaison, disturbance handler, negotiator and more [15]. These roles are attributed both to leaders and management, because "managers will need to learn how to lead while leaders have to learn how to manage" [15:8]. Consequently, the interrelation between internal and external communication and acknowledgement and importance of both for management must be highlighted.

## 3. Importance of an Image for Company's Long-Term Communication Orientation

Specifically, Vladutescu [16] argued that communication is especially needed in enterprises to:

a) establish and disseminate the goals of an enterprise;
b) develop plans for their achievement;
c) organize human and other resources in the most effective and efficient way;
d) select, develop and appraise members of the enterprise;
e) lead, direct, motivate and create a climate in which people want to contribute;
f) control performance;
g) inform management about regulatory activities of government;
h) be aware of the concerns of the immediate community;
i) reduce incidence of industrial unrest if properly managed;
j) establish and disseminate the goals of an enterprise;
k) develop plans for their achievement;
l) organize human and other resources in the most effective and efficient way;
m) lead, direct, motivate and create a climate in which people want to contribute;
n) control performance.

There are more detailed factors that shape everyday communication in enterprises, but mostly they are attributed to management as their responsibility. Vitorrio Coda emphasized the role of communication that has impact for the firm's ability to generate sustainable profits, that is based on competitive success and the consensus of different stakeholders (other than customers) and maintaining competitive success and consensus that are key ingredients for fuelling, rebuilding, and growing the technological, commercial, financial, organizational, human, and social assets of the firm [17:146]. Internal and external communication plays a key role "in producing intangibles which enhance the attractiveness of the offering of goods or services to customers or the offering of collaboration to a given stakeholder; in selling these offerings to interested stakeholders; in crafting and disseminating a powerful, attractive image (for the company and/or the product). As this image paves the way for commercial and social consensus outside the firm, it fuels a positive pressure to work for the functionality and development of the firm" [17:146].

Communication and its management in the organization can be oriented towards short-term or long-term perspective and this is also crucial while bridging theoretical connection between strategy and communication.

From a short-term perspective communication is implemented and managed fragmentally and episodically without unifying goal and business strategy which links all communication activities and express companies' functionality and sustainable development. While long-term perspective is a basic strategic orientation transformed into integrated management of myriad communication activities which are mutually coherent over time and which aspire to build and enhance a powerful and attractive corporate image [17:151]. According to Coda, company's image is essential when communication is managed with a long-term orientation ([17:151].

Therefore, how the company's image is being constructed, maintained, and represented to the customers really depend on the

efficient management of internal and external communication, which should be reflected in communication strategy that later is integrated in company's business strategy.

## 4. Organization's Image and Position of Brand Communication

Since the image of a company is reflected through the brand, IMC as a communication management tool can help to strategically manage all brand messages that could build its strong position in the market. IMC is also "centered on building and leveraging customer and consumer interests and relationships" [9:12]. Following Reid, Luxton and Mavondo approach [9], the strategic dimension of IMC can be grouped under five broader dimensions - see Table 1.

Customer and stakeholder connectivity already shows the importance of IMC and putting customer first that enhances customer connectivity and organizational responsiveness. While strategic consistency is more related with the brand entity that all parts would send a message to customers and other stakeholders from whatever source, and they should be consistent in order to protect the brand image [9:14].

Cross-functional integration puts attention to the internal and external integration of the organization. It means that top management needs to be involved to drive the process and at the same time internal marketing must be managed in order to provide the foundation for effective IMC planning and reporting [9:15]. If we look more into external communication and especially how the brand can help communicate the image, brand orientation "seeks to add value to an existing or new product or service to give it a competitive advantage and a reason for customers to choose it" [9:16].

Therefore, IMC can be used to form brand distinctiveness and value, which later transforms into an image that influence the consumer and the way he sees it. Nevertheless the value "is increasingly being created outside the physical product by such factors as interactions between the customer and organization, responsiveness to complaints, and customer needs and expectations" [9]. It leads to effective internal and external communication, which is important not only for marketing departments, but to all employees that represent the brand and the company.

Since more and more processes have to be integrated for consistent management and communication of company's image, not only information, knowledge and changes should be managed and communicated internally (variety of authors have emphasized the concepts of knowledge management, learning organization, leadership communication and change management, which are connected to the strategically oriented organization – for instance, Refs. [18-20,14], but communication strategy and strategically oriented communication could help to orient organization culture, communication policy and shape everyday processes, which could bring innovations and maintain organization's efficiency.

It is important to emphasize industry lifecycle stages that shape every organization's management decision related to strategic goals and their implementation. Depending in which stage the company is, whether it is pioneering (fragmentation), shakeout, maturity or decline, different strategic goals must be set. For instance, if the company is facing the pioneering stage, which is characterized by low market entry barriers when many small start-up firms start competing for the exploitation of business opportunities, it can already prepare internal and external communication strategy, which can be integrated into business strategy and start pursuing strategic goals. On the other hand, this stage is followed by a shakeout when barriers-to-entry begin to rise and "mergers and acquisitions become more common as aspiring market leaders pursue growth and fight for an increased market share by establishing larger operations that can yield scale efficiencies and increased profitability" [21:15], therefore, previous strategy has to be updated and changed according to new market situation.

## 5. Importance of Strategic Communication

Since digital technologies are rapidly developing and user-driven marketing is still very important, there are questions how future companies would look and operate in this environment. For instance, Zanda highlights wide range of networks that future modern companies will have to focus on, because relationships create opportunities for development and economic success [22:178]. Also she emphasize how this very different environment requires new policies and marketing techniques, because fundamental orientation of marketing tends to change: "it is no longer focused on the product, market share and the need to sell a particular product to global consumers, but on the need to create a lasting commercial relationship with the single consumer, to whom various products or services must be sold over time" [22:179]. Now the fundamental objective is to develop lasting relationships with the individual consumer and encourage her/him to purchase a wide range of products or services that could be called "relationship-based marketing".

This type of commercial relationships constitutes the basis for creating value. Zanda explains more that, "according to the new orientation, it is no longer so necessary for companies to stimulate masses of consumers to purchase to increase total market share. It is more important to develop and put into practice corporate policy aimed at getting to know individual customers, binding them to the firm and its business and creating loyalty to build stable relationships that are reciprocally advantageous. In this way, the company enters the life of customers, supports them, follows them and, gradually, even becomes pervasive in their lives by sharing values and behaviours" [22]. Therefore, the brand is of paramount importance.

To coordinate communication internally and externally and make the brand stronger and maintain the image of the company it is crucial to establish a certain system where communication would be planned and strategically oriented since "organizations that have effective communication systems are known to be more e ective than organizations with poor communication systems" [23:21].

For example, Barnard mention that, the organizational structure has influence for the planning process [24:59]. Moreover, "system must operate in a coordinated manner so that decisions and opera-

Table 1. Strategic and Tactical Characteristics of Integrated Marketing Communication (IMC). Adapted according to Ref. [9:14]

| Strategic | Tactical |
|---|---|
| Driven by market-based assets and financial expectations | Campaign-level consistency |
| Customer and stakeholder Connectivity | Campaign-level clarity |
| Strategic consistency | Campaign-level coordination |
| Cross-functional integration | |
| Resource commitment for IMC | |

tions at the various organization levels (from bottom to top) are coordinated and thus consistent with each other and directed together towards the ultimate goals of the company system" [24]. It shows that certain strategy is required and as P. Drucker would say, "every organization operates on a *Theory of the Business*, ... , that is, a set of assumptions as to what its business is, what its objectives are, how it defines results, who its customers are, what the customers value and pay for. Strategy converts this *Theory of the Business* into performance. Its purpose is to enable an organization to achieve its desired results in an unpredictable environment" [10:43].

Organization structure shapes its internal and external processes, accordingly, not only business strategy has to be established and implemented, but communication strategy as part of business strategy as well. V. Coda defines communication strategy as it "should unify and integrate customer and employee needs, the demand for profitability and environmental protection, and so on. Such a strategy translates into events and messages which combine and converge, capturing the public's attention and building consensus around the firm's strategic plan" [17:165].

Following Coda, communication strategy has certain objectives, which are: awareness (modification the attitudes of certain publics), mobilization (engagement of personnel), commercial (undertaking initiatives that target customers in some way), disseminating reputation (carrying out initiatives to increase recognition of the company or brand in new markets) [17:165-166]. These objectives and initiatives are followed by the combination of four elements that are: the messages to convey, the publics to address, the means or channels to use to reach target audience, and the timing to carry out a given initiative [17:165]. Once again, these combinations presumably lead to external communication planning, therefore, possibility to unite external and internal communication should come from the connection of explicit communication activities with the corporate/business strategy. It means that the objectives and prerequisites of communication strategy and business strategy must be properly communicated, along with the results it promises and the commitment it requires [17:174].

If the companies are trying to maintain new ways of communication and relationship with clients, the role of management in determining and integrating communication strategy into general strategy of the firm is essential. Coda offer these modes of management that can be useful while managing communication strategy [17:174]:

1) be more or less rooted in a solid strategy and shaped by transparency and coherence;
2) use communication mainly for contingent objectives and/or primarily for strategic purposes;
3) consume image and other intangible resources accumulated in the past or, on the contrary, strive to enhance and enrich them;
4) be an expression of management aware of its role and of the professional skills necessary to play this role effectively today.

Without management being involved in all levels of maintenance of communication strategy, there would be less possibilities for the employees to understand communication policy and it would affect the organization's culture in a sense of brands and image representation to the target audience. If it is not clear, what are you representing and how should you maintain the image of the organization it sooner or later transforms into everyday communication. Since user-driven marketing environment has the influence for the company's sales and economic performance and the impact of ICT's for communication might continue to rise, strategy and communication should go along and strategic communication should not be oriented only towards external, but internal goals too. This market situation and future perspectives to develop organizations raise the role of strategic management to consider how to reach better efficiency and control rising challenges with strategic approach to communication.

## Conclusions

Information and communication technologies are shaping not only everyday communication, but the environment of organizations and markets that they operate in. The role of communication in 21st century is closely connected not only with internal and external communication, but the outcome it brings for successful networking between clients, stakeholders, and businesses. Even though networks started to be important already in the beginning of 2000s, but their relevance might be more important in future development of organizations, because relationship-based marketing brings the possibilities to engage customers and employees of the organizations in a new type of communication that can affect organization's brand and image.

Since marketing is adapting to relationship-based mode between clients and companies it brings the question about internal and external communication interrelation and its impact for the image and brand of organizations that create products and provide services. To bring innovations and maintain the image of the company it is necessary to orient both internal and external communication towards strategy and implement strategic communication within all levels of management. The task for strategic management is also related with building up strong connection between communication and strategy and later integrate it into business strategy.

## Abbreviations

| | | |
|---|---|---|
| ICT | - | Information and Communication Technologies |
| IMC | - | Integrated Marketing Communication |

## References


1. Castells M. (2001) The Rise of the Network Society. – Wiley-Blackwell.
2. Castells M. (2000) End of Millennium: The Information Age. – *Economy, Society and Culture* Vol. III.
3. Buble M. (2015) Tendencies in Evolution of 21st Century Management. – *Management* Vol. 20, 2015, Special issue, pp. 1-17.
4. Strikwerda J. (2012) Organization Design in 21st Century: From Structure Follows Strategy to Process Follows Proposition. – Amsterdam Business School.
5. Grant Robert M. (2005) Contemporary Strategy Analysis: Concepts, Techniques, Applications. A Guide for Instructors. – Malden: BlackWell Publishing.
6. Moses Onyesom, Anthonia Ewere Onyesom (2015) Dimensions of Communication in the 21st Century Organizations: A Conceptual Review. – *Journal of Language and Communication* Vol. 2, No. 1, 7-20.
7. Conrad D. Newberry R. (2012) Identification and instruction of important business communication skills for graduate business education. – *Journal of Education for Business* 87: 112-120. DOI: 10.1080/08832323.2011.576280.
8. Kitchen P., Burgmann I. (2010) Integrated Marketing Communication. – Wiley International Encyclopedia of Marketing.

9. Reid Luxton, Mavondo (2005) The relationship between integrated marketing communication, market orientation, and brand orientation. – *Journal of Advertising* 34:4, p. 11-23.
10. Drucker P. F. (1999) Management Challenges for the 21st Century. – PerfectBound.
11. Udeoba A.A. (2012) Effectiveness of organizational communication in the global age. – *Mediterranean Journal of Social Science* 3(15), 31-41. Doi:10.5901/mjss.2012.v3n15p31.
12. Dolphin R.R. (2005) International Communications: Today's Strategic Imperative. – *Journal of Marketing Communications* Vol. 11 No. 3 pp. 171-190.
13. Dianne Elaine L. Ozaeta1 , Imelda L. An2 (2014) Use of Internal Communication Tools among AG&P Employees. – *Journal of Education and Literature* Vol. 1, No. 1, 20-30.
14. Voehl Harrington (2016) Change management. Manage the change or it will manage you. – *CRC Press.*
15. Harrison Christian (2018) Leadership Theory and Research. A Critical Approach to New and Existing Paradigms. – Palgrave Macmillan
16. Vladutescu S. (2013) Three diachronic paradigms of communication. – *Internationa Journal of Education and Research* 1(12), 1-6.
17. Coda V (2010) Enterpreneurial values and strategic management. – Essays in management theory. Palgrave Macmillan.
18. Edmund Pawłowski, Stefan Trzcieliński, Hanna Włodarkiewicz-Klimek, Joanna Kałkowska in Lewandowski J., Jałmużna I., Walaszczyk A. [eds.] (2011) Organisational Structures in Knowledgedbased Enterprises In Contemporary and future trends in management. Monograph. – Technical University of Lodz, Lodz; ISBN: 978-83-7283-431-7; 107-124.
19. Suffian Hadi Ayub, Norhayati Abd Manaf, Mohammad Rezal Hamzah (2014) Leadership: Communicating Strategically in the 21st Century. – *Procedia - Social and Behavioral Sciences* 155, 502-506.
20. Dike Odiwe, Ehujor (2015) Leadership and Management in the 21st Century Organizations: A Practical Approach. – *World Journal of Social Science Research* 2(2), 139.
21. Cunningham Stuart; Silver, Jon (2013) Screen Distribution and the New King Kongs of the Online World. – Basingstoke: Palgrave Macmillan.
22. Zanda G. (2012) Corporate management in a knowledge-based economy. – Palgrave Macmillan.
23. Miller P. (2012) Leadership Communication – the three levels. – In: Today's Manager. – Singapore Institute of Management, Issue February-March, p 19-21.
24. Zanda G. (2018) Building Efficient Management and Leadership Practices The Contemporary Relevance of Chester I. Barnarda lations Thought in the Context of the Knowledge-Based Economy. – Springer International Publishing.

*Article*

# Application of Evolutionary Artificial Intelligence. An Exploratory Literature Review

Nijole Maknickiene[1,2*]
[1] Department of Business, Vilnius Business College, Kalvarijų g. 129-401, LT-08221 Vilnius, Lithuania
[2] Department of Financial Engineering, Faculty of Business and Management, Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania
[*] Corresponding author, E-mail: *nijole.maknickiene@vilniustech.lt*

Received: 01 May 2022
Accepted: 11 June 2022
Online: 31 August 2022

**Abstract.** Evolutionary processes found in nature are of interest to developers and practitioners of artificial intelligence because of the ability to optimize, detect, classify, and predict complex man-made processes. Evolutionary artificial intelligence (EAI) is examined from various perspectives to evaluate the main research directions and the trend of the decade. Co-occurrence networks were used to visualize data and find key sub-themes in a dataset consisting of article titles. The literature review covers the following aspects of EAI applications: methods, detection, data, approach, and colony. The resulting co-occurrence networks show a huge increase in diversity in research methods, data and function application possibilities, and approaches. Although simulating the behaviour of colonies is not as popular as it was a decade ago, the scope of applications for known algorithms has not been diminished.

**JEL:** C6; C8.

**Keywords:** colony; co-occurrence network; detection; differential evolution; evolution; multi-objective optimization; swarm intelligence.

