Ministry of Education and Science of Ukraine Sumy State University
Educational and Scientific Institute of Business, Economics and Management
Department of Economic Cybernetics

BACHELOR'S QUALIFICATION WORK

on the topic " Modelling the potential convergence of the cybersecurity system and combating money laundering"

Completed: 4 year student with
specialty 051 "Economics"
(Business Analytics)
Svitlychna Alona

Head: Doctor of Science, Associate Professor of the
Economic Cybernetics Department
Yarovenko Hanna

Sumy 2022

Ministry of Education and Science of Ukraine
Sumy State University
Educational and Scientific Institute of Business, Economics and Management
Department of Economic Cybernetics

APPROVE
Head of the Department
Dr. Econ. Sciences, Professor
_____ Kuzmenko O.V.
“__”_____2022

TASK
FOR THE BACHELOR'S QUALIFICATION WORK
in the direction of training 051 Economic (Business analytics)
student 4th year of the group АБ-81а.en

Svitlychna Alona

1. Topic of the work: modelling the potential convergence of the cybersecurity system and combating money laundering by order of the university 0382-III from 15.03.2022.

2. The deadline for the student to submit the completed work "_ "_____ 2022.

3. The purpose of the work is the analysis of the process of convergence of two systems: cybersecurity systems and financial monitoring systems.

4.The object of the study is the relationship between the cybersecurity system and combating financial fraud.

5. The subject of research is a scientific approach to the selection, evaluation and formation of indicators that characterize the cybersecurity system and combating financial fraud.

6. Article in a professional journal of category B.

7. Indicative plan of qualification work, terms of submission of sections to the head and the maintenance of tasks for performance of the set purpose

8. Consultations on work:

Chapter	Consultant	Signature, data	
		Task issued by	Task accepted by
1		H.M. Yarovenko.	A.O.Svitlychna
2		H.M. Yarovenko.	A.O.Svitlychna.
3		H.M. Yarovenko.	A.O.Svitlychna.

9. Date of issue of the task“___”_____20__ p.

Supervisor _____ _H.M. Yarovenko __

Received the task to perform _____ _A.O.Svitlychna ____

ABSTRACT

of the qualifying work

for obtaining the educational and qualification level “bachelor”

Svitlychna Alona

MODELLING THE POTENTIAL CONVERGENCE OF THE CYBERSECURITY SYSTEM AND COMBATING MONEY LAUNDERING

The urgency of this work is due to the serious threat to humanity from cyberterrorism. The current experience of the world community is insufficient to fully counter this threat and speaks of a guaranteed vulnerability of any state. This is directly related to the fact that cyberterrorism is a transnational phenomenon, the participants of which have the opportunity to threaten information systems from anywhere in the world. After all, using the global Internet, terrorists can gather detailed information about the targets of attacks, their location and the collection of money to support terrorist acts.

Financial crimes and the growing complexity of cyberattacks have become a widespread problem for financial institutions. A leading provider of corporate cybersecurity solutions, whose products and services are recognized worldwide, will help strengthen your bank's cybersecurity. The introduction of information security tools in the banking infrastructure allows to protect data and resources and create a solid foundation for compliance with regulatory requirements. The data security strategy should be comprehensive, covering people, processes and technologies. The process of convergence of the cybersecurity system and financial crimes will ensure reliability and security in the financial sphere for the country.

The object of the study is the relationship between the cybersecurity system and combating financial fraud.

The subject of the research is a scientific approach to the selection, evaluation and formation of indicators that characterize the cybersecurity system and combating financial fraud.

Methods of research – analysis of data related to the cybersecurity and financial monitoring system for 76 countries.

The information and factual base consisted of: a set of empirical economic indicators of the world (76 observations, 12 variables), based on which the analysis and modeling; documentation in the Python programming language used to perform calculations.

The work contains an introduction, three sections, conclusions, a list of references. The first section describes the theoretical basis for the convergence of the cybersecurity system and combating money laundering. The second section provides a statistical analysis of the potential convergence of the cybersecurity system and the fight against financial crime. In the third section, a neural network model of potential convergence of the cybersecurity system and combating money laundering is built.

The total volume of the work is 40 pages, 30 of them are the main text, 3 are the list of links, 7 are the appendices. 28 illustrations were used to illustrate the study. 28 literary sources are processed in the work.

The results of the study were tested by publishing 1 article in a professional journal of category B and implementation in the discipline of the educational process "Introduction to Business Analytics" and "Forecasting of socio-economic processes".

Key words: convergence, cybersecurity, legalization, financial fraud, criminal proceeds, principal components method, neural network.

CONTENT

INTRODUCTION

The urgency of this work is due to the serious threat to humanity from cyberterrorism. The current experience of the world community is insufficient to fully counter this threat and speaks of a guaranteed vulnerability of any state [1]. This is directly related to the fact that cyberterrorism is a transnational phenomenon, the participants of which have the opportunity to threaten information systems from anywhere in the world. After all, using the global Internet, terrorists can gather detailed information about the targets of attacks, their location and the collection of money to support terrorist acts.

It should be noted that one of the main factors in the development of socio-political system is the production and use of information. In modern conditions, it plays a key role in the functioning of not only public and state institutions, but the life of every person [2].

In the conditions of digitalization and development of technologies, the number of cybercrimes is increasing, as well as the number of financial frauds is growing. The level and quality of cybercrime is constantly improving, criminals do not stop there, and therefore financial crime is also a type of cyber fraud. Achieving a certain level of security can be fully achieved through the interaction of such areas as the analysis of the fight against financial fraud and cybersecurity [3]. This process is important and necessary in the context of digitalization and technical development, and is important for finance, as it concerns the personal accounts of citizens who are at risk of being victims of cybercrime, and public resources are under threat. The process of convergence of the cybersecurity system and financial crimes will ensure reliability and security in the financial sphere for the country [4].

The object of the study is the relationship between the cybersecurity system and combating financial fraud. The subject of the research is a scientific approach to the selection, evaluation and formation of indicators that characterize the cybersecurity system and combating financial fraud.

The aim of the study is to identify the level of significance of the potential convergence of the cybersecurity system and to combat financial fraud.

Objectives: to choose the factors for the analysis; to conduct statistical and visual analysis of data; using canonical analysis to identify the impact of cybersecurity factors on the fight against financial crime; choose the most influential factors; to analyze the selected factors using the neural network and regression analysis.

The study uses methods such as statistical analysis; principal components method - to eliminate multicollinearity, neural network method and regression analysis.

1. CONVERGENCE OF THE CYBERSECURITY SYSTEM AND COMBATING FINANCIAL CRIME

1.1 The essence of the convergence of the cybersecurity system and combating financial crime

Financial crimes and the growing complexity of cyberattacks have become a widespread problem for financial institutions [5]. A leading provider of corporate cybersecurity solutions, whose products and services are recognized worldwide, will help strengthen your bank's cybersecurity. The introduction of information security tools in the banking infrastructure allows to protect data and resources and create a solid foundation for compliance with regulatory requirements. The data security strategy should be comprehensive, covering people, processes and technologies [6].

Developing and approving appropriate control measures and policies to reduce cybersecurity risks is no less important than organizational culture than selecting and deploying good tools. In other words, information security should be at the forefront in all departments of the enterprise. The development of a reliable cybersecurity system not only allows you to clearly see the threats, but also helps to ensure regulatory compliance [7]. To meet regulatory requirements and counter the growing number of cyber threats and frauds, the financial industry needs tools equipped with artificial intelligence, as well as a multi-level system of protection against cyber threats that supports rapid and scalable detection and resolution of problems.

The sphere of finance - the area of circulation of monetary and currency values, as well as securities - the most important element of the domestic economy, which is actively developing [1]. Given the freedom of economic relations and the imperfections of their legal regulation, the financial sector has become one of the most attractive for criminal action. A significant number of different criminal financial transactions continue to take place here [5]. The criminal activity of the

financial sphere is characterized by the commission of a set of illegal actions invested in interfering with the movement of mainly funds or their substitutes, not related to or detached from the movement of other commodity values [1].

The concept of "financial crimes", being forensic, includes a very large group of crimes, have similarities in forensic characteristics. This concept is mostly associated with criminal acts as fraud [4].

The subject of criminal activity of this type are monetary resources that provide settlement operations [3]. There is an introduction in this area of methods of settlement operations using technologies based on the use of communication technology [1]. Large amounts of financial information are stored in the form of "electronic" documents. Criminal acts with such documents are often associated with unauthorized access to computer information [1]. It is important to keep in mind that criminal techniques can be distinguished from the real ones only with the use of special knowledge and techniques [7].

Ways of committing financial crimes are very diverse. Criminal financial transactions can be conditionally grouped as follows [9]:

• operations in the field of settlement mechanisms for money (including currency) circulation, which use the imperfection of the legal regulation of the settlement mechanism between counterparties or the lack of official control over its operation [1];

• operations in the field of circulation of payment documents or securities based on the imperfection of organizational legal and technical methods of protection of these financial instruments, banking products, etc. [1];

• operations in the field of borrowed resources, bank lending, based on the illegal receipt of funds in the form of borrowed resources, their misuse or misappropriation [8];

• fictitious lending operations accompanied by bribery of responsible bank employees and distribution of borrowed funds among criminals [1];

• operations in the field of information financial computer technologies, based on the imperfection of the means and mechanisms to protect such information

systems of financial institutions from unauthorized access to and management of such information from the outside [9].

It is characteristic of financial crimes that their actual commission is carried out in a very short time, the organizers of the crime, setting a specific goal, stop criminal acts immediately after their achievement [9].

Criminals' knowledge of banking technology allows them to avoid the rapid detection of illegal transactions and gain access to funds as quickly as possible [10]. In this case, banking instruments are actively used both to commit theft, and the introduction of criminal capital legal turnover [11].

After the crime is committed, active actions are taken to conceal it, often involving the liquidation of enterprises and financial institutions, their fictitious or intentional bankruptcy, destruction of documents, distortion of accounting, statistical and other reporting, transfer to other positions or dismissal of persons who knew something on the progress of the financial transaction [4].

The situation created or created by criminals for the possibility of committing these crimes, first of all, is formed under the influence of various inconsistencies, contradictory, undeveloped provisions of the legislation governing the financial sphere; ill-considered individual decisions and the mechanism of their implementation in the actions of relevant officials, etc. [3].

Selfish crimes are often facilitated by a situation of weak control over the order of operation of systems, weak protection of systems from unauthorized access. For example, insufficient protection against fraudulent actions of bank payment documents, insufficient quality of their production and security [1].

It is under such conditions that an effective solution would be to converge the cybersecurity system and combat financial crime.

After all, in such conditions, when the financial system of the state is vulnerable in the field of financial security and has an underdeveloped system of cyber defense - peace and reliability are not guaranteed.

The importance of cybersecurity is growing. In fact, our society is more technologically dependent than ever, and there is no indication that this trend will slow.

One of the benefits of converging cybersecurity and preventing financial crime is protecting networks and data from unauthorized access.

Financial data needs reliable protection against criminal use, which is a serious threat to the operation and existence of financial systems. Therefore, it is important to protect data from unauthorized access.

1.2 Characteristics of factors that characterize cybersecurity systems and combating financial crime.

When discussing data and information, a triad of factors must be considered. The CCA triad refers to the information security model, which consists of three main components: confidentiality, integrity and accessibility. Each component represents the main task of information security.

Therefore, how can the three components of the CDC be described:

• Confidentiality: This component is often associated with confidentiality and the use of encryption. Confidentiality in this context means that data is only available to authorized persons. If the information was confidential, it means that it was not compromised by other parties; Confidential data is not disclosed to people who do not need it or who should not have access to it. Ensuring confidentiality means that information is organized in terms of who should have access, as well as the confidentiality of data. Violations of privacy can occur in various ways, such as hacking or social engineering.

• Integrity. Data integrity refers to the assurance that data is not falsified or degraded during or after submission. This is the belief that the data has not been tampered with, intentionally or unintentionally. During the transfer process, there

are two points during which integrity may be compromised: when downloading or transferring data, or when storing a document in a database or collection.