**Citation:** Nijole Maknickiene (2022) Application of Evolutionary Artificial Intelligence. An Exploratory Literature Review. – *Applied Business: Issues & Solutions* 1(2022)22-31 – ISSN 2783-6967.
https://doi.org/10.57005/ab.2022.1.4

## Introduction

The advent of programming languages has created new opportunities for human expression and communication with machine. Mathematical logic can be supplemented by logic created from nature using programming language operators, loops, objects, and so on. Evolutionary processes in nature characterize change in the heritable characteristics of biological populations over successive generations and ensure biodiversity. Why did nature choose such a goal? Why not copy/paste? It would be a simple algorithm to replicate the most perfect population and its behaviour; in an ever-changing world, however, a biological species must adapt to survive. Imitation of evolution by artificial intelligence is the main object of investigation.

The purpose of this article is to classify and evaluate different aspects of simulating evolutionary processes in scientific articles of the Scopus database [1]. One database contains not all scientific works on a certain topic, but rather the unified management of the bibliography of articles; the search system provides equal opportunities for selected articles to enter the author's field of interest. The first section of this paper discusses the methodology of finding main aspects in certain topic of articles. Text analysis and co-occurrence network algorithms are used for analysis and visualization. The second section describes the data found in the Scopus database. The data includes the titles of articles from 2021–2022 in the Scopus database, sorted by keywords 'application', 'evolutionary', 'artificial', 'intelligence'. This data was compared with data from 2011–2012 with the aim of identifying trends in scientific research topics. The third section includes a literature review from five different perspectives.

## 1. Application of textual analysis to review

For exploratory literature review, we selected a co-occurrence network [2]; this visualisation method is typically used for investigation of social networks. Research directions, processes and changes are also analysed using the networks. Notably, Rajita et al. [3] used machine learning to explore the collaborative relationships among researchers with the goal of predicting changes in research topics. Classical collaborative distance (CCD) and refined classical collaborative distance (RCCD), together with other community-specific estimates, are sufficiently accurate predictors of change in studies. Networks of co-authorship of researchers in different disciplines were drawn by Newman [4]; Zhang et al. [5] analysed the ethical issues of artificial intelligence using the co-authorship network method. An article by Elmsili and Outtaj [6] inspired us to use co-occurrence networks to select scientific papers on evolutionary artificial intelligence for review.

Titles of articles during the two periods were taken for the study with the aim of visualizing trends in selected research aspects. The algorithm organizes the text, removes stop words, and creates a bag-of-word matrix using a model. This action allows us to assign certain numerical characteristics to words, such as the frequency of repetition of the word. Counting word co-occurrence estimates relationships between words. Graphs model relationships between words visually: nodes are assigned to words, and edges are assigned to connections. The strength of the connection between the nodes is determined by the weight and thickness of the line. The code presents a limited number of nodes and connections in the final visualization because the comprehensive graph is too confusing and uninformative, so it allows the user to select a keyword in the set and find its neighbours.

## 2. Overview of data used in the study

We selected articles exclusively from the Scopus database [1] for review to ensure that the data would be selected and classified in the same way. For 2011–2012, the database returned 260 articles; for 2021 and approximately the first half of 2022, it returned 206

Table 1. Distributions of articles according to scientific field.

| Scientific Field | Part, % 2011–2012 | Part, % 2021–2022 |
|---|---|---|
| Computer Science | 49 | 35 |
| Mathematics | 23 | 15 |
| Engineering | 15 | 26 |
| Decision Sciences | 2 | 5 |
| Biochemistry, Genetics, Molecular Biology | 2 | 4 |
| Energy | 2 | 3 |
| Social Sciences | 2 | 4 |
| Agriculture and Biological Sciences | 1 | 1 |
| Environmental Sciences | 1 | 3 |
| Physics and Astronomy | 1 | 5 |

articles by selected keywords: *application*, *evolutionary*, *artificial*, *intelligence*. The distribution of articles by assigned topics is presented in Table 1. The last ten years have seen an expanded application of Evolutionary Artificial Intelligence (EAI) in a variety of fields. Ten years ago, research was more focused on computer science, mathematics, and engineering, whereas in recent years, application in almost all other fields of science has doubled. This trend indicates that newly developed models of EAI are being successfully applied in various fields.

*Evolution* represents the term that describes the unique object of a selected article that unites all aspects of Evolutionary Artificial Intelligence. The network of titles of scientific articles (Fig. 1) obtained by Matlab code [2] revealed that the centre contains nodes with higher weights close to the keywords: *evolution* (52), *intelligence* (31), *based* (30), *artificial* (26). Important nodes are found farther from the centre: *differential* (26), *method* (20), *detection* (24), *research* (23), *analysis* (20), *data* (20), *optimization* (20), *colony* (12), *approach* (17).

In addition to the most popular words in the centre, Fig. 1 shows three larger and two smaller wings of word combinations, each representing different aspects of the topic.

1. The first large group covers method and is mostly related to the word 'differential'. It is easy to infer that many scientists are developing new (or improving upon known) methods of evolutionary artificial intelligence (EAI). The differential evolution method is one of the most popular in this field.
2. The second large group combines articles with 'feature' and 'detection'. EAI is usually associated with optimization, so detection is an interesting area of research.
3. The third large group of articles is related to 'date', 'research' and 'analysis'. Data is used in every study, but in recent years new aspects of data use, security and analysis have emerged.
4. The first smaller group deals with 'approach'. This finding was surprising and made us look at EAI as a heuristic or even meta-heuristic application.
5. The second smaller group is related to 'optimization' and 'colony'. This aspect is closely related to the imitation of evolutionary processes in nature.

The resulting network became the basis for the structure of the literature analysis. It is important to note that in the initial phase of the study, more word networks were obtained with the keywords 'application', 'evolutionary', 'artificial', 'intelligence' and others. The information in these networks is interesting, but it remains insufficiently clear.

## 3. Literature review

### 3.1. Methods

Co-occurrence network by the word 'methods', presented in Fig. 2, clearly shows that in ten years, the variety of application of methods has expanded significantly. Thicker lines in the 2011–2012 graph show two popular methods: differential evolution and ant colony optimization; in 2021–2022, more methods are used, which already need to be classified.

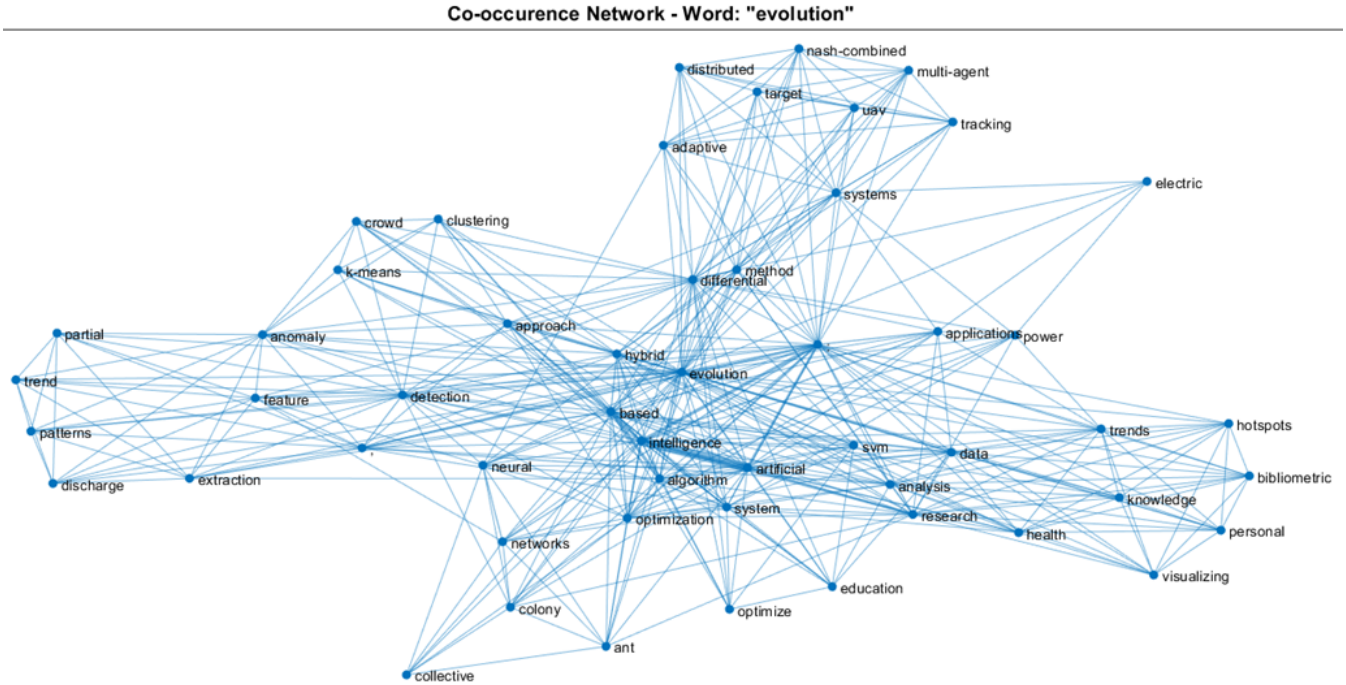



Fig. 1. Co-occurrence network for 'evolution'.

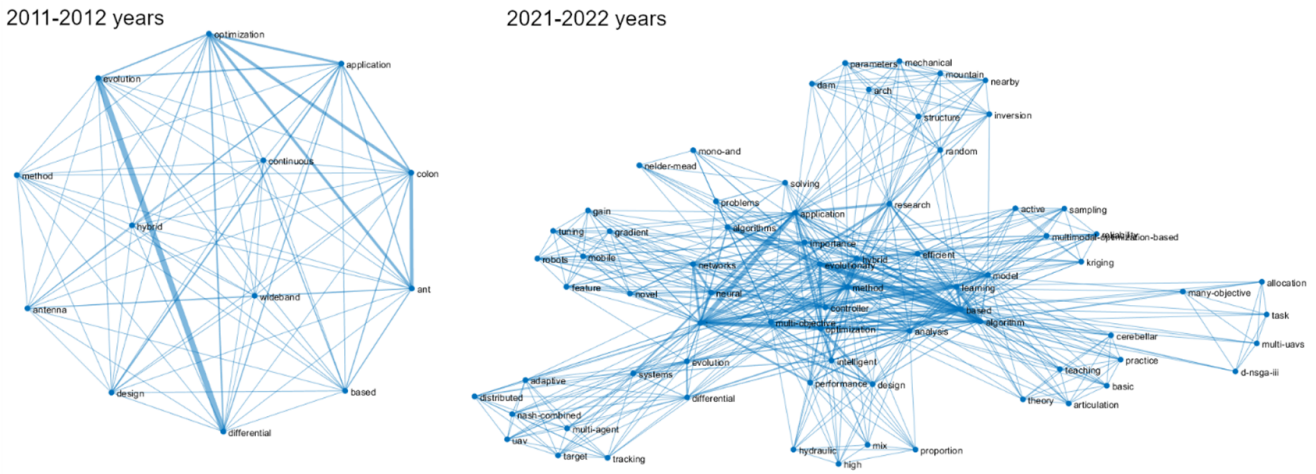



Fig. 2. Co-occurrence network for 'method' (left: 2011–2012, right: 2021–2022).

Evolutionary computation is a subset of artificial intelligence that is most commonly used for optimization by extracting certain variances from large data sets using several important estimates. These are usually genetic algorithms, evolutionary or genetic programming, or swarm intelligence. Because all these methods are heuristic, researchers are constantly striving to improve them and adapt them to new problems. Ahli et al. [7] combined three evolutionary computational methods – global optimization methods (GOM), the genetic algorithm (GA), and particle swarm optimization (PSO) – and applied them to hydrological modelling. Differential evolution (DE) is a method that solves the problem of optimization, in the form of multiple iterations, in order to improve a possible solution. DE first was proposed by Storn and Price [8] and outperformed other optimization algorithms, because it is a very simple and straightforward strategy. DE is one of the most popular evolutionary methods due to its easy adaptation, simple design, low number of parameters and fast operation. Jebaraj [9] adapted DE to address energy issues, namely: reactive power planning, congestion management, transmission capability, cost-effective load shedding, generating equipment commitment, power flow optimization, and optimal reactive power supply. Yan et al. [10] adopted this method for the design optimization of x-ray source beam optics. Long and Gao [11] studied an artificial intelligence training system consisting of training need, training data, training characteristics, model, and application to training.

In that research, DE initiates the population and then uses mutation, crossover, and selection to obtain the optimal combination of SVM parameters. In recent years, researchers have sought to combine genetic algorithms with other known methods with the goal of increasing the efficiency of solutions, the speed of calculations, and adapting them to new data sets. The resulting hybrids facilitate solving a wider range of problems. Liu et al. [12] combined the DE method with the K-mean classification algorithm to identify anomalies in artificial crowd intelligence. The differential evolution of parameter adaptive schemes is known as adaptive differential evolution (ADE). Another hybrid was proposed by Yu et al. [13]; it combines the ADE algorithm with Nash optimization, and then proposes a Nash-combined ADE method for unmanned aerial vehicle (UAV) systems. The MAP-elite algorithm looks for the best solution in a large space by evaluating certain criteria at each point in the space. Choi and Togelius [14] merged DE with MAP-elites, thereby obtaining higher-quality solutions.

In real life, it is not enough to have a clear goal when making decisions; oftentimes the decision-maker must compromise. Sometimes the goals contradict each other, and sometimes optimization results only in a set of 'good-enough' decisions, and the choice of the result depends on the decision-maker. In optimization tasks, this problem is solved by classifying the tasks into single-objective optimization (SOO) and multi-objective optimization (MOO). The use of evolutionary algorithms in combination with SOO and MOO algorithms has allowed Lambrinidis and Tsantili-Kakoulidou [15] to increase their chances of success in developing new, more effective drugs. For the distribution of various unmanned aircraft tasks, a multi-objective evolutionary algorithm named D-NAGA-III was proposed, which effectively optimizes military tasks. Boukhari et al. [16] improved the hybrid approach of evolutionary strategies and multi-objective optimization to accelerate the speed of convergence and applied it to two-objective portfolio optimization.

A new method of MMO was proposed by the Nie and Luo [17]. They did not aim to simplify the multi-objective task; rather, they developed a method for finding solutions in parallel. The complex model of rare events can be solved by the complex model proposed by Wang et al. [18], known as evolutionary multimodal-based multi-objective optimization (EMO-MMO), which recognizes the most probable failure points. For antenna engineers, alignment can be greatly simplified using the methodology proposed by De Melo et al. [19], which incorporates the non-dominated sorting genetic algorithm (NSGA-II) and multi-objective evolutionary algorithm based on decomposition (MOEA / D) to easily reconcile several objectives. In recent years, more scientific articles have been written in the field of computer science than in any other field (Table 1). Such articles usually involve the development of methods or methodologies that combine the construction of hybrid models with known models.

## 3.2. Detection

There is also a clear difference in the variety of research aimed at detecting certain properties using EAI between a decade ago and now (Fig. 3).