• Accessibility: this means that information is available to authorized users when needed. For a system to demonstrate accessibility, it must have well-functioning computing systems, security controls, and communication channels. Systems that are identified as critical (power generation, medical equipment, security systems) often have extraordinary accessibility requirements. These systems must be resistant to cyber threats and have protection against power outages, equipment failures and other events that may affect the availability of the system.

Stability, accessibility and security are also three important factors to consider when creating a cyberspace, and this is especially important in finance and financial transactions.

Accessibility is a serious problem in a collaborative environment, as such an environment must be stable and constantly maintained. Such systems should also provide users with access to the necessary information with little waiting time. There may be backup systems that provide a high failure rate. The concept of availability can also refer to the ease of use of the system.

Information security of the financial sphere means maintaining integrity and secrecy during the storage or transmission of information. Information security breaches occur when information is accessed by unauthorized persons or parties. Violations may be the result of hackers, intelligence services, criminals, competitors, employees or others. In addition, individuals who value and want to maintain their confidentiality are interested in information security.

The CCA Triad describes three important components of data and information protection that can be used as a guide to establish a security policy in an organization. Establishing and maintaining an organization's security policy can be a challenge, but using a three-pillar strategic approach to cybersecurity can help you identify and manage cybersecurity risks methodically and comprehensively.

The financial services sector is a particularly important target for cyberattacks and is strictly regulated by jurisdictions around the world. Faced with constant

attempts at intrusion and other attacks, financial service providers often face difficulties in moving from a reactive position on cybersecurity to a precautionary one. Achieving this goal is hampered by the constant increase in the number of areas of attack resulting from the use of new technologies introduced through digital innovation initiatives. In addition to this complexity, there is a need to comply with the growing number of regulations on the use of financial and personal data [6].

Some factors that can counter cybercrime in the field of finance have been identified, such as:

- Cost reduction,
- Tracking
- Operational efficiency
- Flexibility
- Reporting on compliance with the requirements [1].

Regarding cost reduction, we can say that financial institutions are constantly under pressure to limit and reduce the cost of maintaining their IT environment [13]. Due to the limitation of budgets for cybersecurity, it is necessary to use a strategic approach to the allocation of financial resources as well as human resources. Due to the limited money and time of staff, a strategy is needed to limit the risk limit, as well as compromise. These problems are obviously exacerbated by a shortage of cybersecurity staff, which in turn complicates and increases the cost of finding specialists, and calls into question their ability to find them [12].

If we talk about such a factor as tracking, it is obvious that the sphere of attacks is constantly growing and increasing, thus complicating the process and the ability to protect against threats. The introduction of multi-cloud solutions for business services and the use of mobile customers leads to a rapid and widespread increase in the number of areas of attack. Given this factor, companies providing financial services need to implement more and more specialized remedies to address the problems that may arise due to the growing number of such attacks [14]. The resulting repositories have a negative impact on tracking, increasing inefficiencies and increasing risk.

The factor of operational efficiency is just as important, he understands that without the same efficiency, there is a lack of integration between different elements of security and fragmentation of the architecture, thereby increasing operational inefficiency [15]. In the absence of integration, many workflows have to be managed manually. In addition to delaying the detection, prevention and response to threats, architectural repositories create redundancy, increase operating costs and lead to potential gaps in the organization's cybersecurity system [4].

In the area of finance, the security architecture must be flexible enough to ensure high speed, security and interoperability of public, private and hybrid cloud services while protecting local services, all of which are integral to successfully countering the increase in cybercrime. financial services are increasingly using cloud programs and infrastructure, and this area is therefore more vulnerable and needs good, reliable cybersecurity.

Compliance reporting is an important factor that monitors how well the system is working and whether there are any urges to believe that there is an element of instability or error that could cause serious problems such as information leakage, resource theft and similar problems. in this area [16]. Financial services are one of the most rigorous industries in the world, and all financial data, personal and corporate, is stored online, from campus to data center, peripherals and the cloud. Organizations must demonstrate compliance with several norms and standards. Employees performing strategic tasks should not be involved in preparing safety reports manually.

There are also some factors that characterize the convergence of the cybersecurity system and the fight against financial crime. These factors include

• The Global Cybersecurity Index (GCI), a reliable benchmark that measures countries' commitment to cybersecurity globally - to raise awareness of the importance and different dimensions of the problem. Because cybersecurity has a wide scope, covering many industries and different sectors, each country's level of development or involvement is assessed in five pillars: legal, technical,

organizational, capacity building and cooperation - and then summarized in an overall assessment [17]. .

• The Network Readiness Index is an index published annually by the World Economic Forum in collaboration with INSEAD as part of their annual report on global information technology. It aims to measure the degree of readiness of countries to use the opportunities offered by information and information. communication technologies [12].

• The National Cyber Security Index measures a country's level of cybersecurity, its readiness to prevent cyber threats, and its readiness to manage cyber incidents, crime, and large-scale crises. NCSI's vision is to develop a comprehensive cybersecurity measurement tool that provides accurate and up-to-date public information on national cybersecurity [18]. The NCSI focuses on measurable aspects of cybersecurity implemented by the central government and aims to identify which policy and strategy gaps need to be filled to improve a country's cybersecurity [13].

• The level of digital transformation is the process of completely replacing manual, traditional and outdated ways of doing business with the latest digital alternatives. This innovation affects all aspects of business, not just technology. This feature opens up much more than just improving individual processes. It allows you to transform any industry on a large scale [14].

• Political Stability Index - an institutional indicator that shows stability in political and governmental issues in countries. This also includes political violence [15].

• The Government Performance Index is an index developed by the World Bank Group that measures the quality of public services, civil service, policy making, policy implementation and confidence in the government's commitment to improving or maintaining these qualities at a high level [15].

• Ease of Doing Business Index is an index published by the World Bank. This is an aggregate indicator that includes various parameters that determine the ease of doing business in the country [19].

• The crime index is a powerful but easy to understand crime rating on a scale from 0 to 100, where 100 is the safest. The index is based on the crime rate per 1,000 population for all crimes in a given district or city [16].

• The Global Terrorism Index (GTI) is a comprehensive study that analyzes the impact of terrorism for 163 countries, covering 99.7% of the world's population, and is used to measure the impact of terrorism [17].

• The Financial Secrets Index (FSI) is a report published by the human rights organization Tax Justice Network, which ranks countries by financial secrecy indicators weighted by each country's economic flows [17].

• The consumer price index is an indicator of the average change over time in the prices paid by urban consumers for the market basket of consumer goods and services [17].

1.3 Conceptual research model

The conceptual model is a representation and description of the system [20]. A model consists of concepts that are used to help people learn, understand, or model an object that represents a model.

The analysis of the urgency of the problem of convergence of the cybersecurity system and counteraction to financial crimes aims to identify the need for the use of innovations, changes that are possible in the course of identifying the relationships between these systems.

Statistical analysis is performed to study and present certain amounts of data, as well as to identify key patterns and trends. In this case, statistical analysis is conducted to study the behavior of factors related to the cybersecurity system and the system of combating financial problems and fraud in different countries.

The conceptual model of the study is shown in Figure 1.1

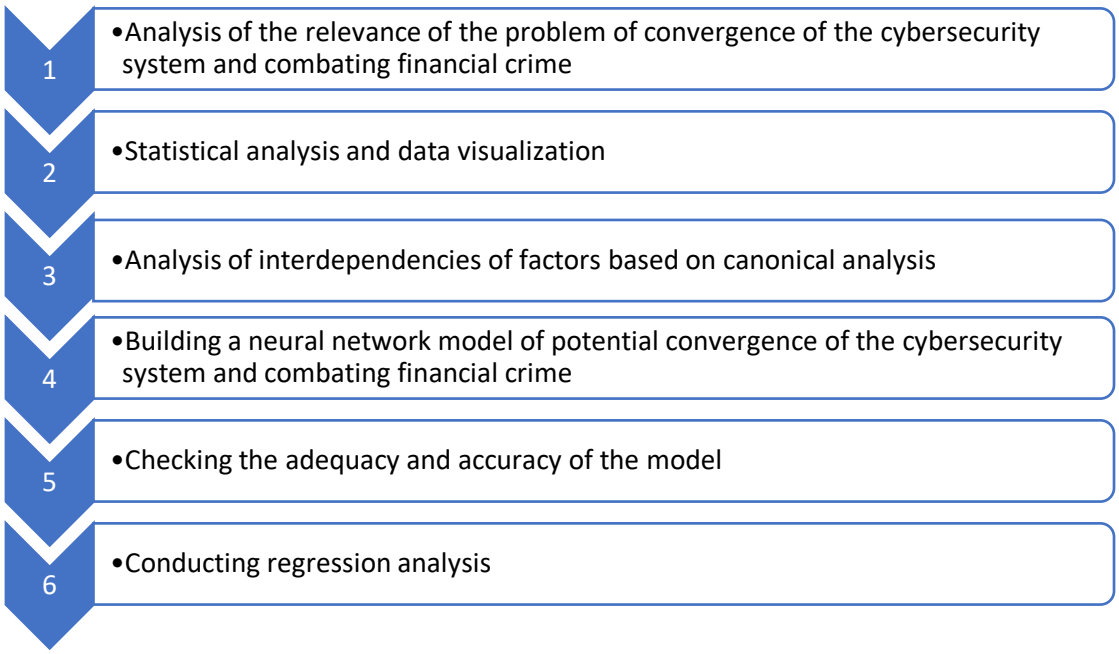


Figure 1.1 - conceptual model of the study

Data visualization is a graphical representation of information and data. Using visual elements such as charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, deviations, and patterns in data. Data visualization is also performed to better understand the distribution of indicators, how often, how and where certain factors predominate.

Canonical analysis is a multifactor technique that aims to determine the relationships between groups of variables in a data set. This type of analysis is used to understand the dependence of cybersecurity factors and counter financial fraud. With the help of canonical analysis, the most influential factors can be identified. This method works by finding a linear combination of variables X, ie X1, X2, etc., and a linear combination of variables Y, ie Y1, Y2, etc., which are most correlated [21].

Neural networks are prediction methods based on simple mathematical models. In the study, the neural model can be used to analyze and classify data according to specified parameters, to form analytical predictions based on input information, as well as to compare and recognize identical data.

Regression analysis is a reliable method of determining which variables are most influential. The regression process allows you to confidently determine which

factors are most important, which factors can be ignored and how these factors affect each other.

2. STATISTICAL ANALYSIS OF THE POTENTIAL CONVERGENCE OF THE CYBERSECURITY SYSTEM AND COMBATING FINANCIAL CRIME

2.1 Calculation and analysis of basic statistics

The analysis of basic statistics was carried out using statistical methods of data processing, their systematization, visual representation of both tables and graphs, as well as quantitative description of data using a system of statistical indicators.

Statistical analysis was performed using the Python programming language.

Python is a high-level general-purpose programming language with dynamic rigorous typing and automatic memory management, focused on improving developer performance, code reading and quality, as well as ensuring the portability of programs written on it [7]. In this paper, the Python programming language was used to calculate basic statistics and visualize data on the convergence of cybersecurity and countering financial crime [22].

The first step was to import the necessary libraries, such as: pandas, numpy, preprocessing, matplotlib.pyplot and some others (Appendix A). The pandas library is used to support Python not only for data collection and purification, but also for data analysis and modeling tasks, without switching to more statistical-specific language processing [10].

The NumPy library provides implementations of computational algorithms (in the form of functions and operators), optimized for working with multidimensional arrays [11].

Matplotlib is a Python programming language library for data visualization with two-dimensional graphics [12].