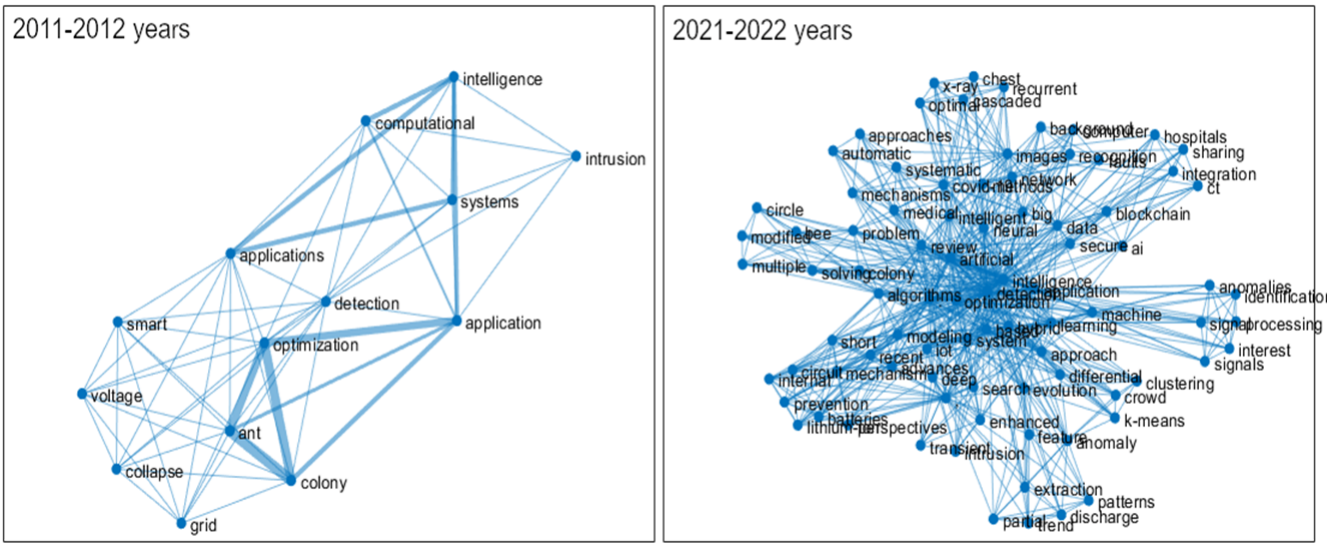



Fig. 3. Co-occurrence network for 'detection' (left: 2011–2012, right: 2021–2022).

Detection is the specific cognitive ability of a person to recognize useful information in the flow of information. This operation is difficult to describe with mathematical formulas or statistics, although a large amount of data is available in this case. Artificial intelligence algorithms are successfully applied to detect near other imitations of human cognitive functions. An important area of application for detection algorithms is medicine, where early detection of disease is very important. Kumar et al. [20] proposed how to identify cancer-affected regions at an early stage in magnetic resonance imaging. This work combined deep learning with a blockchain and applied the bat algorithm. The benefits are early cancer detection and easy information sharing without compromising client privacy.

Furthermore, the global COVID-19 pandemic has challenged doctors to quickly diagnose the virus. Shankar et al. [21] proposed a barnacle mating optimization (BMO) algorithm with a cascaded recurrent neural network (CRNN) model named BMO-CRNN, which can detect a virus from chest x-ray images using an algorithm that simulates the barnacle life cycle from eggs to adult life. The proposed BMO-CRNN model detects COVID-19 with an average accuracy of 94.82%. Afza et al. [22] proposed a more accurate and faster method for detection skin cancer by incorporating a hybrid deep features selection (HDFS) method in a classification algorithm.

Detection is also very important in identifying fraudulent behaviour. The detection of cyber-attacks in the field of the Internet of Things (IoT) has been investigated by Fatani et al. [23]. Their transient search optimization (TSO) generates a population, and operators such as recombination and mutation of the differential evolution (DE) algorithm are used in conjunction with convolutional neural networks that act as extractors of certain properties. The result of this combination of algorithms is a significantly improved accuracy in classifying fraudulent events. A multi-step process consisting of an evolutionary bag-of-ngram approach, a genetic algorithm (GA), and a convolutional neural network (CNN) is used to detect malware in cloud computing [24].

The action of detection is related to rare events, also called anomalies. Florkowski [25] proposed segmentation techniques for feature extraction, anomaly detection, and trend evolution using convolutional neural networks for the detection of coherent forms in the images. With the help of business analytics, companies, suppliers, and customers are connected into a single system where information and data are shared. It is therefore important that computer networks run smoothly. Computer network failure detection is the focus of a study by Ge [26], who improved particle swarm optimization (PSO) and merged it with radial neural network function (RBF). Improvements in evolutionary algorithms have resulted in good (89.3%) accuracy as well as shorter computer network failure detection times.

Business analytics also seeks to extract information from a huge stream of textual data. Textual and sentiment analysis allows us to measure customer or employee opinions and make more useful decisions. Erfanian et al. [27] used textual data from Twitter for the purpose of detecting certain events and their evolution on the social network. Their proposed evolutionary approach involves two steps: in the first stage, events are detected in the text, and in the second stage, changes in the terms associated with the events are detected. Social networks are also of interest to researchers as a source of community dissemination and exchange.

The mechanism of community discovery on social networks is combined with community change in the work of Rajita et al. [28]. In their work, community change is detected as an event using the Cuckoo search algorithm, and it involves three steps: detection of the communities for each timeframe, identification of a proper fitness function, and computation of the similarity and events. Knowledge of social networking communities can be successfully applied in marketing, politics, disaster recognition and management, and more.

EAI algorithms improve solutions to field-specific problems, such as the circle detection problem in images by using Bee Colony algorithms [29], detection of internal short circuit (ISC) within lithium-ion batteries [30], and detection of students who complete their undergraduate studies on time by predicting student's attrition rates [31]. In conclusion, it can be said that EAIs are not only optimization algorithms but are also increasingly applied to detection.

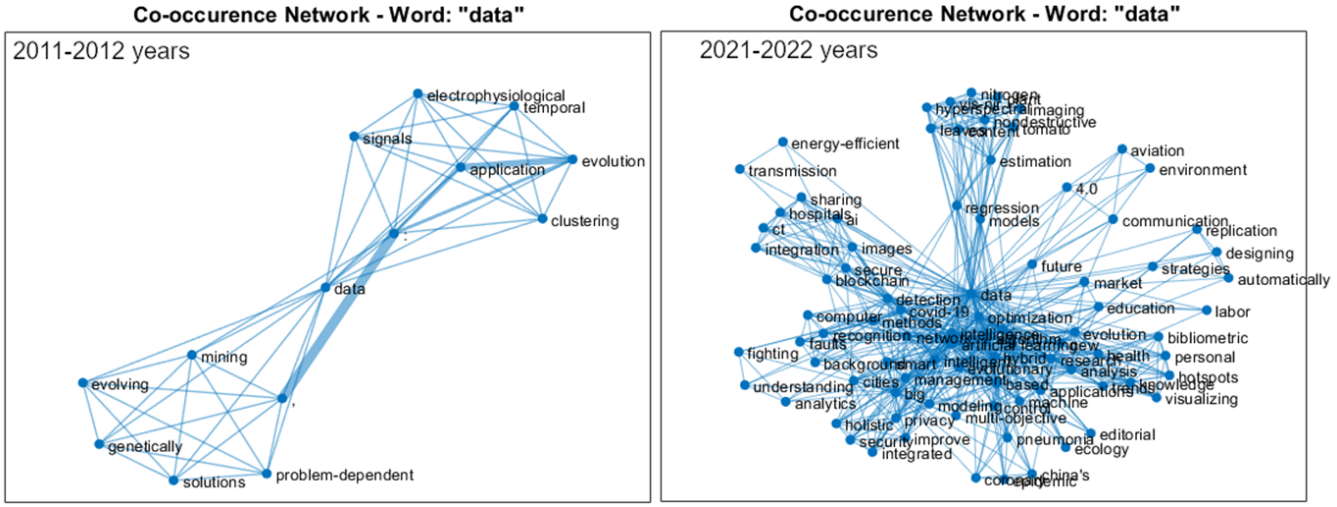



Fig. 4. Co-occurrence network for 'data (left: 2011–2012, right: 2021–2022.

## 3.3. Data

In the scientific articles of 2011–2012, data were analysed in two directions: data extraction and application. In 2021–2022, we even see several clusters of nodes (Fig. 4) that are associated with different methods, new technologies and novel application possibilities.

In each study, the choice, availability, or lack of data plays an important role. In the application of EAI, the data determine not only the application sector but also the heuristic nature of the solutions. The growing use of big data for community purposes is also creating new challenges, such as data integrity, privacy, security, and reliability. Data usage issues are particularly acute in smart city applications. Chen et al. [32] used evolutionary algorithms to solve data management problems, adopting good data availability and process optimization practices from big data analytics. The use of big data faces data transfer challenges such as security and resource savings. Chaitra and RaviKumar [33] solve these problems by creating a model and incorporating evolutionary algorithms into it. Secure data sharing is also very important in medicine. Kumar et al. [20] combined blockchain technology with convolutional deep learning algorithms, and applied data pre-processing techniques to the data, which include the bat algorithm and data augmentation technique. This adaptation made it possible to eliminate noise and improve the efficiency of the calculations. This study is useful for hospitals, testing labs, research centres and other institutions.

In computer networks, where countless requests occur simultaneously, reliability is very important in data distribution systems. Bokhari and Theel [34] solve this problem by using evolutionary algorithms that make it possible to reduce cost without significantly reducing data availability. Data communication, data transfer and data security are very important in building the aviation of the FUTURE 4.0 Project. These data usage issues have been addressed by Sekera and Novák [35], who reviewed the possibilities of using new technologies. The use of personal health data (PHD) in the three aspects of medicine, computer science, and management has been examined by Gong et al. [36]. Although their article examines how knowledge about PHD has evolved over the past 20 years, it remains relevant to the development of new safe PHD technologies. Scientific research uses not only numerical data, but also other forms of data such as pictures [21,37], hyperspectral imaging [38] and mapping [39-40]. In the information age, data challenges users to balance the cost of resources used and return opportunities, data privacy, security, and fast and convenient data availability.

## 3.4. Approach

Networks according to the word approach are shown in Fig. 5. In 2011–2012, the most popular approach was ant colony optimization, and in 2021–2022, we can identify several clusters: optimization, artificial intelligence, machine learning, proteins, and classification.

The optimization approach means that the solution method does not guarantee an exact solution with mathematical formulas, but rather provides a potentially useful heuristic approach to the optimal solution. Almosnino and Cappelletto [41] proposed an optimization system based on an evolutionary algorithm that allows the modelling of human motion functions. The obtained scenarios mimic various doctors' decisions and facilitate predictions of the course of treatment. The many layers transfer learning genetic algorithm (MLTLGA), developed and tested by De Lima Mendes et al. [42], is an evolutionary approach that has found application in medicine for the classification of pneumonia from chest x-ray images. The application of evolutionary and particle swarm optimization algorithms by Polkowski et al. [43] allowed them to obtain a better optimization approach during observation of query plans along with database servers. The machine learning approach is perceived as a rough learning from the data presented, lacking reference to a predetermined equation or model. Adeleke et al. [44] used compounds of genetic algorithms to determine seasonal changes in the composition of physical waste, which could be very beneficial for optimizing municipalities' waste management approaches, although the study was conducted only in South Africa.

The artificial intelligence (AI) approach includes the capabilities of computer algorithms such as acting humanly, thinking humanly, thinking rationally, and acting rationally. Nayeri et al. [45] classifies techniques based on artificial intelligence into three groups: evolutionary algorithms, machine learning algorithms, and combinatorial algorithms; they further divide the application of evolutionary algorithms in the fog computer into swarm intelligence, genomic

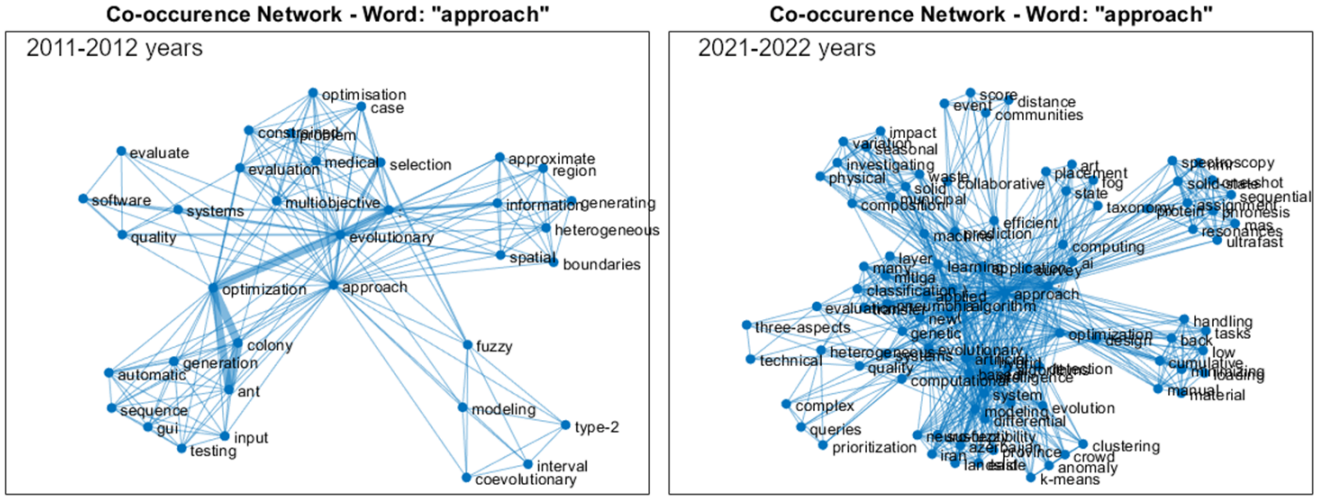



Fig. 5. Co-occurrence network for 'approach' (left: 2011–2012, right: 2021–2022).

mimetic and search based. The review article summarizes the research of many authors to solve fog computing problems and concluded that evolutionary algorithms have focused on providing efficient load balancing methods, whereas EAI solves fog computing problems best in conjunction with machine learning algorithms.

Abdollahizad et al. [46] modelled susceptibility to landslides by comparing three metaheuristic methods including grey wolf optimization (GWO), particle swarm optimization (PSO), and shuffled frog leaping algorithm (SFLA). The approach of artificial intelligence can successfully predict phenomena such as landslides.

Knowledge of microbiology has inspired scientists to create artificial intelligence systems that mimic processes in protein cells. Repecka et al. [47] developed an artificial intelligence network that generates new protein sequences from a complex family of amino acid sequences. Rives et al. [48] integrated the knowledge of protein sequences into a model based on unsupervised AI, which can convey information about the main properties of proteins. Gopinath et al. [49] performed multidimensional experiments by creating artificial multidimensional pulse sequences.

The classification approach is also associated with EAI. Czajkowski et al. [50] developed an evolutionary heterogeneous decision tree, whose application to cancer detection yielded relatively good accuracy. Afza et al. [22] used classification by extreme learning machine for detecting and classifying skin cancer.

## 3.5. Colony

Fig. 6 shows that ten years ago, the word 'colony' was significantly more popular than in recent years. The reason may be that many new algorithms simulating the behaviour of living things were developed ten years ago, and now researchers are already using a selection of these known and tested colony algorithms.

*Swarm intelligence* is a term derived from biology; it describes the collective behaviour of living things – specifically their self-regulation in optimizing the use of resources and in preserving and propagating the population. With the development of programming languages, logical operators can use not only mathematical logic but also logic created by nature. Particle swarm optimization (PSO), colony optimization (CO) and their variants are the most common heuristic descriptions of artificial intelligence algorithms [51], although colony simulation is not limited to optimization, also being used for detection and prediction. Kadkol [52] has explored the mathematical side of the PSO algorithm in depth. Mittal et al. [53] compared the characteristics and application possibilities of different PSO modifications and hybrids with other optimization algorithms. Chen et al. [54] conducted a comprehensive review of the optimization capabilities of PSO algorithms.

Ant colony optimization (ACO) is an algorithm that mimics ants' ability to transfer pheromones to food sources outside the anthill. Artificial intelligence mimics this behaviour through the interaction of agents. Since the late 1980s, when the ACO algorithm was first developed, ACO has been used to graphically find the best path to a goal. In recent years, however, researchers have been tackling other challenges. Liu et al. [55] use ACO to optimize two goals: distance and time. For this purpose, agents are assigned different properties of the ant elite and the ant lion using the mutation algorithm. Li [56] improved the selection of path nodes and the regulation and control of pheromone concentration in ACO, which improves the quality of path optimization. Kounte et al. [57] examined the evolution of evolutionary algorithms with a stronger focus on solving the vehicle routing problem and traveling salesman problem with ACO. Qi et al. [58] combined ACO with genetic algorithms and obtained an improved algorithm that can develop each individual's collective intelligence and force them to evolve along with swarms.