The .head () function was used to display the data (Fig. 2.1).

```
df.head()
```

	Country	GCI	ICTDI	NRI	NCSI	DDL	PSI	GEI	EDB	CI	CPI	GTI	FCI
0	Australia	89	82	79	59.74	80.49	1.00	1.60	80.14	42.55	77	2.827	244.358302
1	Austria	83	80	77	68.83	78.67	0.91	1.45	78.54	20.41	76	1.852	310.412705
2	Bahrain	59	76	73	25.97	74.43	-0.84	0.18	68.03	36.96	36	3.883	490.706709
3	Barbados	17	73	0	15.58	73.10	0.92	0.43	56.78	51.31	68	0.000	230.952985
4	Belgium	81	78	77	85.71	77.62	0.41	1.17	71.71	42.17	75	4.060	212.965184

Figure 2.1 - Display of input data.

Figure 2.1 presents data that characterize the potential process of convergence of the cybersecurity system and combating financial crime.

There are 14 columns, the first column indicates the serial number, the column called "Country" contains a list of countries, from 3 to 14 columns presents statistics such as GCI, ICTDI, NRI, NCSI, DDL, PSI, GEI, EDB, CI, CPI, GTI, FCI, respectively.

GCI - Global Cyber Security Index

ICTDI - Index of Information and Communication Technology Development

NRI - Network Readiness Index

NCSI - National Cyber Security Index

DDL - Level of digital transformation

PSI - Index of Political Stability

GEI - Government Performance Index

EDB - Ease of Doing Business Index

CI - Crime Index

GTI - Global Terrorism Index

CPI - Consumer Price Index

FCI - Index of Financial Secrecy

The next step was to calculate the basic statistics for each of these indicators. The main statistics include: the total number of observations, mean, standard deviation, minimum value, and maximum value. (Fig.2.2)

The .describe () function calculates and displays summary statistics. (Fig.2.2).

The average value of the sample characterizes the location of the values of the random variable and indicates the center of data scattering.

Standard deviation is the most common indicator of the scattering of values of a random variable relative to its mathematical expectation.

Minimum value - indicates the smallest value of the sample in the specified data interval.

The maximum value, respectively, indicates the largest value of the sample.

```
df.describe()
```

	GCI	ICTDI	NRI	NCSI	DDL	PSI	GEI	EDB	CI	CPI	GTI	FCI
count	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000	76.000000
mean	66.078947	65.078947	61.894737	54.255000	65.576184	0.322763	0.633684	70.199342	42.055000	55.342105	2.143276	284.696287
std	24.289786	18.071534	19.904826	23.195725	13.973113	0.779067	0.848850	10.234283	14.360367	18.745704	2.317043	279.032311
min	2.000000	0.000000	0.000000	3.900000	28.100000	-1.860000	-1.580000	30.850000	13.100000	18.000000	0.000000	27.860721
25%	56.000000	56.000000	57.000000	35.060000	57.725000	-0.227500	0.040000	64.265000	33.897500	40.500000	0.052250	127.230995
50%	75.000000	69.500000	63.500000	57.140000	66.815000	0.465000	0.495000	71.825000	40.170000	55.000000	1.011500	208.255223
75%	85.000000	79.000000	77.000000	71.755000	78.145000	0.950000	1.272500	78.095000	49.292500	72.250000	3.958000	355.705963
max	93.000000	90.000000	86.000000	96.100000	85.130000	1.540000	2.230000	86.590000	83.600000	88.000000	7.568000	1589.573888

Figure 2.2 - Basic statistics.

The figure shows that the total number of observations for each indicator is 76. The mean, standard deviation, maximum and minimum values are different.

For GCI:

• average value - 66.078947

• standard deviation - 24.289786

• the minimum value is 2.000000

• maximum value - 93.000000

For ICTDI:

• average value - 65.078947

• standard deviation - 18.071534

• the minimum value is 0.000000

• maximum value - 90.000000

For NRI:

• average value - 61.894737

• standard deviation - 19.904826

• the minimum value is 0.000000

• maximum value - 86.000000

For NCSI:

• average value - 54.255000

• standard deviation - 23.195725

• minimum value - 3.9 million

• maximum value - 96,100,000

For DDL:

• average value - 65.576184

• standard deviation - 13.973113

• the minimum value is 28,100,000

• the maximum value is 85.130000

For PSI:

• average value - 0.322763

• standard deviation - 0.779067

• the minimum value is -1.860000

• maximum value - 1.540000

For GEI:

• average value - 0.633684

• standard deviation - 0.848850

• the minimum value is 1.580000

• the maximum value is 2.230000

For EDB:

• average value - 70.199342

• standard deviation - 10.234283

• the minimum value is 30.850000

• maximum value - 86.590000

For CI:

• average value - 42.055000

• standard deviation - 14.360367

• the minimum value is 13,100,000

• maximum value - 83.6 million

For CPI:

• average value - 55.342105

• standard deviation - 18.745704

• the minimum value is 18.000000

• maximum value - 88.000000

For GTI:

• average value - 2.143276

• standard deviation - 2.317043

• the minimum value is 0.000000

• the maximum value is 7,568,000

For FCI:

• average value - 284.696287

• standard deviation - 279.032311

• the minimum value is 27.860721

• maximum value - 1589.573888

2.2 Visualization of the main factors

Data visualization is the presentation of data in a form that provides the most efficient work of the person who studies them. Data visualization is widely used in many areas, such as: scientific and statistical research, in pedagogical design for teaching and testing, in news reports and analytical reviews [17].

Data visualization helps to achieve results, assess the value of information or data. Data visualization refers to the presentation of information in graphical form, for example, in the form of a pie chart, graph or visual representation of another type [12].

Graphs are convenient to use if you want to depict the nature or general trend of the phenomenon or phenomena. The lines are convenient for the image of several time series of their comparison, when you need to compare the growth rate [16].

Histogram is one of the few ways to graphically represent data, the availability and ease of perception of which is beyond doubt. It is excellent for describing large data sets, as well as for characterizing a small numerical series [16].

Without exaggeration, histograms are one of the most important tools for data analysis. Presenting the results of observations with the help allows you to evaluate a number of statistical indicators, draw conclusions about the distribution functions and identify possible deviations, as well as compare data sets [16].

The matplotlib library was used to visualize pandas data. With its help you can easily build charts [14]. Using the already imported matplotlib module. pyplot and the plot () method were plotted.

```
df.plot(figsize=(12,6))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7faa7d4aad50>
```

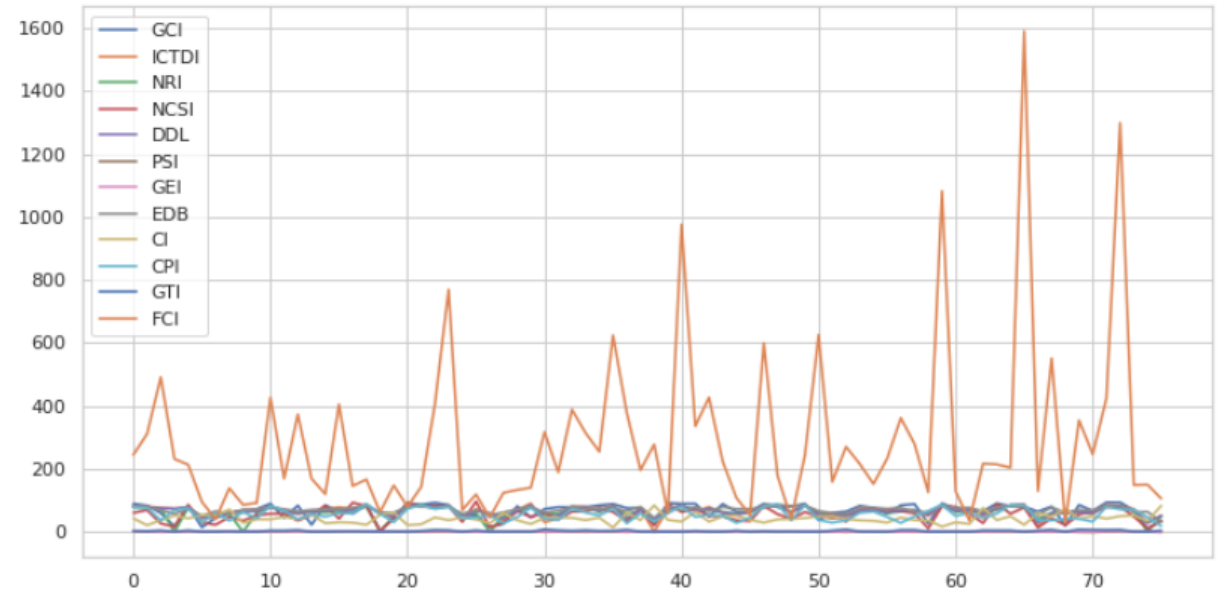


Figure 2.3 - Graph of data distribution.

Figure 2.3 shows the distribution of all data. As you can see, the indicators, except for the FCI, are approximately within the same range. The FCI is different, the value of this characteristic is much higher than the values of others. The maximum and minimum values of this indicator are 1589.573888, 27.860721,

respectively. In order to better understand the distribution of other indicators, another graph was made, it describes all the input data, except FCI. (Fig.2.4).

```
df.plot( y=[ "GTI", "ICTDI","NRI","NCSI","PSI","GEI",
             "EDB","CI","CPI","GTI"])
plt.ylabel("Frequency")
plt.show()
```

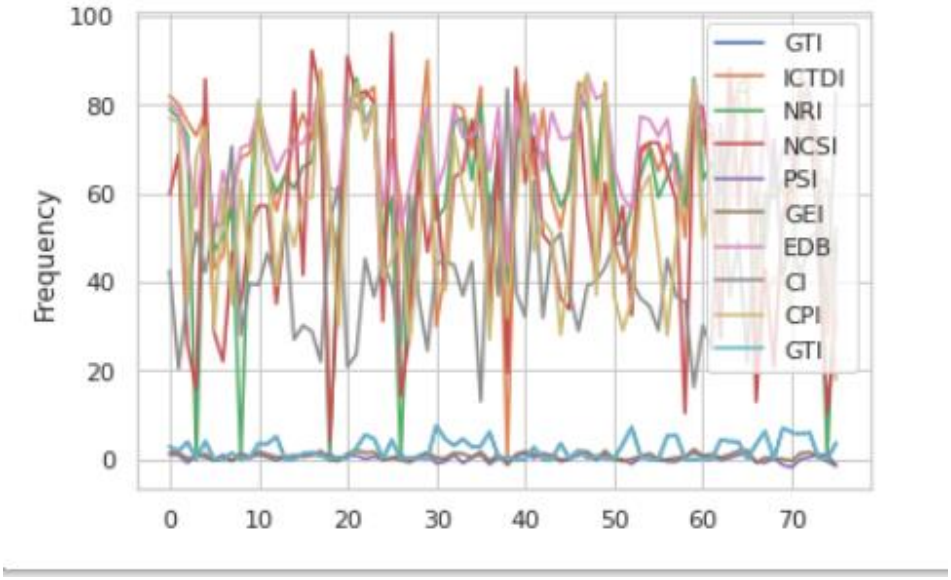


Figure 2.4 - Graph of data distribution.

Figure 2.4 describes the data distribution. The total number of observations described in the figure is 76, the values of indicators vary from 0 to 100.

The next step is to build a histogram for each indicator separately.

```
median_column.plot(kind="hist")
plt.title("GCI")
```

```
Text(0.5, 1.0, 'GCI')
```

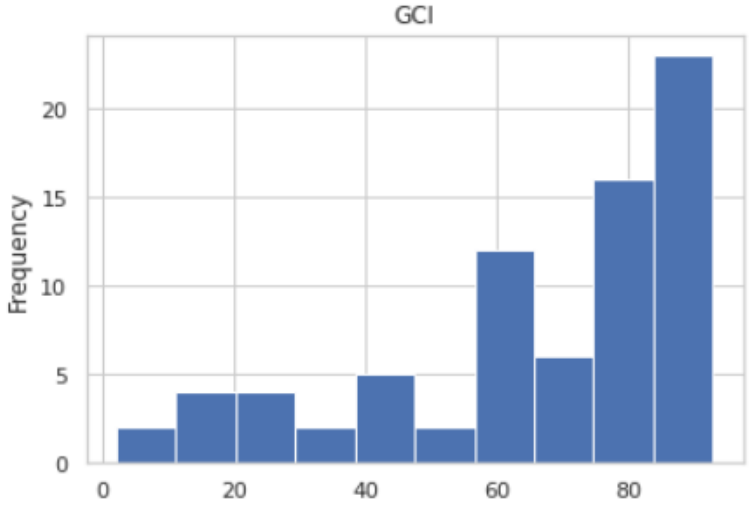


Figure 2.5 - GCI distribution histogram

As can be seen from Figure 2.5, most values are in the range of 60-93. The number of observations from 0 to 60 is much less repeated.

```
df.hist('ICTDI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

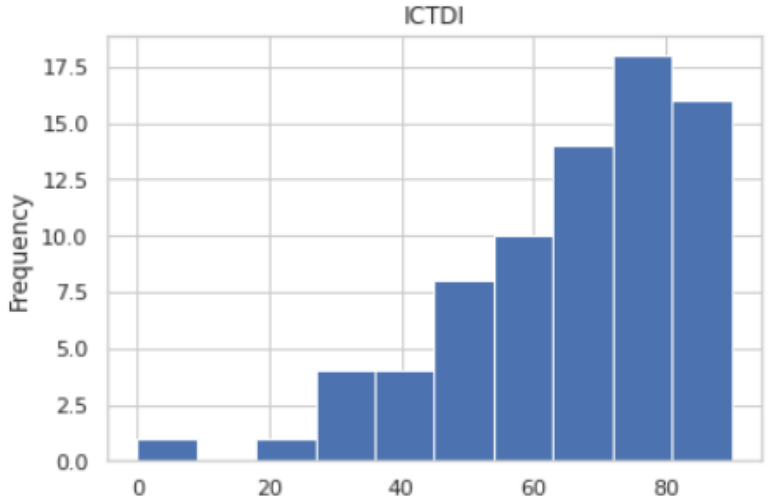


Figure 2.6 - Histogram of the distribution of the indicator "ICTDI"

Figure 2.6 shows that the largest number of observations is concentrated in the range from 70 to 80. That is, these values are most often repeated in the sample.

```
df.hist('NRI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

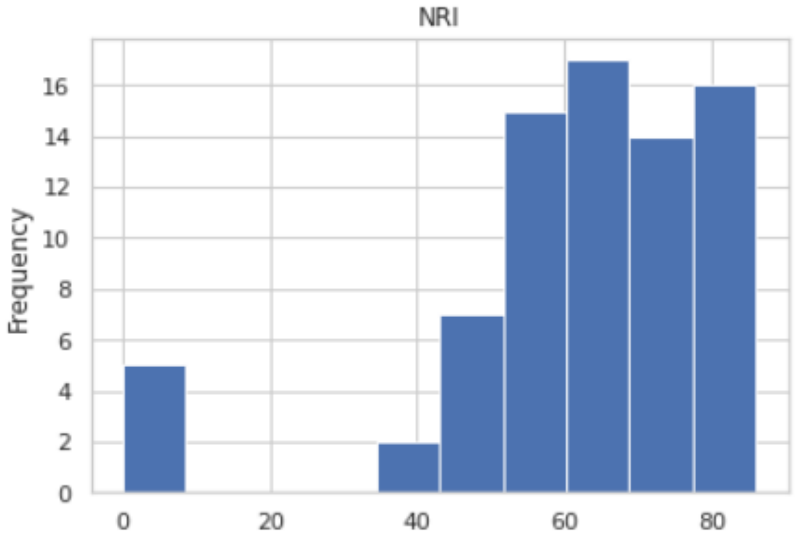


Figure 2.7 - Histogram of the distribution of the indicator "NRI"

Figure 2.7 describes the characteristics of "NRI". The largest values are in the range from 50 to 90, and are repeated more times than the values from 0 to 50.

```
df.hist('NCSI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

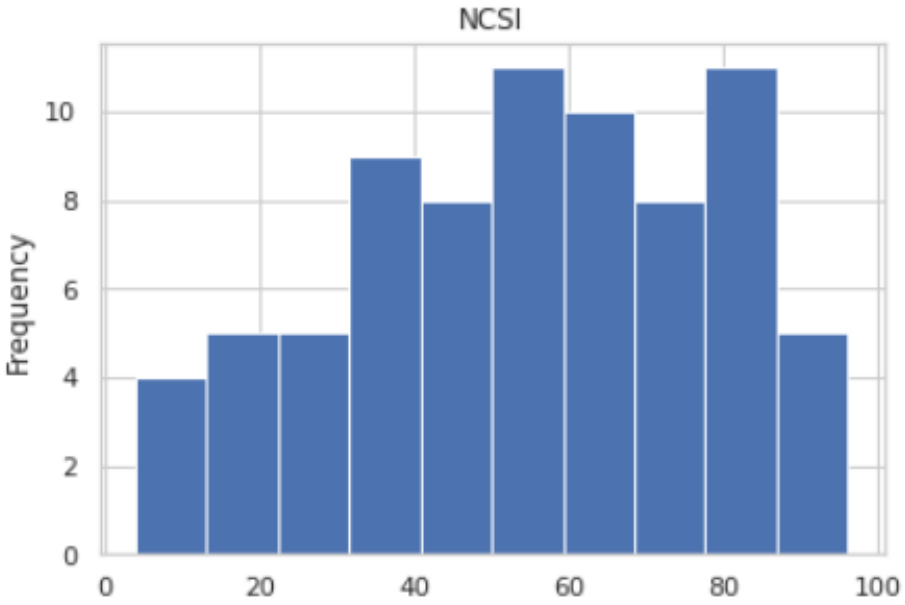


Figure 2.8 - Histogram of the distribution of the indicator "NCSI"

In Figure 2.8, compared to the previous ones, the data are distributed more evenly, the values from 40 to 90 are most often repeated. Other values are less common.

Figures 2.9, 2.10, 2.11, 2.12, 2.13 show the peculiarities of the distribution of factors "PSI", "EDB", "CI", "CPI", "GTI", respectively.

```
df.hist('PSI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

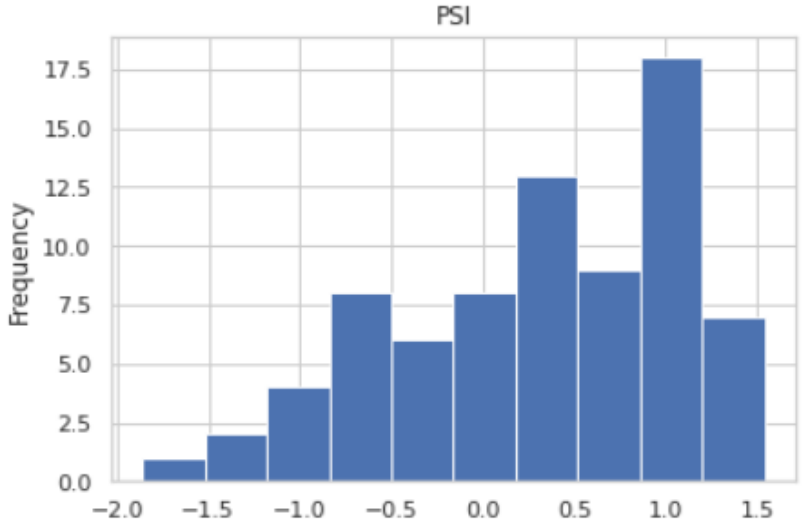


Figure 2.9 - Histogram of the distribution of the indicator "PSI"

```
df.hist('EDB')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

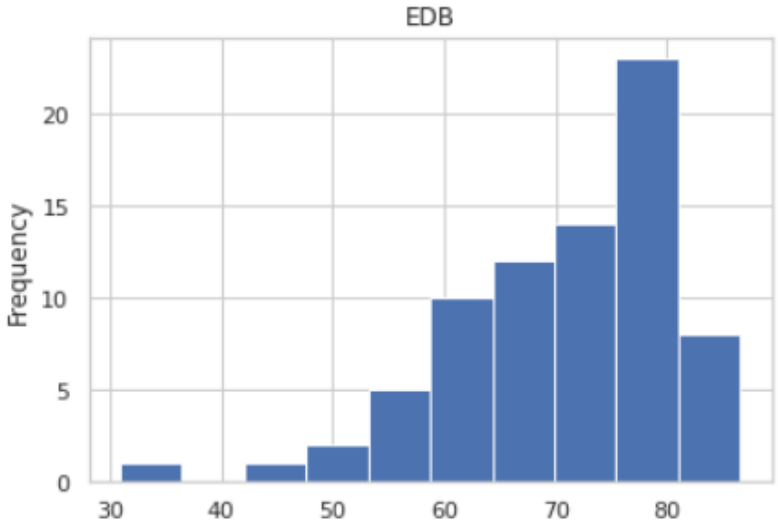


Figure 2.10 - Histogram of the distribution of the indicator "EDB"

```
df.hist('CI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

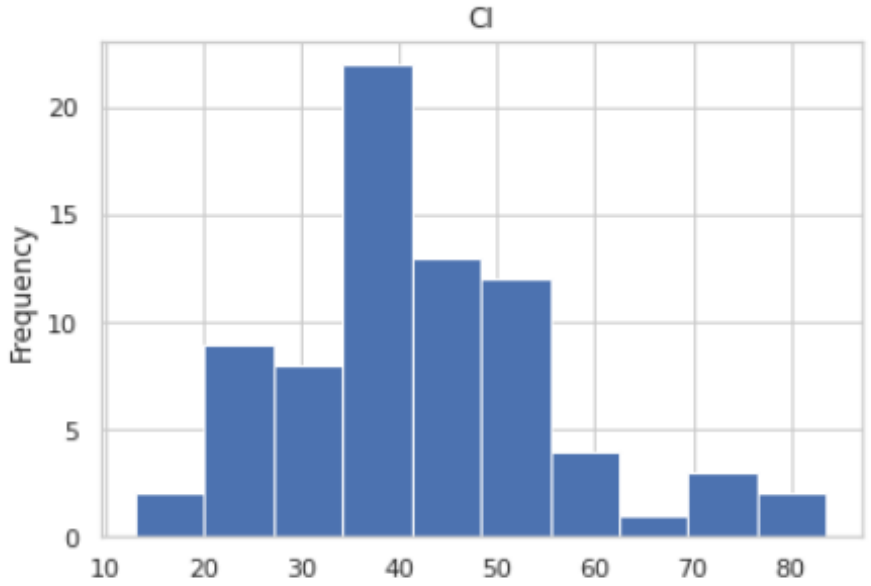


Figure 2.11 - Histogram of the distribution of the indicator "CI"

```
df.hist('CPI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

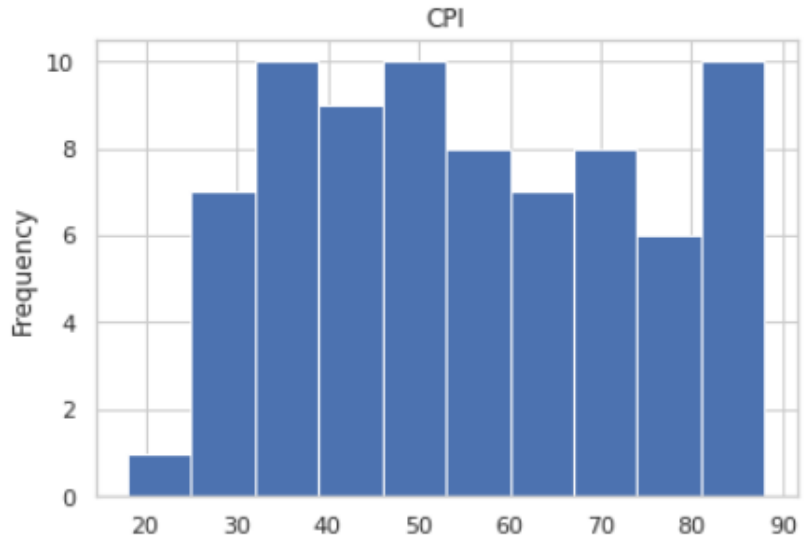


Figure 2.12 - Histogram of the distribution of the indicator "CPI"