The artificial bee colony (ABC) algorithm was first proposed by Karaboga [59] for single- and multi-target optimization. Artificial bees, which have different targets, fly in a multidimensional space and search for food sources with the greatest amount of food. If one bee finds an amount of food greater than that found by the other bees, it captures that location and passes the information on to other bees. In recent years, researchers have been refining the ABC algorithm. Zhang et al. [60] applied the improved fitness function and improved update strategy of food source and improved the quality of the solution. Aslan [29] supplemented the ABC algorithm with two improvements; the first used only the most abundant food sources found by artificial bees, and the second used all food sources, even those left by bees. The combination of these improved solutions enables the use of an algorithm not only for optimization but also for detection. Over time, scientists have developed many different

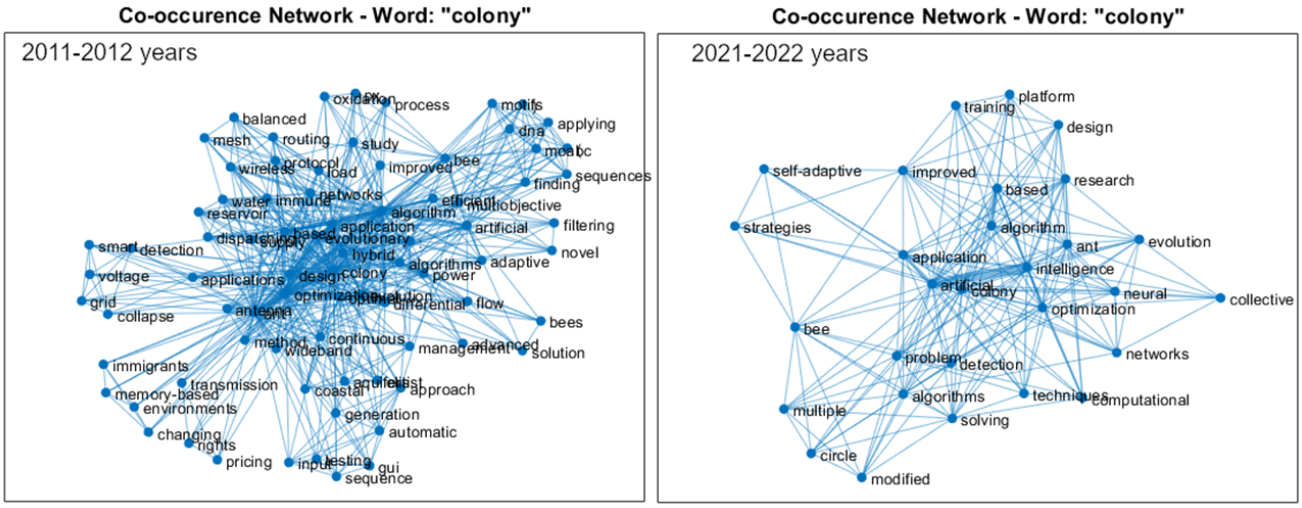



Fig. 6. Co-occurrence network for 'colony' (left: 2011–2012, right: 2021–2022)

algorithms. Solgi and Loáiciga [61] compared the bee system (BS), mating bee optimization (MBO), bee colony optimization (BCO), bee evolution for genetic algorithms (BEGA), bee algorithm (BA), artificial bee colony (ABC), and bee swarm optimization (BSO) global optimization and perceived that ABC is best suited for all seven purposes, whereas others specialize in optimizing certain features.

The firefly algorithm (FA) was proposed by Yang [62]. Kaur et al [63] investigated FA to efficiently apply this algorithm to the relay coordination problem. The authors have studied and applied to the new field three main behavioural features of fireflies.

1. One firefly attracts the other irrespective of their sex because they are unisex in nature.
2. Fireflies' attractiveness is inversely proportional to distance and directly proportional to brightness.
3. The brightness obtained depends upon the fitness function of the firefly.

Yang and Deb [64] developed a cuckoo search algorithm based on information collected by biologists about cuckoos' aggressive breeding strategy. This optimization algorithm can be multi-objective because the nest can contain several eggs that represent different solutions. Rajita et al. [28] applied the cuckoo search (CS) algorithm to social network research, which first detects a certain community, then identifies selected features, and then counts community group similarities and events. The authors compared CS with the already-known particle swarm optimization (PSO) and ant colony optimization (ASO) algorithms for solving a selected social network problem.

Yang [65] developed a bat algorithm (BA) that mimics microbats' use of echolocation, flight, and the ability to emit and recognize sounds of different frequencies. Huang et al. [66] applied the bat algorithm to improve the study process by using multiple intelligence tasks and assessment methods. The monarch butterfly optimization (MBO) algorithm was descripted by Wang et al. [67] and Ghetas et al. [68]; its novelty is that it relied on the migration behaviour of individuals. Self-adaptive crossover (SAC) operator creates new, more perfect individuals in the population, as selecting the best can significantly speed up processes. Feng et al. [69] summarized the contributions of many authors in improving MBO, creating new modifications and hybrids, and evaluated the field of application of this algorithm. Pierezan and Coelho [70] were the first to develop and study an optimization algorithm based on coyote behaviour. The proposed coyote optimization algorithm (COA) is an extension of grey wolf optimizer (GWO), from which it differs in that it does not use hierarchy and related rules in the population, but rather divides it into small groups that exchange experience rather than simply hunting for prey. Li et al. [71] improved COA and applied it to image segmentation based on fuzzy multilayer thresholding. The results obtained show a very good potential for their use in medicine, where accuracy is crucial. Sulaiman et al. [72] proposed a new barnacle mating optimizer, which achieved efficient optimization results by testing 23 mathematical equations. Barnacle is a marine crustacean with an external shell, which feed by filtering particles from the water using their modified feathery legs. Shankar et al. [21] applied BMO in an artificial intelligence-based diagnosis model for Covid-19.

Microbiology and medical knowledge have also inspired scientists to create algorithms that simulate the behaviour of viruses [73] and the processes of changes in immunity [74].

The use of colony logic in algorithms has been criticized by scientists for using easy-to-remember, imaginative names borrowed from nature that can obscure scientific novelty. However, the marketing appeal of such algorithms makes it easier to recognize the differences and purposefully choose them for application.

## 4. Discussion and limitations

Several choices were made that determined the content of this article. The first choice is that of selecting the Scopus database of scientific articles, which stands out for its high-quality requirements and wide coverage of topics. Because there are many other databases available, the choice to use a single database is a major limitation of this article. In addition, because it takes longer to publish articles than to give a seminar, the latest scientific achievements are perhaps first presented at conferences. The price of publications also plays a significant role, which is especially relevant for scientists with fewer economic resources.

A second choice was the use of a collaborative network for article classification and selection. The main dilemma here is whether to leave the review writing to a machine. Before writing the review, the author had a different idea of the distribution of articles and a different understanding of the main subtopics. The network created other priorities, and comparing articles from a decade ago and now, aspects of EAI from network are on the newer side.

The third choice is the use of the keyword 'evolution'; this choice is the most subjective. The keywords by which the articles were selected are not suitable here, so a word was searched to combine the topics. As a result, five sub-themes were obtained. Methods, colony, and detection are the topics most often associated with EAI. Data is also a very important aspect of EAI, but articles in this area have given rise to nuances such as data security and privacy. Approach is the most unexpected – but no less interesting – subsection.

A graphical comparison of co-occurrence networks a decade ago and now reveals a growing diversity of research directions. The single topic of swarm intelligence has narrowed.

Our research can be extended in several directions. This could be done first by looking at more databases of scientific articles, or by comparing this database articles to the research results of another very well-known database, Clarivate Analytics Web of Science. Another approach would be to look through the mapping of EAI application sectors.

## Conclusions

In this study, we used information from the Scopus database on 206 articles from 2021–2022 and compared them with scientific articles from a decade ago. For research use, co-occurrence networks have identified five important research subtopics in recent years: methods, detection, data, colonies, and approach. During the decade, research directions increased in all subtopics, except for the field of colonial intelligence.

The differential evolution method remains the most popular; however, there is a need to use not only single-objective optimization but also multi-objective optimization. Furthermore, researchers use hybrid methods, combining EAI with other methods.

EAI has optimization as well as detection functions. It has been successfully applied in medicine, in recognizing anomalies and fraudulent behaviour, and in identifying and solving business and social problems.

In recent years, scientists have helped businesses confront the challenge of balancing data privacy, security, and reliability with the opportunities that big data provides for businesses and customers.

In scientific articles, EAI is found in different approaches: optimization, machine learning, artificial intelligence, classification, and protein. All of them reflect the heuristic or meta-heuristic nature of EAI.

Algorithms that simulate the behaviour of swarms or colonies are assigned the name of a particular animal species by scientists. The best known are the ant colony optimization and the artificial bee colony algorithms. Algorithms that are less well known – but which nevertheless offer specific advantages – are the firefly, cuckoo search, and bat algorithms. Recent colony optimization algorithms include monarch butterfly optimization (MBO), which simulates the migration phenomenon, and the coyote optimization algorithm, which abandons the colony hierarchy.

Co-occurrence networks revealed that the scope of topics, methods, and application areas of evolutionary artificial intelligence is growing rapidly.

## Abbreviations

| | | |
|---|---|---|
| ABC | - | Artificial Bee Colony |
| AI | - | Artificial Intelligence |
| BA | - | Bee Algorithm |
| BCO | - | Bee Colony Optimization |
| BEGA | - | Bee Evolution for Genetic Algorithms |
| BS | - | Bee System |
| BSO | - | Bee Swarm Optimization |
| EAI | - | Evolutionary Artificial Intelligence |
| IoT | - | Internet of Things |
| MBO | - | Mating Bee Optimization |

## References


1. SCOPUS database, <https://www.elsevier.com/solutions/scopus>, retrieved 2022.06.01.
2. Matlab (n.d.) https://ch.mathworks.com/help/textanalytics/ug/create-co-occurrence-network.html
3. Rajita B. S. A. S., Narwa B. S., Panda S. (2021) An efficient approach for event prediction using collaborative distance score of communities. – – In: International Conference on Distributed Computing and Internet Technology (pp. 271–279). Springer, Cham.
4. Newman M. E. (2004). Coauthorship networks and patterns of scientific collaboration. – *Proceedings of the National Academy of Sciences* 101(suppl_-1), 5200–5205.
5. Zhang Y., Wu M., Tian G. Y., Zhang G., Lu J. (2021) Ethics and privacy of artificial intelligence: Understandings from bibliometrics. – *Knowledge-Based Systems* 222, 106994.
6. Elmsili B., Outtaj B. (2018) Artificial neural networks applications in economics and management research: An exploratory literature review. – In: 2018 4th International Conference on Optimization and Applications (ICOA) (pp. 1–6). IEEE.
7. Ahli, H., Merabtene, T., Seddique, M. (2021, December). Optimization of a Conceptual Rainfall-Runoff Model using Evolutionary Computing methods. – In: 2021 14th International Conference on Developments in eSystems Engineering (DeSE) (pp. 424–431). IEEE.
8. Storn R., Price K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. – *Journal of Global Optimization* 11(4), 341–359.
9. Jebaraj L. (2022) Applications of Differential Evolution in Electric Power Systems. – In: Differential Evolution: From Theory to Practice (pp. 265–296). Springer, Singapore.
10. Yan W., Bai Y., Xu R., Neculaes V. B. (2022) X-ray source design optimization using differential evolution algorithms – A case study. – *Review of Scientific Instruments* 93(5), 053101.
11. Long W., Gao Y. (2022) Artificial Intelligence Education System Based on Differential Evolution Algorithm to Optimize SVM. – *Scientific Programming*.
12. Liu J., Liang B., Ji W. (2021) An anomaly detection approach based on hybrid differential evolution and K-means clustering in crowd intelligence. – *International Journal of Crowd Science*.
13. Yu Y., Wang H., Liu S., Guo L., Yeoh P. L., Vucetic B., Li, Y. (2021). Distributed multi-agent target tracking: A Nash-combined adaptive differential evolution method for UAV systems. – *IEEE Transactions on Vehicular Technology* 70(8), 8122–8133.
14. Choi T. J., Togelius J. (2021) Self-referential quality diversity through differential MAP-Elites. – In: Proceedings of the Genetic and Evolutionary Computation Conference (pp. 502–509).