```
df.hist('GTI')
plt.ylabel("Frequency")
```

```
Text(0, 0.5, 'Frequency')
```

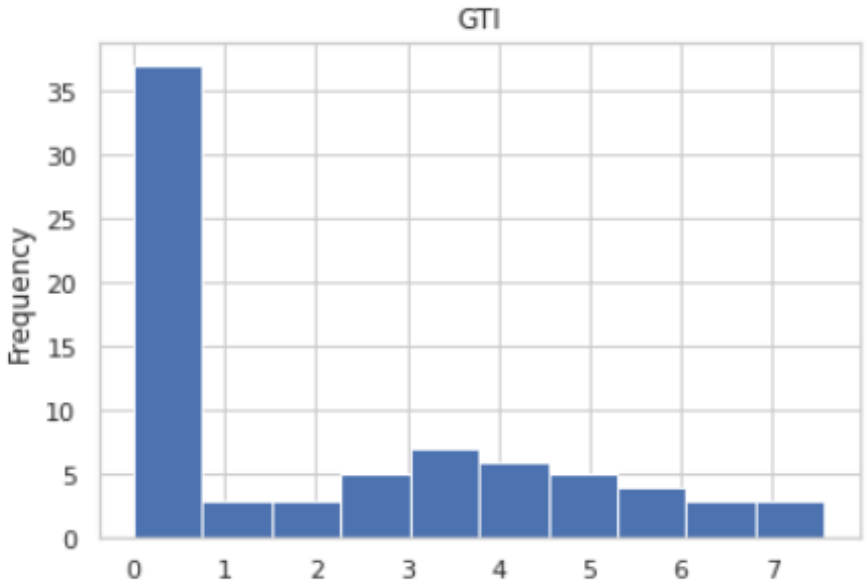


Figure 2.13 - Histogram of the distribution of the indicator "GTI"

2.3 Analysis of interdependencies of factors based on canonical analysis

Canonical analysis is a multidimensional method of analysis that involves determining the relationships between groups of variables in a data set. The main purpose of canonical analysis is to find the maximum correlations between groups of source variables [18].

Canonical data analysis was performed using the STATISTICA program.

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .96635 Chi²(35)=297.75 p=0.0000 Left Set	**Right Set**
No. of variables	5	7
Variance extracted	100.000%	85.9457%
Total redundancy	58.7705%	58.7748%
Variables: 1	GCI	PSI
2	ICTDI	GEI
3	NRI	EDB
4	NCSI	CI
5	DDL	CPI
6		GTI
7		FCI

Figure 2.14 - the results of canonical analysis

The obtained value of the canonical R is large enough and equal to 0.96635. The canonical value of R, shown in Figure 2.14, suggests that there is a strong correlation between the factors that characterize cybersecurity and financial crime prevention systems. Pearson's test, which in this case is equal to 297.75, confirms the statistical significance of the correlation coefficient, and the level of significance of this coefficient does not exceed 0.05 (p = 0.0000). The value of the left set, which was formed from the indices of the cybersecurity system, is 58.7705%. This value suggests that the factors described in the right set, which characterize the level of resistance to financial crime, explain by 58.7705% the variability of factors in the cybersecurity system [27]. The system of counteraction to financial crimes to some extent depends on the system of cybersecurity in the country, apparently, the factors of the cybersecurity system by 58.7748% explain the variability of factors that characterize the level of counteraction to financial fraud. The analysis of indicators shows that the factors of the cybersecurity system have an impact on the process of combating financial crime [27].

The next step was to analyze the impact of each factor of the cybersecurity system on combating financial fraud.

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .62413 Chi²(7)=34.794 p=.00001 **Left Set**	Right Set
No. of variables	1	7
Variance extracted	100.000%	18.2672%
Total redundancy	38.9537%	7.11577%
Variables: 1	GCI	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Figure 2.15 - Results of canonical analysis.

Figure 2.15 shows the results of the canonical analysis of the GCI cybersecurity factor and anti-fraud factors.

As can be seen, the obtained value of the canonical R is not large enough (R = 0.62413). This suggests a low correlation between the Global Cybersecurity Index and the Prevention of Financial Crime. Pearson's test, which is 34,794, also confirms the statistical insignificance of the correlation coefficient. The value of the left set is 38.9537% and suggests that the factors that describe the fight against financial crime, 38.9537% explain the variability of the global cybersecurity index [27].

As can be seen from Figure 2.15, a very small percentage of countering financial crime factors depend on the chosen factor of the cybersecurity system. The factor of the global cybersecurity index of only 7.11577% explains the variability of factors in combating financial crime [27]. The value obtained suggests that the impact of the global cybersecurity index on combating financial fraud is low and has no significant impact on the system of combating financial fraud [27].

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .75906 Chi²(7)=60.519 p=0.0000 Left Set	Right Set
No. of variables	1	7
Variance extracted	100.000%	21.8918%
Total redundancy	57.6170%	12.6134%
Variables: 1	ICTDI	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Figure 2.16 - Results of canonical analysis.

Figure 2.16 shows the results of the analysis of the impact of the index of development of information and communication technologies on combating financial fraud [27]. As can be seen from Figure 2.16, the obtained value of R is low, which indicates a low correlation between the selected factors. Pearson's coefficient, in this case, is 60,519, which confirms the statistical insignificance of the correlation coefficient. The value of the factors of the left set is 57.6170%. This shows that the factors of counteraction to financial fraud by 57.6170% describe the variability of the factor of the cybersecurity system [27]. The value of the factors of the right set is equal to 12.6134%. This describes that the index of development of information and communication technologies by 12.6134% describes the variability of factors in combating financial crime. The obtained values indicate that the impact of the index of development of information and communication technologies on combating financial crimes is, but it is not so great.

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .71093 Chi²(7)=49.636 p=0.0000 Left Set	Right Set
No. of variables	1	7
Variance extracted	100.000%	15.8106%
Total redundancy	50.5428%	7.99112%
Variables: 1	NRI	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Figure 2.17 - Results of canonical analysis.

Figure 2.17 describes the impact of the Network Readiness Index on financial crime response factors. The value of the obtained R is low, and this indicates a low correlation between factors [27]. Pearson's coefficient confirms the assumption of statistical insignificance of the correlation coefficient. The value of the factors of the left set is 50.5428%. This suggests that the factors of counteraction to financial fraud by 50.5428% describe the variability of the network readiness index. The value of the factors of the right set is equal to 7.99112%. This suggests that the network readiness index of 7.99112% describes the variability of factors in combating financial crime [28].

The obtained results indicate a low influence of the network readiness factor on the factors of counteraction to financial crimes.

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .74559 Chi²(7)=57.225 p=0.0000	
	Left Set	Right Set
No. of variables	1	7
Variance extracted	100.000%	42.1733%
Total redundancy	55.5897%	23.4440%
Variables: 1	NCSI	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Figure 2.18 - Results of canonical analysis.

The correlation between the factors of counteraction to financial crimes and the national cybersecurity index is weak, as evidenced by the correlation coefficient equal to 0.74559 (Fig.2.18). Pearson's criterion is 57,225 and confirms that the correlation coefficient is not statistically significant.

The value of redundancy for the left set, which consists of a factor of the cybersecurity system, namely the factor "National Cybersecurity Index" is 55.5897%. This means that the factors of the right set, which consist of indices for combating financial crimes, by 55.5897% explain the variability of the cybersecurity system [27]. The fight against financial fraud depends in part on the national cybersecurity index, as the factor of the cybersecurity system by 23.4440% describes the variability of the system of combating financial crimes. Although the values obtained are moderate, this is enough to prove the small impact of the National Cyber Security Index on combating financial fraud in countries [28].

N=76	Canonical Analysis Summary (Convergensy.sta) Canonical R: .95472 Chi²(7)=170.94 p=0.0000 Left Set	Right Set
No. of variables	1	7
Variance extracted	100.000%	43.6225%
Total redundancy	91.1491%	39.7615%
Variables: 1	DDL	PSI
2		GEI
3		EDB
4		CI
5		CPI
6		GTI
7		FCI

Figure 2.19 - Results of canonical analysis.

Figure 2.19 shows the results of the analysis of the impact of the factor "Level of digital transformation" on combating financial crime in countries.

As can be seen, the obtained value of the canonical R = 0.95472. This suggests that there is a strong correlation between the factors that characterize the level of digital transformation and the fight against financial crime.

Pearson's test, which is equal to 170.94, and the level of significance of which does not exceed 0.05 (p = 0.0000), confirms the statistical significance of the correlation coefficient. The redundancy value for the left set, which consists of a factor of the cybersecurity system, namely the "Level of Digital Transformation", is 91.1491%. This indicates that the factors of the right set, which describe the fight against financial crimes, 91.1491% explain the variability of the index of the level of digital transformation, which indicates a high value of influence. The process of combating financial fraud in the country depends on the cyber protection of financial systems, as the index of digital transformation at 39.7615% describes the variability of factors that characterize the fight against financial fraud in countries [28]. The value obtained is high and this indicates that the cybersecurity system (digital transformation level index) has a strong influence on combating financial fraud. Thus, in the process of analyzing the impact of cybersecurity factors, one index was

identified that has a strong influence on the factors that characterize the fight against financial crime, this factor is the "level of digital transformation".

3. BUILDING A NEURAL NETWORK MODEL OF POTENTIAL CONVERGENCE OF THE CYBERSECURITY SYSTEM AND COMBATING FINANCIAL CRIME

3.1 Elimination of multicollinearity of factors using the method of principal components

The principal components method is a technique for reducing the dimensionality of data sets, increasing interpretation, but at the same time minimizing information loss [26]. This is done by creating new uncorrelated variables that consistently maximize variance [18].

The principal components method and further analysis are performed in the Python programming language. The necessary step was to assign values of X and Y.

	DDL	PSI	GEI	EDB	CI	CPI	GTI	FCI
0	80.49	1.00	1.60	80.14	42.55	77	2.827	244.358302
1	78.67	0.91	1.45	78.54	20.41	76	1.852	310.412705
2	74.43	-0.84	0.18	68.03	36.96	36	3.883	490.706709
3	73.10	0.92	0.43	56.78	51.31	68	0.000	230.952985
4	77.62	0.41	1.17	71.71	42.17	75	4.060	212.965184

Figure 3.1 - Input data

The value of X was assigned the index "DDL", which is responsible for the level of digital transformation, and characterizes the cybersecurity system. This index was chosen in the process of canonical analysis as the most influential factor in combating financial fraud (Fig. 2.21).

The value of B was assigned indices that characterize the fight against financial fraud in countries, namely: "PSI", "GEI", "EDB", "CI", "CPI", "GTI", "FCI".

The indicators that were selected have a high level of multicollinearity. Since multicollinear indicators cannot be further processed with this data, a neural network

cannot be constructed with multicollinear factors. The method of the main components was used for further analysis. The essence of this method is to reduce the data set and create new components that do not have multicollinearity. After plotting and performing calculations, it is seen that only 4 of the 7 components were obtained. As can be seen from the table (Fig. 3.3), 4 components accumulate a variation of 93% and the level of significance of each should not be less than 0.05.

This method was used to build a neural model based on new factors.

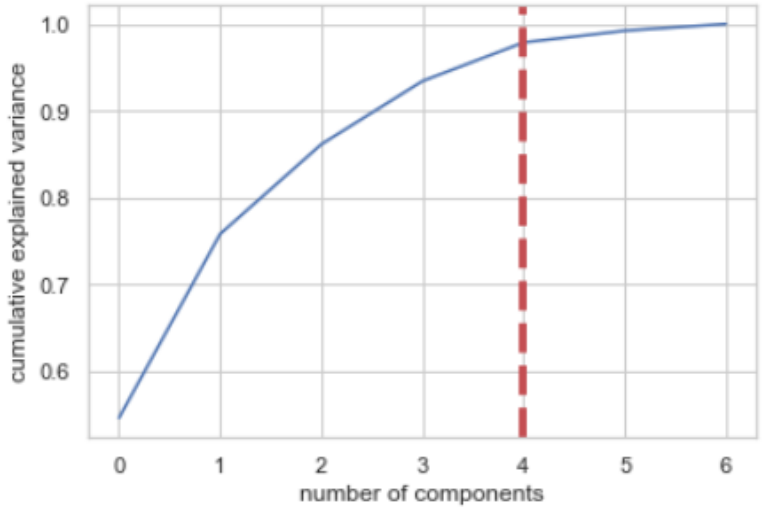


Figure 3.2 – The graph of new components.

	Cumulative Variance Ratio	Explained Variance Ratio
0	0.545733	0.545733
1	0.757906	0.212173
2	0.861269	0.103363
3	0.934204	0.072935

Figure 3.3 - Results of the calculation.

3.2 Construction of a neural network model

A neural network is a series of algorithms designed to recognize the basic connections in a data set.

Neural networks can adapt to changing input data; thus the network generates the best possible result without the need to redesign the output criteria.

The relu activation function was used to construct the neural network.

ReLU is a nonlinear activation function. This feature is the most commonly used feature. It is used for convolutional neural networks and deep learning for all layers except the original [24].

Figure 3.4 shows a neural network that shows actual and predicted values, as well as an estimate. It turns out that the criterion of determination is equal to 0.799 according to the test results and the assessment of the training coefficient of determination is equal to 0.821 (Fig. 3.5). The model contains three layers, each layer contains 45 nodes. This model is made manually by selection (Appendix B).

Estimates such as, mean absolute error (equal to 3,47495173501873) and mean square error (equal to 26,07682730734608) describe a comparison of the quality of the predicted data and the actual. As can be seen from Figure 3.5, the errors are insignificant and this indicates a high quality forecast.

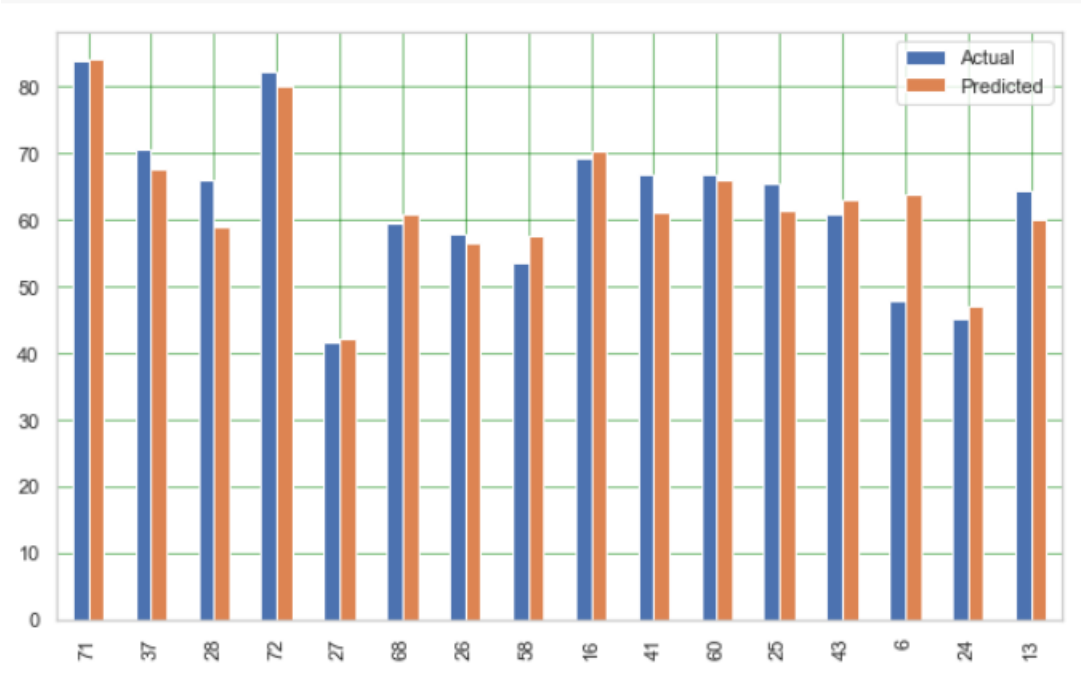


Figure 3.4 - Neural network