15. Lambrinidis G., Tsantili-Kakoulidou A. (2021) Multi-objective optimization methods in novel drug design. – *Expert Opinion on Drug Discovery* 16(6), 647–658.
16. Boukhari N., Debbat F., Monmarché N., Slimane M. (2021) Solving Mono- and Multi-Objective Problems Using Hybrid Evolutionary Algorithms and Nelder-Mead Method. – *International Journal of Applied Metaheuristic Computing (IJAMC)* 12(4), 98–116.
17. Nie X., Luo J. (2021) The hybrid intelligent optimization algorithm and multi-objective optimization based on big data. – In: Journal of Physics: Conference Series (Vol. 1757, No. 1, p. 012132). IOP Publishing.
18. Wang T., Yang X., Mi C. (2021) An efficient hybrid reliability analysis method based on active learning Kriging model and multimodal optimization-based importance sampling. – *International Journal for Numerical Methods in Engineering* 122(24), 7664–7682.
19. De Melo M. C., Santos P. B., Faustino E., Bastos-Filho C. J., Sodré A. C. (2021) Computational Intelligence-Based Methodology for Antenna Development. – *IEEE Access* 10, 1860–1870.
20. Kumar R., Wang W., Kumar J., Yang T., Khan A., Ali W., Ali I. (2021) An integration of blockchain and AI for secure data sharing and detection of CT images for the hospitals. – *Computerized Medical Imaging and Graphics* 87, 101812.
21. Shankar K., Perumal E., Díaz V. G., Tiwari P., Gupta D., Saudagar A. K. J., Muhammad K. (2021) An optimal cascaded recurrent neural network for intelligent COVID-19 detection using Chest X-ray images. – *Applied Soft Computing* 113, 107878.
22. Afza F., Sharif M., Khan M. A., Tariq U., Yong H. S., Cha J. (2022) Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine. – *Sensors* 22(3), 799.
23. Fatani A., Abd Elaziz M., Dahou A., Al-Qaness M. A., Lu S. (2021) IoT intrusion detection system using deep learning and enhanced transient search optimization. – *IEEE Access* 9, 123448–123464.
24. Mishra P., Gupta A., Aggarwal P., Pilli E. S. (2022) vServiceInspector: Introspection-assisted evolutionary bag-of-ngram approach to detect malware in cloud servers. – *Ad Hoc Networks* 131, 102836.
25. Florkowski M. (2021) Anomaly Detection, Trend Evolution, and Feature Extraction in Partial Discharge Patterns. – *Energies* 14(13), 3886.
26. Ge M. (2022) Recognition and Detection Methods of Artificial Intelligence in Computer Network Faults under the Background of Big Data. – *Wireless Communications and Mobile Computing* 2022.
27. Erfanian P. Y., Cami B. R., Hassanpour H. (2022) An evolutionary event detection model using the Matrix Decomposition Oriented Dirichlet Process. – *Expert Systems with Applications* 189, 116086.
28. Rajita B. S. A. S., Bansal M., Narwa B. S., Panda, S. (2022) Cuckoo search in threshold optimization for better event detection in social networks. – *Social Network Analysis and Mining* 12(1), 1–19.
29. Aslan S. (2021) Modified artificial bee colony algorithms for solving multiple circle detection problem. – *The Visual Computer* 37(4), 843–856.
30. Lai X., Jin C., Yi W., Han X., Feng X., Zheng Y., Ouyang M. (2021) Mechanism, modeling, detection, and prevention of the internal short circuit in lithium-ion batteries: recent advances and perspectives. – *Energy Storage Materials* 35, 470–499.
31. Zahedi L., Ghareh Mohammadi F., Amini M. H. (2022) A2BCF: An Automated ABC-Based Feature Selection Algorithm for Classification Models in an Education Application. – *Applied Sciences* 12(7), 3553.
32. Chen J., Ramanathan, L., Alazab M. (2021). Holistic big data integrated artificial intelligent modeling to improve privacy and security in data management of smart cities. – *Microprocessors and Microsystems* 81, 103722.
33. Chaitra H. V., RaviKumar G. K. (2021) Secure and Energy-Efficient Data Transmission. – In: Advances in Artificial Intelligence and Data Engineering (pp. 1311–1322). Springer, Singapore.
34. Bokhari S. M. A., Theel O. (2020) Designing New Data Replication Strategies Automatically. – In: International Conference on Agents and Artificial Intelligence (pp. 308–331). Springer, Cham.
35. Sekera J., Novák A. (2021) The future of data communication in Aviation 4.0 environment. – *INCAS Bulletin* 13(3), 165–178.
36. Gong J., Sihag V., Kong Q., Zhao L. (2021) Visualizing Knowledge Evolution Trends and Research Hotspots of Personal Health Data Research: Bibliometric Analysis. – *JMIR Medical Informatics* 9(11), e31142.
37. Rahmani A. M., Azhir E., Naserbakht M., Mohammadi M., Aldalwie A. H. M., Majeed M. K., Hosseinzadeh M. (2022) Automatic COVID-19 detection mechanisms and approaches from medical images: a systematic review. – *Multimedia Tools and Applications* 1–20.
38. Pourdarbani R., Sabzi S., Rohban M. H., García-Mateos G., Arribas, J. I. (2021) Nondestructive nitrogen content estimation in tomato plant leaves by Vis-NIR hyperspectral imaging and regression data models. – *Applied Optics* 60(30), 9560–9569.
39. Taghizadeh-Mehrjardi R., Emadi M., Cherati A., Heung B., Mosavi A., Scholten T. (2021) Bio-inspired hybridization of artificial neural networks: An application for mapping the spatial distribution of soil texture fractions. – *Remote Sensing* 13(5), 1025.
40. Uchihira N. (2021) Dialogue Tool for Value Creation in Digital Transformation: Roadmapping for Machine Learning Applications. – In: International Conference on Applied Human Factors and Ergonomics (pp. 518–526). Springer, Cham.
41. Almosnino S., Cappelletto J. (2021) Minimizing low back cumulative loading during design of manual material handling tasks: An optimization approach. – *IISE Transactions on Occupational Ergonomics and Human Factors* 9(3–4), 124–133.
42. De Lima Mendes R., da Silva Alves A. H., de Souza Gomes M., Bertarini P. L. L., do Amaral L. R. (2021) Many Layer Transfer Learning Genetic Algorithm (MLTLGA): a New Evolutionary Transfer Learning Approach Applied To Pneumonia Classification. – In: 2021 IEEE Congress on Evolutionary Computation (CEC) (pp. 2476–2482). IEEE.
43. Polkowski Z., Mishra J. P., Mishra S. K. (2021) Prioritization of complex heterogeneous queries using evolutionary and computational approach. – In: 2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) (pp. 1–5). IEEE.
44. Adeleke O., Akinlabi S., Jen T. C., Dunmade I. (2022) A machine learning approach for investigating the impact of seasonal variation on physical composition of municipal solid waste. – *Journal of Reliable Intelligent Environments* 1–20.
45. Nayeri Z. M., Ghafarian T., Javadi B. (2021) Application placement in Fog computing with AI approach: Taxonomy and a state of the art survey. – *Journal of Network and Computer Applications* 185, 103078.
46. Abdollahizad S., Balafar M. A., Feizizadeh B., Babazadeh Sangar A., Samadzamini K. (2021) Using hybrid artificial intelligence approach based on a neuro-fuzzy system and evolutionary algorithms for modeling landslide susceptibility in East Azerbaijan Province, Iran. – *Earth Science Informatics* 14(4), 1861–1882.
47. Repecka D., Jauniskis V., Karpus L., Rembeza E., Rokaitis I., Zrimec J., Zelezniak A. (2021) Expanding functional protein sequence spaces using generative adversarial networks. – *Nature Machine Intelligence* 3(4), 324–333.
48. Rives A., Meier J., Sercu T., Goyal S., Lin Z., Liu J., Fergus R. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. – *Proceedings of the National Academy of Sciences*, 118(15), e2016239118.
49. Gopinath T., Manu V. S., Weber D. K., Veglia G. (2022) PHRONESIS: a one-shot approach for sequential assignment of protein resonances by ultrafast MAS solid-state NMR spectroscopy. – *ChemPhysChem*
50. Czajkowski M., Jurczuk K., Kretowski M. (2021) Accelerated evolutionary induction of heterogeneous decision trees for gene expression-based classification. – In: Proceedings of the Genetic and Evolutionary Computation Conference (pp. 946–954).

51. Beni G., Wang, J. (1993). Swarm intelligence in cellular robotic systems. – In: Robots and Biological Systems: Towards a New Bionics? (pp. 703–712). Springer, Berlin, Heidelberg.
52. Kadkol, A. A. (2021) Mathematical model of particle swarm optimization: numerical optimization problems. – In: Applying Particle Swarm Optimization (pp. 73–95). Springer, Cham.
53. Mittal A., Pattnaik A., Tomar A. (2021) Different variants of particle swarm optimization algorithms and its application: A review. – *Metaheuristic and Evolutionary Computation: Algorithms and Applications* 131–163.
54. Chen P. Y., Chen R. B., Wong, W. K. (2022) Particle swarm optimization for searching efficient experimental designs: A review. – *Wiley Interdisciplinary Reviews: Computational Statistics* e1578.
55. Liu, Y., Qin, W., Zhang, J., Li M., Zheng, Q., Wang J. (2021) Multi-Objective Ant Lion Optimizer Based on Time Weight. – *IEICE Transactions on Information and Systems* 104(6), 901–904.
56. Li L., Sun L., Xue Y., Li S., Huang X., Mansour R. F. (2021) Fuzzy multilevel image thresholding based on improved coyote optimization algorithm. – *IEEE Access* 9, 33595–33607.
57. Kounte M.R., Niveditha E., Afrose K., Sai Sudeshna A. (2020) Problem Solving Techniques Using Ant Colony Optimization in Computational Intelligence. – 2nd International Conference on Data Science, Machine Learning and Applications, ICDSMLA 2020. Volume 783, pp. 739–747.
58. Qi X., Gan Z., Liu C., Xu Z., Zhang X., Li W., Ouyang C. (2021) Collective intelligence evolution using ant colony optimization and neural networks. – *Neural Computing and Applications* 33(19), 12721–12735.
59. Karaboga D. (2005) An idea based on honey bee swarm for numerical optimization (Vol. 200, pp. 1–10). Technical Report - tr06, Erciyes University, Engineering Faculty, Computer Engineering Department.
60. Zhang J., Zhang Z., Lin X. (2021) An Improved Artificial Bee Colony with Self-Adaptive Strategies and Application. – In:2021 International Conference on Computer Network, Electronic and Automation (ICCNEA) (pp. 101–104). IEEE.
61. Solgi R., Loáiciga H. A. (2021) Bee-inspired metaheuristics for global optimization: a performance comparison. – *Artificial Intelligence Review* 54(7), 4967–4996.
62. Yang X. S. (2008) Nature-Inspired Metaheuristic Algorithms. Luniver Press.
63. Kaur G., Moulik B., Rao K. U. (2021) Determining the optimum TMS and PS of overcurrent relays using the Firefly Algorithm for solving the relay coordination problem. – In: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1011–1015). IEEE.
64. Yang X. S., Deb S. (2009) Cuckoo search via Lévy flights. – In 2009 World Congress on Nature Biologically Inspired Computing (NaBIC) (pp. 210–214). IEEE.
65. Yang X. S. (2010) A new metaheuristic bat-inspired algorithm. – In Nature-Inspired Cooperative Strategies for Optimization (NICSO 2010) (pp. 65–74). Springer, Berlin, Heidelberg.
66. Huang M., Liu S., Zhang Y., Cui K., Wen Y. (2022) Basic Theory and Practice Teaching Method Based on the Cerebellar Model Articulation Controller Learning Algorithm. – *Wireless Communications and Mobile Computing* 2022.
67. Wang G. G., Zhao X., Deb S. (2015) A novel monarch butterfly optimization with greedy strategy and self-adaptive. – In: 2015 Second International Conference on Soft Computing and Machine Intelligence (ISCMI) (pp. 45–50). IEEE.
68. Ghetas M., Yong C. H., Sumari P. (2015) Harmony-based monarch butterfly optimization algorithm. – In: 2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE) (pp. 156–161). IEEE.
69. Feng Y., Deb S., Wang G. G., Alavi A. H. (2021) Monarch butterfly optimization: a comprehensive review. Expert Systems with Applications, 168, 114418.
70. Pierezan, J., Coelho, L. D. S. (2018, July). Coyote optimization algorithm: a new metaheuristic for global optimization problems. – In: 2018 IEEE Congress on Evolutionary Computation (CEC) (pp. 1–8). IEEE.
71. Li F. (2021) Research and Design of Artificial Intelligence Training Platform Based on Improved ant Colony Algorithm. – In: 2021 International Conference on Aviation Safety and Information Technology (pp. 860–863).
72. Sulaiman M. H., Mustaffa Z., Saari M. M., Daniyal H. (2020) Barnacles mating optimizer: a new bio-inspired algorithm for solving engineering optimization problems. – *Engineering Applications of Artificial Intelligence* 87, 103330.
73. Fountas N. A., Vaxevanidis N. M. (2021) Optimization of abrasive flow nano-finishing processes by adopting artificial viral intelligence. – *Journal of Manufacturing and Materials Processing* 5(1), 22.
74. Wu S. (2022) Application of Artificial Immune Algorithm in Evolutionary Creation. – In: International Conference on Cognitive-based Information Processing and Applications (CIPA 2021) (pp. 766–771). Springer, Singapore.

*Article*

# Forecasting of Successful Completion of University Study Programs: Data Pre-processing and Optimization of LAMA BPO Algorithm

Aleksei Iurasov[1,2*], Artur Iurasov[3]
[1] Department of Business, Vilnius Business College, Kalvarijų g. 129-401, LT-08221 Vilnius, Lithuania
[2] Department of Business Technologies and Entrepreneurship, Faculty of Business Management, Vilnius Gediminas Technical University, Saulėtekio 11, Vilnius, Lithuania,
[3] Ltd "Nacionalinis švietimo centras", Žalgirio g. 92, LT-09303 Vilnius, Lithuania.
[*] Corresponding author, e-mail: *aleksei.iurasov@vilniustech.lt*

Received: 10 August 2022
Accepted: 30 August 2022
Online: 31 August 2022

**Abstract.** Lithuanian school graduates wishing to be admitted to state-funded places at universities undergo a competitive selection based on their final school and state exam grades. The problem of organizing competitive selection is that in Lithuania there are different types and scales of school knowledge assessments. Algorithm developed by LAMA BPO address this problem by adjusting grades into a single scale. But choice of final arithmetic values into which pupil's grades are converted is not justified theoretically. Proposed by the authors algorithm is a development of the LAMA BPO algorithm and allows to achieve a consistently higher accuracy of predicting learning results at the university. The higher accuracy of the models indicates a better capture of the central trend: a positive correlation between the level of performance in individual school disciplines and the results of university education in certain study programs.

**JEL:** C11, C45, C53, C61, I29.

**Keywords:** educational data analytics; educational data mining; learning analytics; post-secondary education.

**Citation:** Aleksei Iurasov, Artur Iurasov (2022) Forecasting of Successful Completion of University Study Programs: Data Pre-processing and Optimization of LAMA BPO Algorithm. – *Applied Business: Issues & Solutions* 1(2022)32–41 – ISSN 2783-6967.
https://doi.org/10.57005/ab.2022.1.5

## Introduction

Lithuanian school graduates have a large variety of study programs and universities to choose from. The problem of choosing a university and study program is complicated by the phenomenon of unique combination of each university and study program. Approximate number of universities could be obtained from Ref. [1]: in EU - 3400; in Europe - 5700; in the World – 31100.

The school graduates do not have information for assessment of their individual opportunities for successful completion of various study programs at various universities. It is leading to the problems of high dropout rates [2-3], students switching from one study program to another [4], increasing the period of obtaining university education [5] and low efficiency of university education [6-8]. At the same time, universities already possess such information, but do not share it with other stakeholders of educational market. Elimination of data silo may create new services for data-driven educational decisions and rise the efficiency of university education [9-10].

The R&D Project: "Advanced data analysis and forecasting in education" [11] is dedicated to solving this problem. The project aims to create pan-European undergraduate study program search system (USPSS) for school graduates. Analysing information from Lithuanian universities and electronic school diaries on individual performance of pupils and students, the system advises the most appropriate study programs to school graduates. Analysis of data on VGTU students confirmed the existence of central tendency, a positive correlation between the level of performance in certain school disciplines and the results of further education in certain university study programs [12]. Accordingly, it is possible to create predictive models of university learning results based on demographic data and school learning results. The USPSS suggestions are based on criteria of successful completion: a) low probability of dropout, failing exams or passing after many retakes, switching study programme; b) high probability of admission, successful graduation, and career etc.

This work is aimed to solve the following tasks.

1. Review the existing literature on the application of Data Science in the education sector to predict students' achievements and learning results.
2. Analyse the student data of Lithuanian universities, including students' school marks and their adjustment into a single scale in accordance with the LAMA BPO algorithm, to identify the correlation between students' characteristics at the time of their admission and their future performance at university.
3. Examine LAMA BPO algorithm, based on students' data from two Lithuanian universities and confirm or refute the possibility of creating an optimal algorithm for adjusting school grades into a single scale based on a positive relationship between the level of performance in certain school disciplines and the results of further education in certain study programs at the university.
4. Suggest further research directions.

## 1. Literature review

Many authors tested different Data Science methods to predict students' achievements and learning results. Existing research can be divided based on data sources used: 1) behavioural data (clickstream analytics, pages visited, time spent, etc.) from online learning systems [13-16]. 2) student's previous grades - articles based on student performance should be considered in more detail, since they have a similar data structure and algorithms.

Alsuwaiket et al. [17] applied *Naive Bayes* and *Random Forest* algorithms to forecast students' second year averages based on their first-year averages. The results obtained by researchers especially interesting as data gathered for research was from similar study programs and departments with similar names as we are going to analyse in case of VGTU (Vilnius Gediminas Technical University) (study programmes *Civil Engineering*, *Computer Science*, *Electrical and Computer Systems*, *Engineering*, *Mathematics*, *Mechanical Engineering*, *Business*). *Random Forest* algorithm was more accurate. Accuracy measured by using Area Under the Curve (AUC) method, that provides an aggregate measure of forecasting model performance.

Similar to previous research Hasan et al. [18] used classification approach to forecast future university learning outcomes, based on relatively small dataset of 1170 students. They applied *K-Nearest Neighbors* and *Decision Tree* (with *ID3* algorithm to determine the root). *Decision Tree* was more accurate in predicting student performance.

Kostopoulos et al. [19] applied *Bayesian network*, *Naïve Bayes*, *Decision tree* (C4.5), *K-Nearest Neighbors* and *Sequential Minimal Optimization* to forecast future marks based on dataset of 340 students. Marks of two previous academic semesters were used as independent variables. The highest accuracy is 72.94% was achieved by the model developed based on *Naïve Bayes*.

As Panessai et al. [20] aimed their study to predict the student's attrition and failing to complete the course, it unites them with this research work. Other similarity is study programme name they analysed: *Software Engineering* at Universiti Pendidikan Sultan Idris. In present research we are going to analyse 595 student's data from VGTU *Software Engineering* programme (Panessai et al. [20] analyzed data of 123 students) along with student's data from other 42 VGTU and 8 LSU study programs. But Panessai et al. [20] used learning outcomes data of already admitted students, grades for: mid-term exam, group project, quizzes, coursework, and other assessments.

Contrary we use school learning outcomes as independent variables to predict student's failing to complete the course and other student's performance results. Further add to differences is classification approach chosen by Panessai et al. [20]. They applied *Naïve Bayes*, *Generalized Linear classification*, and *Decision Tree* methods to predict the final grade (A, B, C, D, E, F). The highest accuracy (79.18%) was achieved by forecasting model developed based on *Decision Tree* (C4.5 algorithm) method.

Li et al. [21] clustered student grades across three college courses (*Physics*, *Career Planning and Management*, *Chinese Language and Literature*) for teachers to make better decisions on delivering quality education. This is an example of unsupervised learning and therefore has no target variable to predict. By applying *Fuzzy C-means* method Li et al. [21] divided learning outcomes of three university courses in four groups: "Great", "Good", "Average" and "Bad". Therefore, the applicability of research results is limited to visualization and analysis of the status quo. The analysis is based on a small dataset of 246 students from Huaqiao University. In conclusion authors propose possibility of application of *Support Vector Machines* and *Artificial Neural Network* to predict students' grades.

Rajab and Ramadan [22] aimed their research at predicting *Grade Point Average* (GPA) based on grades in the first and second semester, non-public and social factors such as living location (dormitory or apartment), and attendance. But instead of employing regression-based forecasting methods for predicting numeric values, researchers used classification methods dividing predicted GPA into four classes: First Class, Upper Class, Lower Class, Pass. As the research is based on a small dataset of 72 student records from the College of Health Science at the State University of Zanzibar, the results obtained may be just random coincidences. For example, only 2 out of 72 rows in the dataset classified as "Pass" (the lowest result), both rows represented female students with very good attendance and good/very good level of secondary school completion, living in apartments, etc. Based on such characteristics, the Data Science model may predict similar female students to get the lowest results. However, for male students, such a forecast is impossible, since there are no corresponding records in the data based on which the forecasting model was trained.

The choice of modeling methods used by researchers looks controversial since all methods are based on a *Decision Tree* algorithm: *C4.5*, *Iterative Dichotomiser 3* (ID3), *Classification and Regression Tree* (CART), and *Chi-square automatic interaction detection* (CHAID). For example, it is not clear why they use ID3 (precursor to C4.5 algorithm) together with its successor C4.5. Predictably, ID3 was least correct, while CART demonstrated an exceptional accuracy of 100%.

Table 1 represents the applicability of university learning results prediction studies to current research.

## 2. Methodology

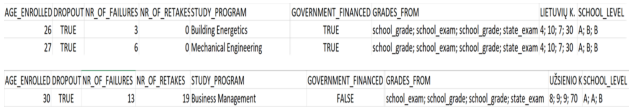
### 2.1. Data preparation: analysis, parsing, transformation, and cleaning

Current research based on the data of 24341 Vilnius Gediminas Technical University (VGTU) and 2137 Lithuanian Sports University (LSU) students. Characteristics and structure of the original dataset contain following statements.

1. Admission year to university Bachelors' study program in (VGTU: from 2009 to 2021, LSU from 2010 to 2019).
2. Country where admitted student studied at school (19 counties).
3. Gender (M/F).
4. Year of school graduation (from 1979 to 2021).
5. Age at the time of the admission to university (from 17 to 58).
6. Drop out of the university ("true" - student dropped-out, "false" - not dropped-out, "changed in the study programme" - student switch to another study program at the same university).
7. Number of courses failed to complete (study courses with the mark lower than 5 in a 10-point system was identified as failed). In the dataset minimum number of failed courses is 0 maximum is 49 of failed courses per student.
8. Number of retakes. In the dataset minimum number of retakes is 0 maximum is 94 per student.
9. Average university grade.
10. Name of study program.
11. Year of university graduation.
12. Type of financing (state funded position / paid by student).
13. Information about school subjects (courses) and school marks includes:
    i) course name;
    ii) level of education: "A", "B", "S", "M", "T";
    iii) type of grade: "annual school grade", "school exam" or "state maturity exam";
    iv) course mark.

Table 1. Applicability of university learning results prediction studies to current research.

| Type of ML | Aim | Target variable | More accurate method/ other methods | Accuracy, % | Independent variables | Dataset size - Number of students | Study programme | Refs. | Applicability |
|---|---|---|---|---|---|---|---|---|---|
| SL | Not specified by authors | Second year average, classified by groups (Fail, First, Lower second, Pass, Third, Upper second) | *Random Forest/ Naive Bayes* | 94.2 | Course marks (module, coursework, exam), other course-related, study program-related attributes | 230 823 | Civil Engineering, Computer Science, Electrical and Computer Systems, Engineering, Mathematics, Mechanical Engineering, Business | [17] | Very limited by di erences in: discretization of target variable, and data structure |
| SL | Admission and scholarships decisions | GPA classified by groups (fail, fair, good, very good) | *Decision Tree/Bayesian Network* | 86 | Grades in previous years by semesters, non-public and social factors such as Family Job, Ethnic, Religion | 20492 | Not specified by the authors | [23] | Limited by di erences in: discretization of target variable, and data structure |
| SL | Decision of additional attention from teacher | GPA classified by groups (First Class, Upper Class, Lower Class, Pass) | *CART/ C4.5*, *Decision Tree (C4.5)*, *Decision Tree (ID3)*, *CHAID* | 100 | Grades in first and second semester, non-public and social factors such as living location (dormitory or apartment), and attendance | 72 | Healthcare | [22] | Questionable due to di erences in: research aims, discretization of the target variable, student's field of study, volume, and the structure of the data used for forecasting |
| SL | Decision of additional attention from teacher | Course final examination mark (numbers classified as groups) | *Decision Tree* (ID3)/ *K-Nearest Neighbors* | 94.88 | Course name, test mark, attendance mark, presentation mark, assignment mark, midterm mark, final examination mark | 1170 | Not specified by the authors | [18] | Very limited by di erences in: research aims, discretization of target variables, and the structure of the data used for forecasting |
| SL | Decision of additional attention from teacher | Course final grade classified by groups (Poor, Good, Very good, Excellent) | *Naïve Bayes/Bayesian network*, *Decision tree* (C4.5), *K-Nearest Neighbors*, *Sequential Minimal Optimization* | 72.94 | Two semester course marks (oral, tests, exam, overall) | 340 | Not specified by the authors, since the dataset is taken from open sources | [19] | Very limited by di erences in: research aims, discretization of target variable, and the structure of the data used for forecasting |
| SL | Decision of additional attention from teacher | Final grade classified by groups (A, B, C, D, E, F) | *Decision Tree (C4.5)/ Naïve Bayes*, *Generalized Linear classification* | 79.18 | Grades: mid-term exam, group project, quizzes, coursework, other assessments | 123 | Software Engineering | [20] | Aim of the study partly coincides with the aim of current research, but applicability of study is limited by di erences in discretization of target variable, and the structure of the data used for forecasting |
| USL | Clustering and visualizing student grades across three courses to make better decisions on delivering quality education | Final grades of three college courses clustered by groups (Great, Good, Average, Bad) | *Fuzzy C-means clustering* | Not used in USL | Grades of three college courses (College Physics, Career Planning and Management for College Students, College Chinese Language and Literature) | 246 | Information Technology | [21] | Not possible due to a fundamental di erence in the research aim |



| AGE | ENROLLED | DROPOUT | NR_OF_FAILURES | NR_OF_RETAKES | STUDY_PROGRAM | GOVERNMENT_FINANCED | GRADES_FROM | LIETUVIŲ K. SCHOOL_LEVEL |
|---|---|---|---|---|---|---|---|---|
| 26 | TRUE | | 3 | 0 | Building Energetics | TRUE | school_grade; school_exam; school_grade; state_exam | 4; 10; 7; 30 A; B; B |
| 27 | TRUE | | 6 | 0 | Mechanical Engineering | TRUE | school_grade; school_exam; school_grade; state_exam | 4; 10; 7; 30 A; B; B |

| AGE | ENROLLED | DROPOUT | NR_OF_FAILURES | NR_OF_RETAKES | STUDY_PROGRAM | GOVERNMENT_FINANCED | GRADES_FROM | UŽSIENIO K SCHOOL_LEVEL |
|---|---|---|---|---|---|---|---|---|
| 30 | TRUE | | 13 | 19 | Business Management | FALSE | school_exam; school_grade; school_grade; state_exam | 8; 9; 9; 70 A; A; B |

Fig. 1. Fragment of the dataset with school marks data.

Fig. 1 shows different types, levels, and scales of school knowledge assessments for courses in "Lithuanian" and "Foreign language". Here we see a difference in the grades of the same school subject. The first student's annual school grades in Lithuanian are 4 (A level) and 7 (B level). In addition, the student has 10 (B level) for school exam and 30 for state exam. Also, we can see another student with double annual school marks in foreign language: 9 (A level) and 9 (B level), and 8 for the school exam (A level) with 70 for the state exam (Fig. 1).

To build correctly functioning forecasting models and predict successful completion of study programme, it is required to adjust all school marks into a single scale. The algorithm developed by Lithuanian Association of Higher Education Institutions for General Admission (LAMA BPO) was chosen for that purpose [24] . LAMA BPO is an Association the purpose of which is to organize and coordinate general admission to Lithuanian institutions implementing study and training programs. This algorithm was implemented in the processing of the data of Lithuanian universities and schools. According to the LAMA BPO algorithm, school final exams and annual grades have several types of adjusting coefficients.

The state exam grades of 16-100 points scale are adjusted into the single scale according to the Eq.(1):

$$Y = 4 + (X - 16) * 0.07143 \quad (1)$$

where $Y$ represents the grade adjusted to the single scale; $X$ represents the grade of the state exam. As can be seen from Table 2, the value of annual school grades is significantly inferior to the value of final school exam grades. According to the algorithm, if there are several grades for a school discipline, then after adjusting them into a single scale, the maximum is selected. This grade is considered an assessment of the pupil's knowledge of the school course. Following the LAMA BPO algorithm, let's determine the grade in Lithuanian based on the data from Figure 1: the annual school grade of A-level will become 3.6 (see Table 2), the annual school grade of B-level will become 2.3, the school exam grade of B-level will become 7.2, and the state exam grade will become 5 - see Eq.(1). The maximum value is 7.2, this value is taken as an assessment of Lithuanian.

To process data and convert school grades into university entry points, based on the LAMA BPO algorithm, KNIME Analytics Platform was used (Fig. 2, 3, 5). Since in the original dataset, all grades of the applicant are in one cell through the separating ";" (Fig. 1), then to process each grade, program divided each cell by number of grades. For this, purpose node "cell Splitter" is used (Fig. 2). Because each grade of school discipline has three types of

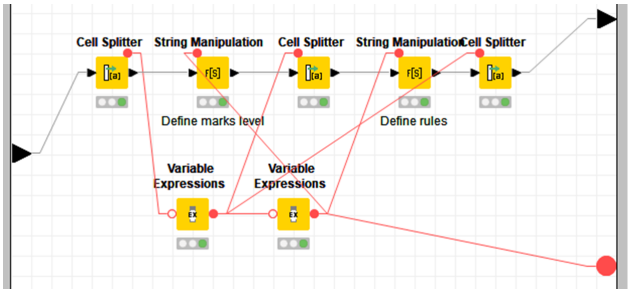



Fig. 2. Converting school marks into a single scale: data preprocessing.

characteristics: grade value, grade type and level, then to process, for example, four different grades in Lithuanian at the end of this step we get 12 columns with grades characteristics. Variable Expressions - assigns the iteration number to the name of each column so that the grades and their characteristics are not confused (Fig. 2).

The program shown in Fig. 3 creates rules to convert school marks by using the LAMA BPO algorithm. Each rule is written by two nodes ("Table Creator" and "String Manipulation") then rules are merged in one table (Fig. 3). Fig. 4 shows conversion table with the rules (conditions) and conversion values.

The last phase of converting school marks into a single scale is implementation of rules from Fig. 4 and Eq.(1) - as presented in Fig. 5. At the final stage of data processing, cleaning was implemented to process missing values, missing students, etc. This form initial dataset for creation of prediction models of successful completion of study programs.

## 2.2. Modelling the academic performance at university

For each of the mentioned criteria of successful completion each study program of each university, it is required to create a predictive model. For that purpose, forecasting factory was developed based on KNIME Analytics Platform. forecasting models were built using data of 24341 Vilnius Gediminas Technical University (VGTU) and 2137 Lithuanian Sports University (LSU) students who studied from 2010 to 2020. At the top level of the loop the forecasting factory iterates over universities, picking students data of new university at each iteration.

The second level of the loop iterates over parameters of successful completion of study programme (GPA, number of fails to complete the courses, etc.). The third level of the loop iterates over study programmes creating predictive models to forecast parameters of successful completion for each study program. For example, the forecasting model to predict the number of fails to complete courses in the study programme "Air Traffic Control" was developed based on *Ensemble of Regression Trees* algorithm (Fig. 6).

Table 2. Adjusted school annual and final exam grades according to the LAMA BPO algorithm.

| Assessment scales | School ten-point scale | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Adjusted A-level, S-level, no-level school exam grade | 4,4 | 5,3 | 6,1 | 7,0 | 7,9 | 8,7 | 9,6 |
| Adjusted B-level school exam grade | 3,3 | 4,0 | 4,6 | 5,3 | 5,9 | 6,5 | 7,2 |
| Adjusted A-level, S-level, no-level annual school grade | 3,6 | 3,9 | 4,2 | 4,5 | 4,8 | 5,1 | 5,4 |
| Adjusted B-level annual school grade | 1,8 | 2,0 | 2,1 | 2,3 | 2,4 | 2,6 | 2,7 |

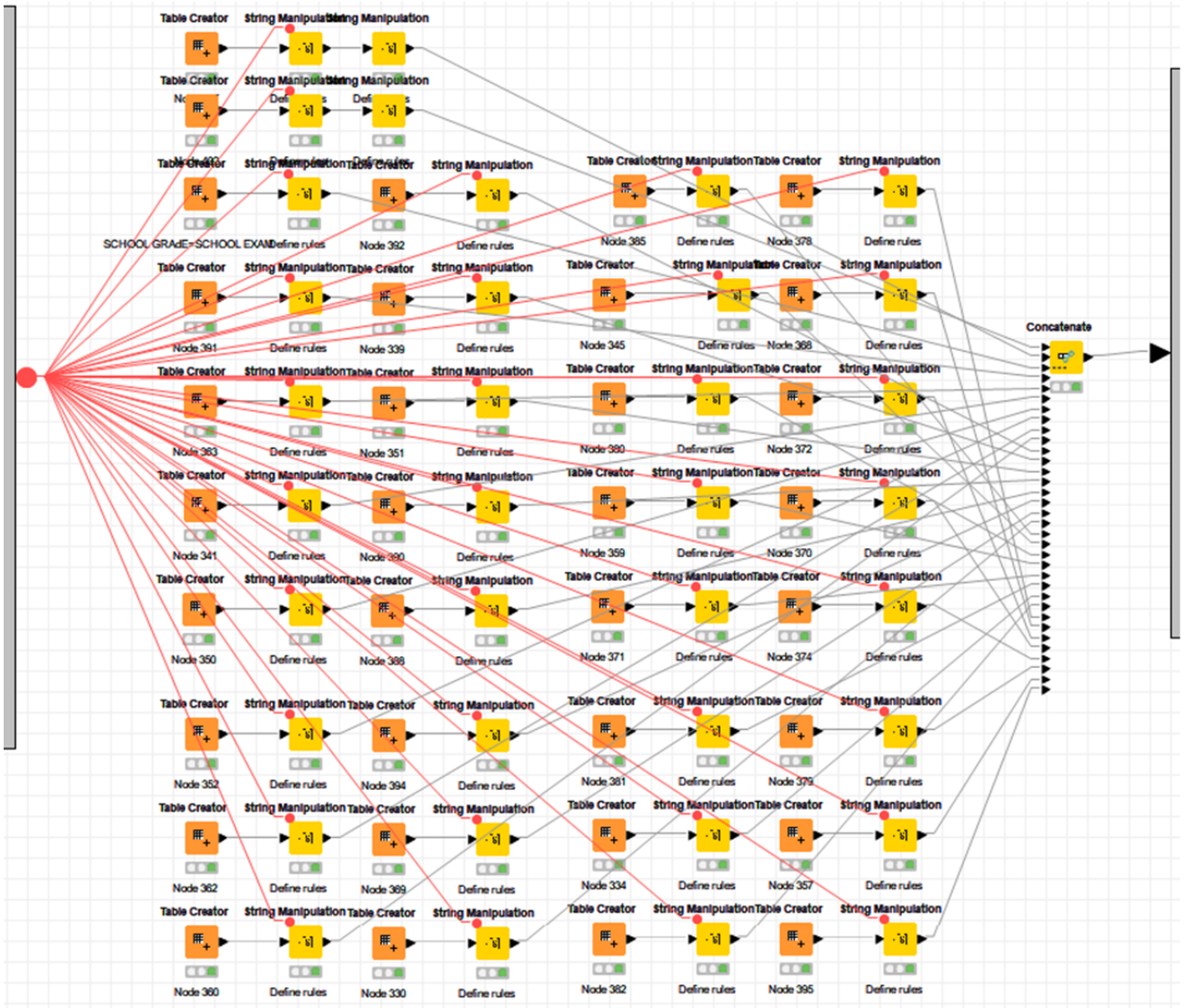



Fig. 3. Converting school marks into a single scale: formation of rules for adjusting school grades into a single scale.