```
Mean Absolute Error: 3.47495173501873
Mean Squared Error: 26.07682730734608
Root Mean Squared Error: 5.106547493889201
Test R^2 Score : 0.799
Training R^2 Score : 0.821
```

Figure 3.5 - The result of the calculation

The next stage is to build a loss curve. One of the most commonly used graphs for neural network debugging is the learning loss curve. Figure 3.6 characterizes the behavior of indicators. This curve measures the error of the model and shows "how badly the model works." All these indicators record the performance of the model, so the higher they are, the better the model becomes. In this case, the number of losses decreases, the curve decreases over time and equalizes to a certain level.

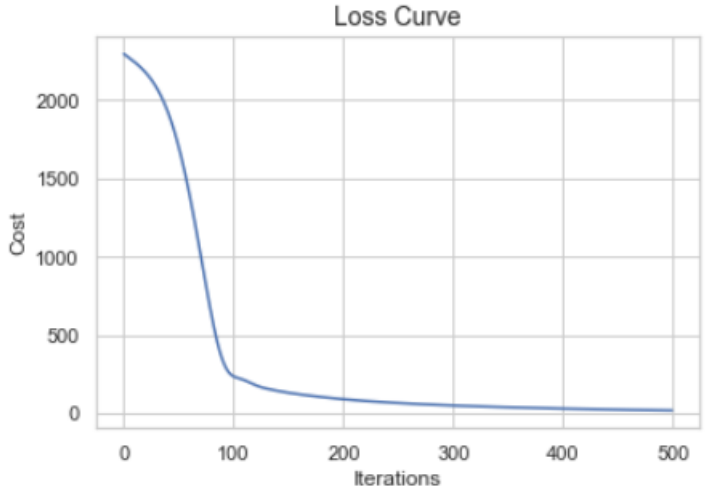


Figure 3.6 - graph of losses

The coefficients obtained for the neural network are shown in Figure 3.7, 3.8.

```
[(4, 45), (45, 45), (45, 45), (45, 1)]
[array([[-0.26004826, -0.12029681,  0.20590884, -0.06945696, -0.3113571 ,
         0.21721447, -0.30771875, -0.10858792, -0.2631856 , -0.21179765,
         0.03922925,  0.34470119,  0.00092399, -0.03584113, -0.21201222,
         0.39837075,  0.21987484, -0.09763449, -0.34011139, -0.13194794,
        -0.32285526,  0.35387933,  0.0824541 , -0.32409041, -0.31574998,
         0.01327073, -0.374209  , -0.2635563 ,  0.23914328, -0.16130783,
        -0.08460061, -0.0878057 , -0.27154981, -0.22982154,  0.18132776,
         0.0032163 , -0.04726616, -0.38168097, -0.05618342, -0.0240545 ,
        -0.01425243,  0.09723332,  0.25665729, -0.27056029,  0.04194362],
       [-0.37137422,  0.29138754,  0.4493272 , -0.09557417,  0.32288134,
        -0.40760883,  0.12893772, -0.281831  , -0.24050574,  0.45293821,
        -0.12016031,  0.40809908,  0.20684486, -0.14231219, -0.36514533,
        -0.4131254 ,  0.36199315, -0.07692699, -0.33685647,  0.35394341,
        -0.43567026,  0.46121382,  0.26172665,  0.32925829,  0.08731418,
        -0.21769305,  0.37403948, -0.31338637,  0.16617805,  0.32513892,
         0.27063393, -0.14837558, -0.05335017, -0.26485246,  0.47923464,
        -0.23381631,  0.43610618, -0.272976  , -0.34756817, -0.39744876,
         0.23763593, -0.327351  , -0.09516424, -0.16217758,  0.4214515 ],
       [-0.34914899,  0.03479934, -0.38569757, -0.05618054, -0.32536316,
         0.3083773 , -0.50888427,  0.03729243,  0.12824579, -0.29934257,
         0.18971017,  0.03014801, -0.32204607, -0.06726018,  0.35458359,
         0.03199527, -0.16551574,  0.27497358, -0.08863406, -0.07647509,
        -0.34039611,  0.16389043,  0.15351833,  0.11692195, -0.07506765,
        -0.0310238 ,  0.02180154,  0.3428187 , -0.37137128, -0.03728706,
        -0.46024447,  0.12702506, -0.55350416,  0.06861506,  0.0110601 ,
         0.08716832, -0.06640432, -0.26295906, -0.23210827, -0.11435545,
        -0.39570294, -0.34034274,  0.08491891, -0.00243018,  0.32409415],
       [ 0.28956792,  0.6455377 ,  0.28963336, -0.00493871, -0.44261102,
        -0.53542088, -0.18523631,  0.47829689,  0.0411324 ,  0.1874174 ,
         0.42292073, -0.36328156, -0.1570596 , -0.23801606, -0.36803248,
        -0.15363543, -0.09741371,  0.60318768,  0.2947905 , -0.05998706,
        -0.11174835,  0.19493713,  0.42810767,  0.1572119 , -0.05909069,
        -0.21803694,  0.13179591, -0.24389973, -0.12475218,  0.03903671,
        -0.3136397 , -0.33920583,  0.0564947 , -0.28340583, -0.40138855,
         0.41286151, -0.02299765, -0.15389607, -0.32725482, -0.36204329,
        -0.24209945, -0.63771505,  0.04305627,  0.13866181,  0.3861812 ]]),
```

```
array([[-2.09553329e-03, -2.83383686e-01, -1.89348098e-01, ...,
         4.10836329e-09,  1.78167763e-01, -2.12532525e-02],
       [-1.56179435e-07,  1.68390463e-01, -4.12112578e-02, ...,
        -7.84874525e-03,  3.62979784e-01,  1.52828725e-01],
       [-5.30306052e-13, -3.59982154e-01, -4.34660216e-01, ...,
        -8.13580897e-03,  2.33873914e-01,  1.51226849e-01],
       ...,
       [-1.00980314e-02,  2.62556370e-01,  1.98985453e-02, ...,
        -4.36060544e-09,  3.31775193e-01, -1.40769837e-01],
       [-7.81491532e-03, -1.47737230e-01, -4.39706741e-01, ...,
        -1.85035924e-03, -1.51867423e-01,  2.72673594e-01],
       [-1.27720020e-03,  1.45879762e-01,  1.30810459e-01, ...,
        -5.32661953e-05,  3.22773706e-01,  2.96754380e-01]]),
array([[ 6.67087614e-03,  2.57824102e-04, -2.10369387e-04, ...,
         1.30100327e-05,  3.10330738e-10, -3.63647439e-03],
       [-2.22895452e-02,  3.96579367e-02, -2.10459127e-01, ...,
         2.29254822e-12,  1.34435617e-01,  1.47489465e-01],
       [-7.98080545e-02,  6.39805649e-02, -2.07827714e-01, ...,
         6.69372477e-02, -2.34963646e-01,  4.64588216e-01],
       ...,
       [-5.49662702e-04,  9.97029747e-04,  2.77062508e-04, ...,
         3.03950320e-13, -3.10835098e-10, -6.57735393e-05],
       [-1.54747232e-01, -6.46594534e-02,  7.36793926e-02, ...,
        -1.13950265e-01, -9.87774397e-02,  3.27221960e-01],
       [ 1.27878535e-02, -8.27450917e-02, -7.72087166e-02, ...,
        -2.36712608e-01,  2.10399132e-01,  5.84909485e-02]]),
```

Figure 3.7 - neural network coefficients

```
array([[-2.25687092e-01],
       [-4.25735261e-02],
       [-3.20274899e-02],
       [-2.52926024e-01],
       [ 4.52188309e-01],
       [ 3.49553936e-01],
       [-8.57636211e-03],
       [-3.17339902e-01],
       [-1.04082447e-01],
       [-2.69530392e-02],
       [-2.28642096e-01],
       [-2.27235498e-01],
       [-2.59030798e-01],
       [-2.94106680e-01],
       [ 4.95360488e-01],
       [ 2.99785542e-01],
       [-3.01470028e-01],
       [ 4.66513231e-01],
       [-2.97192536e-01],
       [-1.10176387e-01],
       [ 2.18239491e-01],
       [ 3.80657163e-01],
       [-1.49110418e-01],
       [-3.89642835e-02],
       [ 4.35854289e-01],
       [ 1.95112094e-01],
       [ 4.82352862e-01],
       [-9.72168594e-02],
       [ 3.83305037e-01],
       [-1.74542005e-02],
       [-2.59554935e-01],
       [-5.32480436e-02],
       [-3.13696354e-01],
       [-3.34995432e-01],
       [-3.39534231e-01],
       [ 4.27805082e-01],
       [ 4.69683121e-01],
       [ 6.01660508e-05],
       [ 4.53446623e-02],
       [-1.43142808e-01],
       [-1.58161829e-01],
       [ 1.62127858e-01],
       [-2.67300605e-01],
       [-1.48890230e-01],
       [ 4.75176427e-01]])]
```

Figure 3.8 - neural network coefficients

Figure 3.4 shows the graph of the neural network, the parameters of the model were selected manually, as you can see, the network turned out quite good, and the estimated values are very close to real values. There is also a method of finding the parameters of the model on the lattice. This method assumes that the selection of hyperparameters is set manually, then a complete search is performed. A popular implementation of this method is Sclearn Grid Search. In this work, the method of parameter selection was performed, three variations of the neural network were tested and the best version of the network was selected (Fig. 3.9). The results of the calculation of the criterion of determination, the mean absolute error, the root mean square error are presented in Figure 3.10.

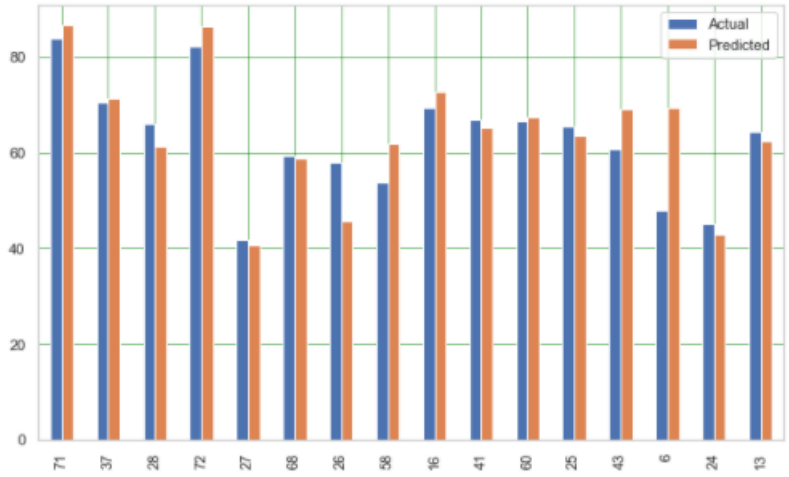


Figure 3.9 - neural network

```
Mean Absolute Error: 4.766978809140717
Mean Squared Error: 51.81089511376816
Root Mean Squared Error: 7.197978543575145
Test R^2 Score : 0.601
Training R^2 Score : 0.889
```

Figure 3.10 - the results of the calculation

As can be seen from Figure 3.9, the predicted values have quite high deviations from the actual values. The obtained coefficient of determination is low, the mean absolute error (4.766978809140717) and the root mean square error (51.81089511376816). Therefore, the first neural network model is better for further use and prediction.

3.3 Regression analysis

Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables [25].

Regression analysis is mainly used for two conceptually different purposes. First, regression analysis is widely used for prediction and prediction, where its use significantly overlaps with the field of machine learning [24]. Second, in some situations, regression analysis can be used to conclude causal relationships between independent and dependent variables [23].

Regression analysis was performed to analyze and predict the impact of cybersecurity factors on the fight against financial fraud.

OLS Regression Results

Dep. Variable:	DDL	R-squared (uncentered):	0.989
Model:	OLS	Adj. R-squared (uncentered):	0.988
Method:	Least Squares	F-statistic:	917.1
Date:	Thu, 20 Jan 2022	Prob (F-statistic):	1.94e-65
Time:	17:50:40	Log-Likelihood:	-254.77
No. Observations:	76	AIC:	523.5
Df Residuals:	69	BIC:	539.9
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
PSI	-4.2121	2.741	-1.537	0.129	-9.680	1.255
GEI	3.3332	2.472	1.349	0.182	-1.598	8.264
EDB	0.5366	0.069	7.748	0.000	0.398	0.675
CI	0.0723	0.071	1.016	0.313	-0.070	0.214
CPI	0.4296	0.103	4.177	0.000	0.224	0.635
GTI	-0.5404	0.650	-0.831	0.409	-1.838	0.757
FCI	0.0047	0.003	1.376	0.173	-0.002	0.012

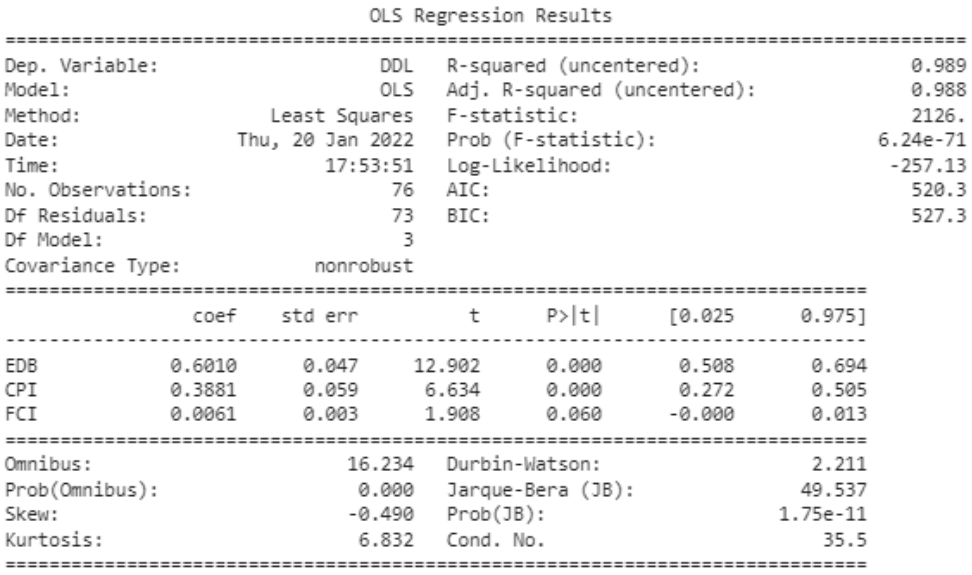
Omnibus:	8.057	Durbin-Watson:	2.162
Prob(Omnibus):	0.018	Jarque-Bera (JB):	13.461
Skew:	-0.286	Prob(JB):	0.00119
Kurtosis:	4.981	Cond. No.	1.48e+03

Figure 3.11 - Results of regression analysis

The regression analysis showed that most of the parameters have a value of P that exceeds 0.05. Accordingly, these parameters do not affect the fight against financial fraud. The next stage is the selection of parameters whose P value is below 0.05.

Selected parameters for regression are: "EDB", "CPI", "FCI".

FCI-this parameter was left because the index of financial secrecy is still important for the system of combating financial crime. The deviation of 0.06 from 0.05 is insignificant, based on this, it was decided to leave this indicator and conduct a regression analysis based on three factors.



OLS Regression Results

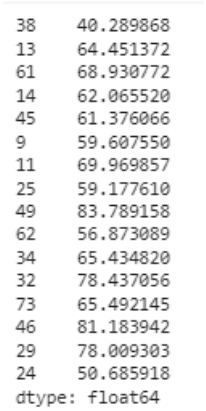
Dep. Variable:	DDL	R-squared (uncentered):	0.989
Model:	OLS	Adj. R-squared (uncentered):	0.988
Method:	Least Squares	F-statistic:	2126.
Date:	Thu, 20 Jan 2022	Prob (F-statistic):	6.24e-71
Time:	17:53:51	Log-Likelihood:	-257.13
No. Observations:	76	AIC:	520.3
Df Residuals:	73	BIC:	527.3
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
EDB	0.6010	0.047	12.902	0.000	0.508	0.694
CPI	0.3881	0.059	6.634	0.000	0.272	0.505
FCI	0.0061	0.003	1.908	0.060	-0.000	0.013

Omnibus:	16.234	Durbin-Watson:	2.211
Prob(Omnibus):	0.000	Jarque-Bera (JB):	49.537
Skew:	-0.490	Prob(JB):	1.75e-11
Kurtosis:	6.832	Cond. No.	35.5

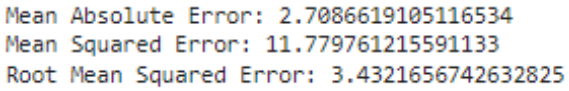
Figure 3.12 - Results of regression analysis

As a result of regression analysis, we have predicted values for the test and training data set (Fig. 3.13, Fig. 3.15).



```
38    40.289868
13    64.451372
61    68.930772
14    62.065520
45    61.376066
9     59.607550
11    69.969857
25    59.177610
49    83.789158
62    56.873089
34    65.434820
32    78.437056
73    65.492145
46    81.183942
29    78.009303
24    50.685918
dtype: float64
```

Figure 3.13 - predicted values for the test data set



```
Mean Absolute Error: 2.7086619105116534
Mean Squared Error: 11.779761215591133
Root Mean Squared Error: 3.4321656742632825
```

Figure 3.14 - calculation results

The average absolute error for the test data set is 2.7086619105116534, which is acceptable. The root mean square error of the forecast is 11.779761215591133, the mean square error is 3.432165674263282.

71	83.116014
7	48.714471
5	41.986723
24	50.685918
38	40.289868
9	59.607550
14	62.065520
37	71.341739
53	71.067008
64	83.077241
16	69.491275
62	56.873089
67	63.852640
55	63.432625
68	52.700596
30	54.270308
1	78.597861
25	59.177610
28	61.899997
3	61.929266
57	57.911556
47	86.897676
46	81.183942
58	63.895372
39	71.660891
0	79.543090
43	68.212335
74	56.676619
59	90.679383
6	62.943810
63	70.461803
34	65.434820
18	59.179764

Figure 3.15 - predicted values for the training data set