| RULE | Result |
|---|---|
| MISSING $LIETUVIU K._Arr[0]$ | $LIETUVIU K._Arr[... |
| $GRADES_FROM_0_Arr[0]$="state_exam" | $LIETUVIU K._Arr[... |
| $LIETUVIU K._Arr[0]$>=2.49 AND $LIETUVIU K._Arr[0]$<3.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 3.3 |
| $LIETUVIU K._Arr[0]$>=3.49 AND $LIETUVIU K._Arr[0]$<4.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 3.6 |
| $LIETUVIU K._Arr[0]$>=4.49 AND $LIETUVIU K._Arr[0]$<5.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 3.9 |
| $LIETUVIU K._Arr[0]$>=5.49 AND $LIETUVIU K._Arr[0]$<6.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 4.2 |
| $LIETUVIU K._Arr[0]$>=6.49 AND $LIETUVIU K._Arr[0]$<7.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 4.5 |
| $LIETUVIU K._Arr[0]$>=7.49 AND $LIETUVIU K._Arr[0]$<8.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 4.8 |
| $LIETUVIU K._Arr[0]$>=8.49 AND $LIETUVIU K._Arr[0]$<9.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 5.1 |
| $LIETUVIU K._Arr[0]$>=9.49 AND $LIETUVIU K._Arr[0]$="school_grade" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 5.4 |
| $LIETUVIU K._Arr[0]$>=2.49 AND $LIETUVIU K._Arr[0]$<3.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 1.5 |
| $LIETUVIU K._Arr[0]$>=3.49 AND $LIETUVIU K._Arr[0]$<4.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 1.8 |
| $LIETUVIU K._Arr[0]$>=4.49 AND $LIETUVIU K._Arr[0]$<5.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 2.0 |
| $LIETUVIU K._Arr[0]$>=5.49 AND $LIETUVIU K._Arr[0]$<6.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 2.1 |
| $LIETUVIU K._Arr[0]$>=6.49 AND $LIETUVIU K._Arr[0]$<7.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 2.3 |
| $LIETUVIU K._Arr[0]$>=7.49 AND $LIETUVIU K._Arr[0]$<8.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 2.4 |
| $LIETUVIU K._Arr[0]$>=8.49 AND $LIETUVIU K._Arr[0]$<9.49 AND $GRADES_FROM_0_Arr[0]$="school_grade" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 2.6 |
| $LIETUVIU K._Arr[0]$>=9.49 AND $LIETUVIU K._Arr[0]$="school_grade" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 2.7 |
| $LIETUVIU K._Arr[0]$>=2.49 AND $LIETUVIU K._Arr[0]$<3.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 2.5 |
| $LIETUVIU K._Arr[0]$>=3.49 AND $LIETUVIU K._Arr[0]$<4.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 3.3 |
| $LIETUVIU K._Arr[0]$>=4.49 AND $LIETUVIU K._Arr[0]$<5.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 4.0 |
| $LIETUVIU K._Arr[0]$>=5.49 AND $LIETUVIU K._Arr[0]$<6.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 4.6 |
| $LIETUVIU K._Arr[0]$>=6.49 AND $LIETUVIU K._Arr[0]$<7.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 5.3 |
| $LIETUVIU K._Arr[0]$>=7.49 AND $LIETUVIU K._Arr[0]$<8.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 5.9 |
| $LIETUVIU K._Arr[0]$>=8.49 AND $LIETUVIU K._Arr[0]$<9.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 6.5 |
| $LIETUVIU K._Arr[0]$>=9.49 AND $LIETUVIU K._Arr[0]$="school_exam" AND $SCHOOL_LEVEL_0_Arr[0]$="B" | 7.2 |
| $LIETUVIU K._Arr[0]$>=2.49 AND $LIETUVIU K._Arr[0]$<3.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 3.3 |
| $LIETUVIU K._Arr[0]$>=3.49 AND $LIETUVIU K._Arr[0]$<4.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 4.4 |
| $LIETUVIU K._Arr[0]$>=4.49 AND $LIETUVIU K._Arr[0]$<5.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 5.3 |
| $LIETUVIU K._Arr[0]$>=5.49 AND $LIETUVIU K._Arr[0]$<6.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 6.1 |
| $LIETUVIU K._Arr[0]$>=6.49 AND $LIETUVIU K._Arr[0]$<7.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 7.0 |
| $LIETUVIU K._Arr[0]$>=7.49 AND $LIETUVIU K._Arr[0]$<8.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 7.9 |
| $LIETUVIU K._Arr[0]$>=8.49 AND $LIETUVIU K._Arr[0]$<9.49 AND $GRADES_FROM_0_Arr[0]$="school_exam" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 8.7 |
| $LIETUVIU K._Arr[0]$>=9.49 AND $LIETUVIU K._Arr[0]$="school_exam" AND ($SCHOOL_LEVEL_0_Arr[0]$="A" XOR $SCHOOL_LEVEL_0_Arr[0]$="S" XOR $SCHOOL_LEVEL_0_Arr[0]$="M" XOR MISSING $SCHOOL_LEVEL_0_Arr[0]$) | 9.6 |

Fig. 4. Conversion rules (based on LAMA BPO algorithm).

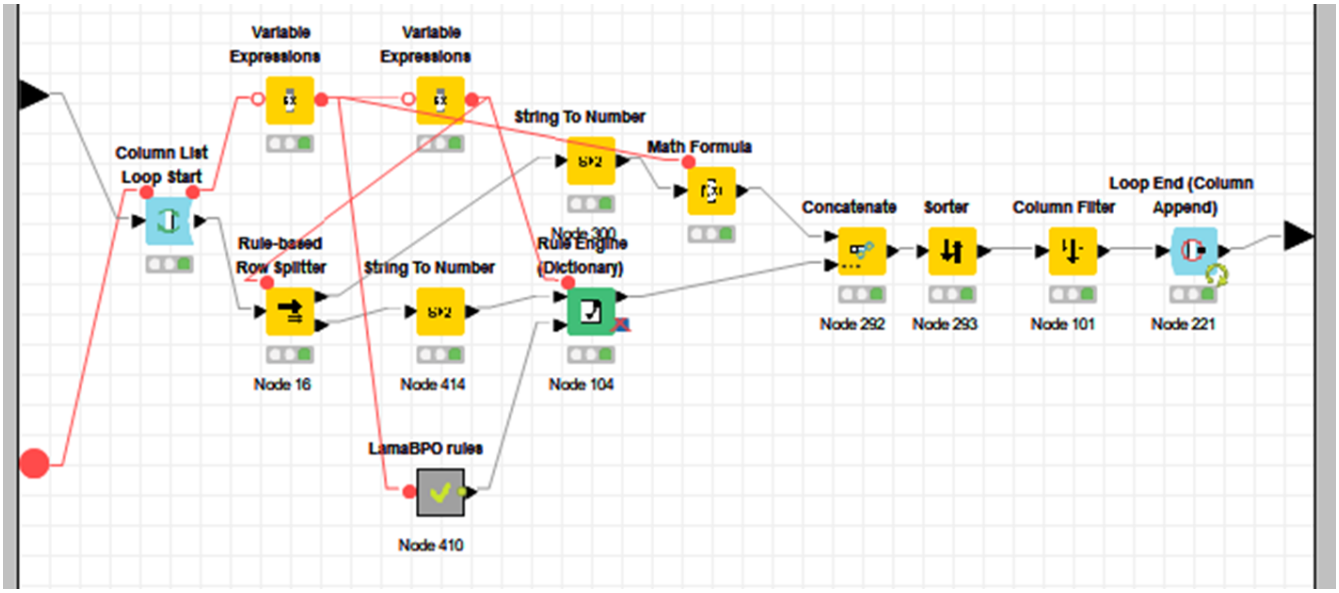



Fig. 5. Converting school marks into a single scale: adjusting school grades into a single scale.

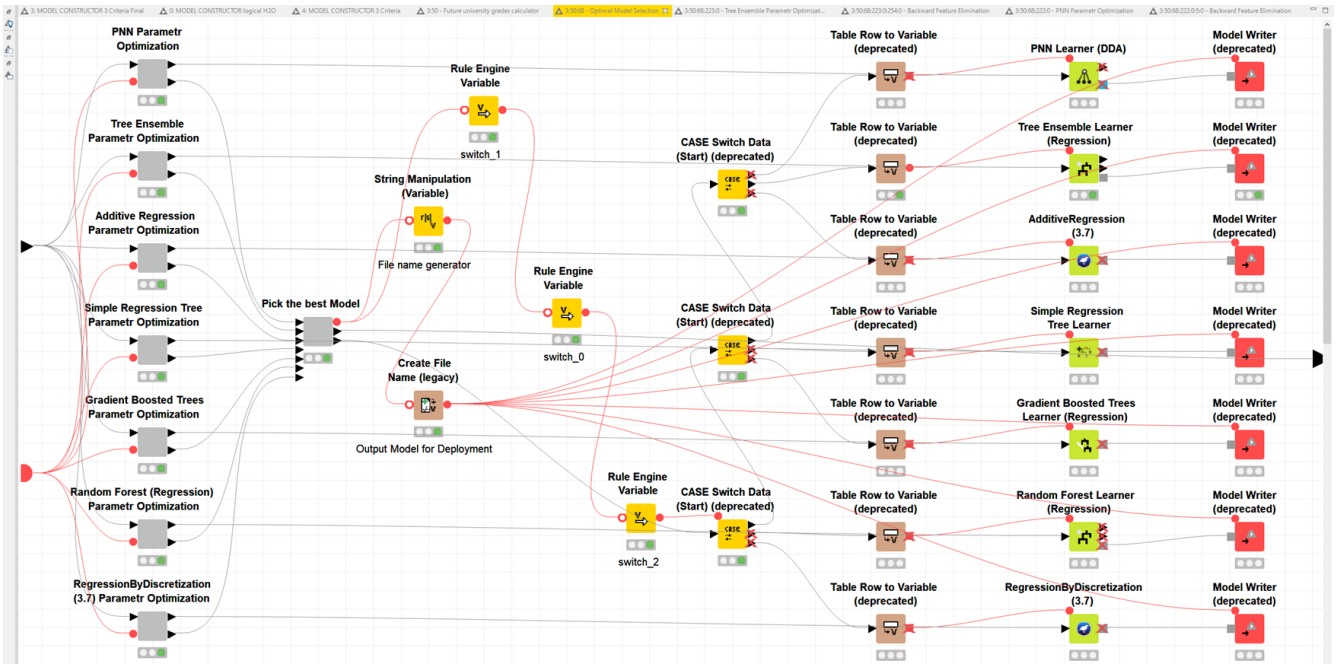



Fig. 6. Predictive model creation to forecast the number of fails to complete courses in the study programme "Air Traffic Control" based on *KNIME Analytics* platform.

For each parameter of successful completion of study programme, models are created and tested, based on seven basic forecasting algorithms (those algorithms differ depending on predicted parameter). The models go through stages of parameter optimization and backward feature elimination to improve their accuracy. As a result, the model with the smallest prediction error is selected to be stored in USPSS.

Iterative work to improve the accuracy of the models created by forecasting factory led to the result after which any of forecasting factory setting updates caused a decrease in the accuracy of the models created. Since the potential to increase the accuracy of forecasting models by further program changes has been depleted, the decision was made to try changes in pre-processing logic of incoming data.

This led the authors of the article to formulate a hypothesis: the LAMA BPO algorithm is non-optimal in adjusting final school grades of Lithuanian school graduates into a single scale. Probable reason is that LAMA BPO algorithm overestimates the level of school knowledge of some pupils and underestimates it for others.

This can reduce university performance, because in some cases, "weaker" pupils are admitted to university instead of more "powerful" ones.

As an optimality criterion for the new algorithm, we choose the minimization of the error of the predictive models of the results of subsequent education at the university.

This criterion assumes that there is a positive relationship between the level of performance in certain school disciplines and the results of further education in certain university study programmes (simplified - the higher the performance at school, the more successful the university education will be). This relation was confirmed based on the analysis of data of VGTU students [12].

It follows that the more accurately we determine school performance, the more accurately we can predict university performance (the error of the predictive model will be less).

To prove the hypothesis, it is enough to find at least one option of adjusting school graduate's grades into a single scale, different from the LAMA BPO algorithm, which will provide consistently more accurate predictions of further learning results at the university. To do this, different options were tested to bring the school grades into a single scale. The generation of models for predicting results of education at the university was carried out, the accuracy parameters of the models were calculated, and compared.

The root-mean-square error (RMSE) was used as a criterion for the assessment of accuracy of the models. RMSE is quadratic mean of differences between predicted values and observed values [25]. It shows how concentrated the predicted and actual values of indicators are [26]. The lower the RMSE of the model, the more accurate the forecast it gives [27].

The difference from the LAMA BPO algorithm was that not only the maximum values of the grades adjusted into a single scale were considered. The decision was made to create predictive models based on minimal, mean, median and max marks for each school course:

1) MINIMUM – from all marks for each school discipline, a minimum is selected;
2) MEAN - the average is calculated for each discipline;
3) MEDIAN - the "middle" of a sorted list of marks;
4) MAX - from all marks for each school discipline, a maximum is selected.

All 4 variants alternative to the LAMA BPO algorithm were developed based on the principle of increasing the weight of annual school grades, with the weight of school and state exams unchanged. The variant that provided the highest accuracy of the forecast, i.e., the smallest RMSE is presented in Table 3.

Accuracy parameters of predictive models (RMSE) are saved for subsequent analysis. Those RMSE of models predicting GPA are presented in Table 4. The method of constructing the most accurate forecasting model is indicated by following abbreviations.

1. ERT, *Ensemble of Regression Trees*, described by Loh [29];
2. BART, *Bayesian Additive Regression Trees*, described by Friedman [30];
3. SGBRT, *Stochastic Gradient Boosted Regression Trees*, described by Friedman [31];
4. RF, *Random Forest Regression*, described by Breiman [32];
5. PNN, *Probabilistic Neural Network*, described by Berthold & Diamond [33].

The table cells with the RMSE of most accurate forecasting models are highlighted in colour (Table 4).

## 3. Results and discussion

It was found that the RMSE of models created based on data adjusted by LAMA BPO algorithm is higher than the similar indicator of the models based on data adjusted to single scale based on algorithm prepared by the authors [28], see Table 4.