```
Mean Absolute Error: 4.424836546812158
Mean Squared Error: 43.873275551329534
Root Mean Squared Error: 6.623690478225076
```

Figure 3.16 - calculation results

The average absolute error for the training data set is 4.424836546812158. The root mean square error of the forecast is 43.873275551329534, the mean square error is 6.623690478225076.

The regression analysis helped to identify factors that, together with the index of the level of digital transformation, are important in the process of combating financial fraud in countries. These factors are the index of ease of doing business, the index of consumer prices, the index of financial secrecy.

CONCLUSION

Today, the problem of the impact of technology development and digitalization of the economy on the growth of the number of cyber frauds in the field of finance around the world is facing humanity. The pace of technology development and the quality of cyber systems is increasing, and, of course, the number of frauds is growing in various fields, especially in the field of finance. At present, financial systems do not have sufficient cybersecurity and are vulnerable to information technology. Information, money in different currencies, securities can be stolen by hackers from around the world. In order to improve the quality of the financial sector, as well as to reduce the number of cybercrimes and combat financial fraud, an analysis of the potential blood energy of the cybersecurity system and the fight against financial fraud was conducted. In the course of the work, statistical and visual analysis described the factors that may affect the level of security of the financial sector. Canonical analysis identified the DDL factor, which characterizes the level of digital transformation in cybersecurity. The process of combating financial fraud in the country depends on the cybersecurity of financial systems, namely the index of the level of digital transformation. This conclusion was made on the basis of calculations made by canonical analysis (Fig. 2.19). Two methods of analysis were used in the work - neural network and regression analysis. This was done in order to compare the two types of analysis and identify which one best describes the potential process of cybersecurity convergence and counter-fanatic fraud. Based on these types of analysis, it was found that important indicators in the field of cybersecurity and combating financial fraud are the factors of digital transformation and three indicators that characterize the index of ease of doing business, consumer price index, index of financial secrecy, respectively.

Thus, the process of convergence of the cybersecurity system and combating financial fraud is a topical issue today and requires detailed study and development. Through the implementation of technology in finance, it is possible to improve the

security of financial institutions and reduce the impact of hackers on financial systems, and. as a result, increase financial security in different countries.

REFERENCES

1. Descriptive data analysis URL: https://www.statmethods.ru/statistics-metody/opisatelnyj-analiz-dannykh/ accsess: 23.01.2022
2. Godin, AM Statistics: a textbook / O. M. Godin. - Moscow: Dashkov and K, 2016. - 451 p.
3. Dudin, MN Statistics: textbook and workshop for the academic bachelor / MN Dudin, NV Lyasnikov, ML Lezina. - Moscow: Yurayt Publishing House, 2019. - 374 p.
4. Cybersecurity URL: https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2016_10/20171004_1610-cybersecurity-curriculum-r.pdf accsess: 27.01.2021
5. CYBER SECURITY IN THE MODERN WORLD URL: https://prepod24.ru/readyworks/130832 / accsess: 27.01.2022
6. Cybersecurity of financial services URL: https://www.fortinet.com/en/solutions/industries/financial-services accsess: 30.01.2021
7. Cybersecurity in the financial sphere URL: https://risk-practice.ru/magazine/112/eau_112_659/ accsess: 30.01.2022
8. Python URL: https://ru.wikipedia.org/wiki/Python accsess: 30.01.2022
9. Pandas URL: https://ru.wikipedia.org/wiki/Pandas accsess: 30.01.2022
10. NumPy URL: https://ru.wikipedia.org/wiki/NumPy accsess: 30.01.2022
11. Global Cybersecurity Index URL https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cybersecurity-index.aspx accsess: 30.01.2022
12. Networked Readiness Index URL https://en.wikipedia.org/wiki/Networked_Readiness_Index#:~:text=The%20Netwo

rked%20Readiness%20Index%20is,by%20information%20and%20communications%20technology. accsess: 30.01.2022

13. National Cyber Security Index URL: https://eucyberdirect.eu/good-cyber-story/national-cyber-security-index accsess: 30.01.2022

14. The level of digital transformation URL: https://www.hpe.com/ru/ru/what-is/digital-transformation.html accsess: 30.01.2022

15. Political Stability Index URL: https://www.igi-global.com/dictionary/political-stability-index/88962 accsess: 30.01.2022

16. Crime index URL: https://help.neighborhoodscout.com/support/solutions/articles/25000001997-what-is-the-crime-index- accsess: 30.01.2022

17. Global Terrorism Index URL: https://www.visionofhumanity.org/maps/global-terrorism-index/#/ accsess: 30.01.2022

18. Principal component analysis URL: https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202 accsess: 30.01.2022

19. NumPy URL: https://ru.wikipedia.org/wiki/NumPy accsess: 30.01.2022

20. Matplotlib: Scientific graphics in Python URL: https://pythonworld.ru/novosti-mira-python/scientific-graphics-in-python.html accsess: 30.01.2022

21. Matplotlib: Scientific graphics in Python URL: https://pythonworld.ru/novosti-mira-python/scientific-graphics-in-python.html accsess: 30.01.2022

22. Data analysis and visualization using Python URL: https://conf.ztu.edu.ua/wp-content/uploads/2017/11/54.pdf accsess: 02.02.2022

23. Histograms. URL: http://sixsigmaonline.ru/baza-znanij/gistogrammy-chto-ehto-kak-postroit-kak-predstavit-dannye-kak-provesti-analiz accsess: 02.02.2022

24. Data visualization URL: https://www.oracle.com/ru/business-analytics/what-is-data-visualization/ accsess: 02.02.2022

25. Canonical correlation analysis URL: https://russianblogs.com/article/92701318727/ accsess: 02.02.2022

26. Canonical correlation analysis URL: http://statsoft.ru/products/STATISTICA_Advanced/canonical-correlation-analysis.php accsess: 02.02.2022

27. Yarovenko G. Canonical analysis of the relationship between information security and socio-economic and political development of the country. Scientific Bulletin of Uzhhorod National University. Series: International Economic Relations and the World Economy. 2020. № 31. S. 160–167

28. Yarovenko G. Kolotilina O., Svitlichna A. (2021). Assessing the level of convergence of the cybersecurity system and combating money laundering. Bulletin of VN Karazin Kharkiv National University. Series: International Relations. Economy. Local lore. Tourism, (14), 119-130. URL: https://doi.org/10.26565/2310-9513-2021-14-12 accsess: 04.02.2022

APPENDIX

Appendix A

```
import pandas as pd
import numpy as np
from sklearn import preprocessing
import matplotlib.pyplot as plt
plt.rc("font", size=14)
import seaborn as sns
sns.set(style="white")
sns.set(style="whitegrid", color_codes=True)
from pandas import read_csv, DataFrame
import statsmodels.api as sm
from statsmodels.iolib.table import SimpleTable
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error
```

Figure A-1 - Import libraries

```
df = pd.read_csv('data_conv.csv', delimiter=";", error_bad_lines=False)
```

Figure A-2 - import input

```
from sklearn.decomposition import PCA
pca_test = PCA(n_components=7)
pca_test.fit(X_scaled) #pca_test.fit(trainX_scaled)
sns.set(style='whitegrid')
plt.plot(np.cumsum(pca_test.explained_variance_ratio_))
plt.xlabel('number of components')
plt.ylabel('cumulative explained variance')
plt.axvline(linewidth=4, color='r', linestyle = '--', x=4, ymin=0, ymax=1)
display(plt.show())
evr = pca_test.explained_variance_ratio_
cvr = np.cumsum(pca_test.explained_variance_ratio_)
pca_df = pd.DataFrame()
pca_df['Cumulative Variance Ratio'] = cvr
pca_df['Explained Variance Ratio'] = evr
display(pca_df.head(4))
```

Figure A-3 - construction of the principal components method.

Appendix B

```
df_temp = df_temp.head(30)
df_temp.plot(kind='bar',figsize=(10,6))
plt.grid(which='major', linestyle='-', linewidth='0.5', color='green')
plt.grid(which='minor', linestyle=':', linewidth='0.5', color='black')
plt.show()
```

Figure B-1 - plotting the graph in Figure 3.4.

```
plt.plot(mlp_reg.loss_curve_)
plt.title("Loss Curve", fontsize=14)
plt.xlabel('Iterations')
plt.ylabel('Cost')
plt.show()
```

Figure B-2 - plotting the graph in Figure 3.6.

```
print([coef.shape for coef in mlp_reg.coefs_])
mlp_reg.coefs_
```

Figure B-3 - construction of neural network coefficients in Figure 3.7.

```
{'cv': 5,
 'error_score': nan,
 'estimator__activation': 'relu',
 'estimator__alpha': 0.0001,
 'estimator__batch_size': 'auto',
 'estimator__beta_1': 0.9,
 'estimator__beta_2': 0.999,
 'estimator__early_stopping': False,
 'estimator__epsilon': 1e-08,
 'estimator__hidden_layer_sizes': (45, 45, 45),
 'estimator__learning_rate': 'constant',
 'estimator__learning_rate_init': 0.001,
 'estimator__max_fun': 15000,
 'estimator__max_iter': 500,
 'estimator__momentum': 0.9,
 'estimator__n_iter_no_change': 10,
 'estimator__nesterovs_momentum': True,
 'estimator__power_t': 0.5,
 'estimator__random_state': None,
 'estimator__shuffle': True,
 'estimator__solver': 'adam',
 'estimator__tol': 0.0001,
 'estimator__validation_fraction': 0.1,
 'estimator__verbose': False,
 'estimator__warm_start': False,
 'estimator': MLPRegressor(activation='relu', alpha=0.0001, batch_size='auto', beta_1=0.9,
              beta_2=0.999, early_stopping=False, epsilon=1e-08,
              hidden_layer_sizes=(45, 45, 45), learning_rate='constant',
              learning_rate_init=0.001, max_fun=15000, max_iter=500,
              momentum=0.9, n_iter_no_change=10, nesterovs_momentum=True,
              power_t=0.5, random_state=None, shuffle=True, solver='adam',
              tol=0.0001, validation_fraction=0.1, verbose=False,
              warm_start=False),
 'iid': 'deprecated',
 'n_jobs': -1,
 'param_grid': {'hidden_layer_sizes': [(40, 40, 40),
   (35, 35, 35),
   (30, 30, 30)],
  'max_iter': [100, 500],
  'activation': ['tanh', 'relu'],
  'solver': ['sgd', 'adam'],
  'alpha': [0.0001, 0.05],
  'learning_rate': ['constant', 'adaptive']},
 'pre_dispatch': '2*n_jobs',
 'refit': True,
 'return_train_score': False,
 'scoring': None,
 'verbose': 0}
```

Figure B-4 - characteristics of neural network parameters.

Appendix C

```
print('Mean Absolute Error:', metrics.mean_absolute_error(testY, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(testY, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(testY, y_pred)))
print('Test R^2 Score : %.3f'%mlp_reg.score(testX, testY)) ## Score method also evaluates accuracy for classification models.
print('Training R^2 Score : %.3f'%mlp_reg.score(trainX, trainY))
```

```
param_grid = {
    'hidden_layer_sizes': [(40,40,40), (35,35,35), (30,30,30)],
    'max_iter': [100, 500],
    'activation': ['tanh', 'relu'],
    'solver': ['sgd', 'adam'],
    'alpha': [0.0001, 0.05],
    'learning_rate': ['constant','adaptive'],
}
```

```
grid = GridSearchCV(mlp_reg, param_grid, n_jobs= -1, cv=5)
grid.fit(trainX, trainY)

print(grid.best_params_)
```

```
print('Mean Absolute Error:', metrics.mean_absolute_error(testY, grid_predictions))
print('Mean Squared Error:', metrics.mean_squared_error(testY, grid_predictions))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(testY, grid_predictions)))
print('Test R^2 Score : %.3f'%grid.score(testX, testY)) ## Score method also evaluates accuracy for classification models.
print('Training R^2 Score : %.3f'%grid.score(trainX, trainY))
```

```
model = sm.OLS(y, X).fit()
predictions = model.predict(X)

print_model = model.summary()
print(print_model)
```

```
X_new = df [['PSI', 'GEI', 'EDB', 'CI', 'CPI','FCI']]
y = df['DDL']
```

```
model = sm.OLS(y, X_new).fit()
predictions = model.predict(X_new)

print_model = model.summary()
print(print_model)
```

```
X_new2 = df [[ 'EDB', 'CPI','FCI']]
y = df['DDL']
```

```
model2 = sm.OLS(y, X_new2).fit()
predictions = model2.predict(X_new2)

print_model2 = model2.summary()
print(print_model2)
```

```
print('Mean Absolute Error:', metrics.mean_absolute_error(testY2, y_pred2))
print('Mean Squared Error:', metrics.mean_squared_error(testY2, y_pred2))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(testY2, y_pred2)))
#print('Test R^2 Score : %.3f'%model2.score(testX2, testY2)) ## Score method also evaluates accuracy for classification models.
#print('Training R^2 Score : %.3f'%model2.score(trainX2, trainY2))
```

Figure C-1 – The code for model calculations.

```
trainX2, testX2, trainY2, testY2 = train_test_split(X_new2, y, test_size = 0.2)
y_pred3 = model2.predict(trainX2)
print(y_pred3)
```

```
print('Mean Absolute Error:', metrics.mean_absolute_error(trainY2, y_pred3))
print('Mean Squared Error:', metrics.mean_squared_error(trainY2, y_pred3))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(trainY2, y_pred3)))
#print('Test R^2 Score : %.3f'%model2.score(trainX2, trainY2)) ## Score method also evaluates accuracy for classification models.
#print('Training R^2 Score : %.3f'%model2.score(trainX2, trainY2))
```

Figure C-2 – The code for model calculations.