Based on Table 4, we can conclude the following outcomes.

1. For the deployment of the models, it is better to choose from models developed on different data sources (Min, Median, Mean, Max mark), e.g., for some study programs higher accuracy could be achieved by considering not Median or Mean mark, but min marks as for study programme *Multimedia Design*.
2. The most accurately determining the level of knowledge of school disciplines, and therefore allowing to predict the success of studying more accurately at the university, are the median and mean of marks values of the grades adjusted to a single scale.
3. The average RMSE of models built based on data adjusted by the LAMA BPO algorithm is the highest and is 0.574 (Table 4). The average RMSE of models built based on data adjusted by the algorithm proposed by the authors of the article is significantly lower. It should be emphasized that our algorithm is more flexible and the system each time selects and saves a model built on such type of data (min, median, mean, max), that allows us most accurately capture the dependence of learning results at the university on learning results at school. In most cases, these are models built on median or mean grades, but sometimes the minimum or maximum grades allow to predict future learning outcomes more accurately.
4. The standardization based on the LAMA BPO algorithm provides a higher error, regardless of which type of the school mark values is considered (min, median, mean, max).
5. The lower RMSE (about 0.08) resulted from the *Probabilistic Neural Network* model based on data of minimum marks of school disciplines. Model was built based on data on 574 students and tested based on data on 143 students (80% and 20% of the initial data on 717 students of the study programme *Multimedia Design*). It indicates the successful determination of central tendency in the data and good forecasting results of the model.

Table 3. Adjusted school annual and final exam grades (prepared by the authors).

| Assessment scales | School ten-point scale | | | | | | |
|---|---|---|---|---|---|---|---|
| | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Adjusted A-level, S-level, no-level school exam grade | 4,4 | 5,3 | 6,1 | 7,0 | 7,9 | 8,7 | 9,6 |
| Adjusted B-level school exam grade | 3,3 | 4,0 | 4,6 | 5,3 | 5,9 | 6,5 | 7,2 |
| Adjusted A-level, S-level, no-level annual school grade | 4,0 | 4,8 | 6,0 | 7,0 | 7,7 | 8,5 | 9,4 |
| Adjusted B-level annual school grade | 3,1 | 4,0 | 4,5 | 5,2 | 5,7 | 6,4 | 7,1 |

Table 4. RMSE of models predicting GPA for VGTU study programmes

| NR | Study programmes | $N_{ST}$ | Method of adjusting school grades into a single scale | | | | | | | | DSA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | LAMA BPO algorithm | | | | Novel algorithm prepared by authors [28] | | | | |
| | | | Min mark | Median mark | Mean mark | Max mark | Min mark | Median mark | Mean mark | Max mark | |
| 1. | Air Traffic Control | 99 | 0.429 | 0.425 | 0.424 | 0.444 | 0.409 | 0.399 | 0.399 | 0.428 | RF |
| 2. | Aircraft Piloting | 139 | 0.536 | 0.488 | 0.518 | 0.529 | 0.468 | 0.452 | 0.452 | 0.478 | SGBRT |
| 3. | Architecture | 658 | 0.607 | 0.567 | 0.584 | 0.591 | 0.556 | 0.562 | 0.568 | 0.567 | ERT |
| 4. | Aviation Mechanics Engineering | 326 | 0.659 | 0.565 | 0.570 | 0.580 | 0.562 | 0.552 | 0.553 | 0.569 | SGBRT |
| 5. | Avionics | 344 | 0.798 | 0.728 | 0.722 | 0.792 | 0.706 | 0.692 | 0.692 | 0.697 | ERT |
| 7. | Building Energetics | 315 | 0.673 | 0.641 | 0.641 | 0.663 | 0.618 | 0.608 | 0.612 | 0.616 | BART |
| 8. | Business Logistics | 211 | 0.560 | 0.583 | 0.586 | 0.557 | 0.525 | 0.532 | 0.538 | 0.509 | SGBRT |
| 9. | Business Management | 649 | 0.592 | 0.585 | 0.583 | 0.664 | 0.548 | 0.547 | 0.547 | 0.548 | PNN |
| 10. | Civil Engineering | 912 | 0.682 | 0.589 | 0.595 | 0.594 | 0.579 | 0.577 | 0.577 | 0.580 | RF |
| 11. | Creative Industries | 818 | 0.565 | 0.554 | 0.564 | 0.609 | 0.549 | 0.532 | 0.526 | 0.537 | BART |
| 12. | Multimedia Design | 717 | 0.743 | 0.521 | 0.578 | 0.690 | 0.081 | 0.14 | 0.14 | 0.14 | PNN |
| ... | ... | | | | | | | | | | |
| 43. | Transport Engineering Economics and Logistics | 825 | 0.732 | 0.799 | 0.734 | 0.733 | 0.711 | 0.705 | 0.702 | 0.621 | RF |
| | Average | | 0.538 | 0.523 | 0.531 | 0.574 | 0.449 | 0.431 | 0.428 | 0.436 | |

$N_{ST}$ Number of students (2010-2020)
DSA Data Science Algorithm for predictive model construction (EW, Mean mark)

The experiment of changing the algorithm of adjusting school marks into a single scale allowed to achieve a significant reduction in RMSE for different VGTU study programs, for example:

a) from 0.690 (LAMA BPO algorithm) to 0.081 (author's algorithm [28]) – study programme *Multimedia Design*.
b) from 0.758 (LAMA BPO algorithm) to 0.216 (author's algorithm [28])) - study programme *Security Systems Engineering*.

This proves research hypothesis that LAMA BPO algorithm is non-optimal in assessing the school performance, it shows the lower accuracy of models based on central tendency: positive correlation between the level of performance in certain school disciplines and the results of further education in certain university study programs [12]. Probable reason is that LAMA BPO algorithm overestimates the level of knowledge for some pupils and underestimates it for others. This can reduce university performance, because in some cases, "weaker" pupils are admitted to university instead of more "powerful" ones.

On the other hand, the author's method [28] of adjusting school grades into a single scale is not optimal too. It should be considered as the first step of Bayesian optimization [34] in which from 100 to 1000 different adjusting algorithms (tables with values of grades adjusted to single scale) will be used to generate data and build forecasting models, assess their accuracy, and guess directions to construct new promising adjusting algorithms, based on the *Tree-structured Parzen Estimator* approach. New promising adjusting algorithms are then evaluated (to minimize RMSE of forecasting models) and search for new promising adjusting algorithms continued. The result will be the optimal adjusting algorithm.

Combining the data on school and university performance from the main Lithuanian universities, we will get a representative database, therefore will be able to develop the optimal algorithm for adjusting school marks into a single scale.

## Conclusions

Several important conclusions can be drawn from this study.

1. The research hypothesis was proven. LAMA BPO algorithm is non-optimal in assessing the school performance and could be improved. Several predictive models have been created according to various methods of adjusting school grades into a single scale. As a result, it was proved that the LAMA BPO algorithm is not the most accurate when it comes to determining the level of knowledge of Lithuanian pupils.
2. The research aim was achieved: a system was developed to forecast a school graduate's university results with a minimum forecasting error.
3. A new hypothesis was put forward: by combining data on school and university performance from the main universities of Lithuania, it is possible to construct a system for calculating the parameters of the optimal algorithm by the criteria of minimizing the errors of forecasting models: successful studies at the university (higher GPA); dropout; not passing exams; retakes; change study program, unemployment and employment prospects, etc.
4. Novel algorithm [28] could be recommended for use by LAMA BPO.

## Acknowledgements

Great thanks to Lithuanian Business Support Agency (VšĮ Lietuvos verslo paramos agentūra) for funding R&D project: "Advanced data analysis and forecasting in education" (J05-LVPA-K-04-0132, "Pažangi duomenų analizė ir prognozavimas švietimo sričiai"). The presented results were by-product of aforementioned R&D project.

## Authors' contributions

Aleksei Iurasov conceived and designed the analysis, reviewed the existing literature, wrote the introduction, literature review and conclusions of the article. Artur Iurasov collected the data, performed the analysis, wrote the methodology, results, and discussion of the article. All authors read and approved the final manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Abbreviations

| | | |
|---|---|---|
| AI | - | Artificial Intelligence |
| AUC | - | Area Under the Curve |
| BART | - | Bayesian Additive Regression Trees [30] |
| ERT | - | Ensemble of Regression Trees [29] |
| GPA | - | Grade Point Average |
| KNIME | - | Konstanz Information Miner |
| LAMA BPO | - | Lietuvos Aukštųjų Mokyklų Asociacija Bendrajam Priėmimui Organizuoti (lith), Association of the General Admission to Lithuanian Universities |
| LSU | - | Lithuanian Sport University |
| ML | - | Machine Learning |
| PNN | - | Probabilistic Neural Network [33] |
| R&D | - | Research and Development |
| RF | - | Random Forest Regression [32] |
| RMSE | - | Root Mean Square Error |
| SGBRT | - | Stochastic Gradient Boosted Regression Trees [31] |
| SL | - | Supervised learning |
| USL | - | Supervised learning |
| USPSS | - | Undergraduate Study Program Search System |
| VGTU | - | Vilnius Gediminas Technical University |

## References


1. Webometrics Ranking of World Universities (2022). Countries arranged by Number of Universities in Top Ranks [online]. http://www.webometrics.info/en/node/54 (Accessed 23 July 2022).
2. Heublein U. (2014) Student Drop-out from German Higher Education Institutions. – *European Journal of Education* Vol. 49 No. 4, pp. 497-513. https://doi.org/10.1111/ejed.12097
3. Hess F. (2018) The college dropout problem [online] https://www.forbes.com/sites/frederickhess/2018/06/06/the-college-dropout-problem/ (Accessed 23 July 2022).
4. Willis M. (2005) Why do students switch from one university to another: The view of students studying for a foreign degree in Hong Kong. – *Journal of Marketing for Higher Education* 15(1), 23-49.
5. Salas-Velasco M. (2007) The transition from higher education to employment in Europe: the analysis of the time to obtain the first job. – *Higher Education* 54(3), 333-360.
6. Salas-Velasco M. (2020) The technical efficiency performance of the higher education systems based on data envelopment analysis with an illustration for the Spanish case. – *Educ. Res. Policy Pract.* 19(2), 159–180
7. Abdalmenem S. A., Owda R. O., Al-Hila A. A., Abu Naser S. S., Al Shobaki M. J. (2018) The Performance Efficiency of University Education between Reality and Expectations. – *International Journal of Academic Management Science Research (IJAMSR)* 2(10), 66-76.
8. Johnes J., Portela M., Thanassoulis E. (2017) Efficiency in education. – *Journal of the Operational Research Society* 68(4), 331-338.
9. Lim L.-A., Gasevic D., Matcha W., Ahmad Uzir N., Dawson S. Impact of learning analytics feedback on self-regulated learning: Triangulating behavioural logs with students' recall – In: 11th International Conference on Learning Analytics and Knowledge: The Impact we Make: The Contributions of Learning Analytics to Learning, LAK 2021; Virtual, Online; United States; 12 April 2021 through 16 April 2021; Code 168184. P. 364-374.
10. Rossen A., Boll C., Wolf A. (2019) Patterns of overeducation in Europe: The role of field of study. IZA J. – *Labor Policy* 9(3), 1–48
11. R&D Project: "Pažangi duomenų analizė ir prognozavimas švietimo sričiai" (Advanced data analysis and forecasting in education, Project number: J05-LVPA-K-04-0132).
12. Iurasov A. (2021) New e-business model: undergraduate study program search system. – *International Journal of Learning and Change* Available at: http://dx.doi.org/10.1504/ijlc.2021.10035252
13. Feng M., Beck J., Heffernan N., Koedinger K. (2008) Can an Intelligent Tutoring System Predict Math Proficiency as Well as a Standarized Test?.In Baker and Beck (Eds.) – In: Proceedings of the 1st international conference on educational data mining (pp. 107–116). Montreal.
14. Puška A., Puška E., Dragić L., Maksimović A., Osmanović N. (2021) Students' satisfaction with E-learning platforms in Bosnia and Herzegovina. – *Technology, Knowledge and Learning* 26(1), 173-191. doi: 10.1007/s10758-020-09446-6
15. Nouri J., Saqr M., Fors U. (2019) Predicting performance of students in a flipped classroom using machine learning: towards automated data-driven formative feedback. – In: 10th International conference on education, training and informatics (ICETI 2019).
16. Poloju K. K., Naidu V. R. (2020) Impact of E-tools in Teaching and Learning for Undergraduate Students. – In: Innovations in Electronics and Communication Engineering (pp. 783-790). Springer, Singapore.
17. Alsuwaiket Mohammed, Blasi Anas and Al-Tarawneh Khawla (2020) Refining Student Marks based on Enrolled Modules' Assessment Methods using Data Mining Techniques. – *Engineering, Technology and Applied Science Research* Vol. 10, pp. 5205-5210.
18. Hasan H.M., Rabby Akm Shahariar Azad, Islam Mohammad, Hossain Syed (2019) Machine Learning Algorithm for Student's Performance Prediction. – In: 10th International Conference on Computing, Communication and Networking Technologies. Kanpur, India.
19. Kostopoulos G., Livieris I. E., Kotsiantis S. and Tampakas V. (2017) Enhancing high school students' performance based on semi-supervised methods. – In: 2017 8th International Conference on Information, Intelligence, Systems and Applications (IISA), pp. 1-6.
20. Panessai I.Y., Lakulu M.M., Rahman M.H., Noor N.A., Salleh N.S., Bilong A.A. (2019) PSAP: Improving Accuracy of Students' Final Grade Prediction using ID3 and C4.5. – *International journal of artificial intelligence* 6, pp. 125-133.
21. Li Y., Gou J., and Fan Z. (2019) Educational data mining for students' performance based on fuzzy C-means clustering. – *The Journal of Engineering* 2019(11), pp. 8245-8250.
22. Rajab A.M., Ramadhan R.M. (2019) Application of Data Mining Techniques in Students' Performance Prediction and Analysis. – *International Journal of Academic Information Systems Research* 3, 1–9.
23. Janecek P. (2007) A comparative analysis of techniques for predicting academic performance. – In: 37th Annual Frontiers in Education Conference - Global Engineering: Knowledge without Borders, Opportunities without Passports.
24. Order of the Minister of Education, Science and Sports Nr. V-718, May 5 2022. – https://www.e-tar.lt/portal/lt/legalAct/d7d19ff0cc4211ec8d9390588bf2de65.
25. Wang W., Lu Y. (2018) Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. – In: IOP conference series: materials science and engineering (Vol. 324, No. 1, p. 012049). IOP Publishing.
26. Chai T., Draxler R. R. (2014) Root mean square error (RMSE) or mean absolute error (MAE). – *Geoscientific Model Development Discussions* 7(1), 1525-1534.
27. Hodson T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. – *Geoscientific Model Development* 15(14), 5481-5487.
28. Aleksei Iurasov (2022) HELM-SCM, v.1. – API
29. Loh W.-Y. (2014) Fifty Years of Classification and Regression Trees. – *International Statistical Review* 82(3) – https://doi.org/10.1111/insr.12016

30. Friedman Jerome H. (2002) Stochastic gradient boosting. – *Computational Statistics and Data Analysis* 38(4), 367–378.
31. Friedman Jerome H. (2001) Greedy function approximation: A gradient boosting machine. – *The Annals of Statistics* 29(5) – https://doi.org/10.1214/aos/1013203451
32. Breiman, L. (2001). Random forests. – *Machine Learning* 45(1) – https://doi.org/10.1023/A:1010933404324
33. Berthold M. R., Diamond J. (1998) Constructive training of probabilistic neural networks. – *Neurocomputing* 19(1–3) – https://doi.org/10.1016/S0925-2312(97)00063-5
34. Mockus J., Eddy W., Reklaitis G. (2013) Bayesian Heuristic approach to discrete and global optimization: Algorithms, visualization, software, and applications (Vol. 17). – Springer Science & Business Media.

*This page intentionally left blank...*

*This page intentionally left blank...*