AR&P sciendo Financial Markets, Institutions and Risks, Volume 7, Issue 4, 2023
ISSN (online) – 2521-1242; ISSN (print) – 2521-1250
Sumy State University

The modeling of the probable behaviour of insider cyber fraudsters in banks

http://doi.org/10.61093/fmir.7(4).155-167.2023

Hanna Yarovenko, https://orcid.org/0000-0002-8760-6835

D.Sc., Visiting Prfessor of the Computer Science and Engineering Department, University Carlos III of Madrid, Spain

Aleksandra Kuzior, https://orcid.org/0000-0001-9764-5320

Professor at the Silesian University of Technology, Faculty of Organization and Management, Department of Applied Social Sciences, Poland; Oleg Balatskyi Department of Management, Sumy State University, Ukraine

Alona Raputa, https://orcid.org/0000-0002-8981-7986

Analyst of operating and application software, "Ascania Trading House" LLC, Ukraine

Corresponding author: h.yarovenko@biem.sumdu.edu.ua

Abstract. *Insider cyber fraud in the banking sector is a serious and complex issue for financial institutions. This form of cyber fraud is particularly insidious due to insiders' inherent access and knowledge, necessitating banks to implement comprehensive strategies for detecting, preventing, and responding to these internal threats. The aim of this study is to develop a scientific and methodological approach to model the probable behaviour of insider cyber fraudsters in banks based on a complex combination of principal component analysis, k-means clustering, and associative analysis. During the analysis of current challenges in the financial sector regarding the evolution of cyber fraud and its implications, the systematization of existing theoretical approaches concerning the examination of cyber fraud in banks was performed. Its result revealed a positive trend in the dynamics of the number of published materials in conferences and articles using keywords "cyber" and "frauds" in the Scopus database from 2000 to 2023. Additionally, utilizing the VOSviewer software facilitated the systematization of keyword combinations used in scholarly publications on the chosen topic, forming clusters to visualize and organize vectors of scientific research. Analytical data from Google Trends on critical issues related to cyber fraud were chosen as input data. Twenty variables were formed, which are the results of search queries, characterizing cyberattacks and decreased trust in financial institutions. The principal components method was used to reduce the dimensionality of the input data array, making it possible to select the nine most significant for the study. Conducting a cluster analysis using the k-means method made it possible to form 3 main groups of search queries, which included 12 of the selected variables. The results of the performed procedures contributed to the implementation of associative analysis for three sets of variables. It has been found that what intrigues potential insider cybercriminals in banks the most is the personal financial information of the client, access to the client's profile in online banking and gaining access to his phone data. The obtained results can be utilized by commercial banks for identifying potential insider cyber fraudsters and ensuring a higher level of client protection against the actions of insider cyber fraudsters, by bank clients for analysing and mitigating potential threats from insider cyber fraudsters, and by law enforcement agencies for prompt responses to potential threats posed by insider cyber fraudsters in banks.*

Keywords: Bank, Cyber Fraud, Insider, Cluster Analysis, Principal Component Analysis, Associative Analysis.

JEL Classification: C38, G21.

Received: 24.10.2023 **Accepted:** 14.12.2023 **Published:** 31.12.2023

Funding: There is no funding for this research.

Publisher: Academic Research and Publishing UG (i. G.) (Germany)

Cite as: Yarovenko, H., Kuzior, A. & Raputa, A. (2023). The modeling of the probable behaviour of insider cyber fraudsters in banks. *Financial Markets, Institutions and Risks*, *7*(4), 155-167. http://doi.org/10.61093/fmir.7(4).155-167.2023

ISSN (online) 2521-1242; ISSN (print) – 2521-1250

Introduction

The banking sector is one of the sections of the economy that generates and accumulates large amounts of information. In addition, it was the banking system that became one of the most popular industries where the concept of "big data" began to be implemented. In recent years, the banking industry has continued to change under the influence of a number of innovations arising from technological progress, psychological aspects of consumers of financial services and regulatory requirements.

With the development of innovative approaches to the transformation of the financial sector, the problem of cyber fraud is becoming increasingly relevant. Thus, between October 2021 and September 2022, malware was the most common type of cyber attack worldwide, affecting 40% of financial and insurance organizations. The second and third most affected organizations in the financial sector are cyber fraud via websites and mobile applications (23%) and intrasystem fraud (20%) (Statista, 2023). Such a trend is rather alarming, as it has a positive character. Thus, in 2022, 1,829 incidents of cyber fraud in the financial industry were reported worldwide, compared to 2,527 in the previous year. The total number of cyber fraud cases in the financial sector during 2013-2022 increased by 46.8% (from 856 cases in 2013 to 1,829 cases in 2022). The lowest level of cyber fraud in the financial sector was observed in 2017 (598 cases) (Statista, 2023). Accordingly, the cases of cyber frauds that were accompanied by a data leak during the presented period changed proportionally to the previous indicator. However, it is worth noting that in percentage terms, the largest number of cyber fraud cases, which were accompanied by data leakage in the financial sphere, occurred in 2020 (64.8%). During 2021-2022, the number of such cyber frauds does not exceed 28%, which indicates an improvement in the level of cyber protection in the financial sector (Statista, 2023).

It should be noted that the initiators of cyber fraud can be not only external sources, but also internal ones caused by insiders - employees of financial institutions. Insider cyber fraud is a serious and complex problem. Unlike external threats, insider cyber fraud involves individuals within an organization using their privileged access to commit fraudulent activities. This type of threat can compromise the integrity of banking systems and undermine the trust of customers and stakeholders. This form of cyber fraud is particularly insidious because of the inherent access and knowledge possessed by insiders, making it imperative for banks to implement comprehensive strategies to detect, prevent and respond to these insider threats. With the development and implementation of digital innovations, the problem of the spread of insider cyber fraud is constantly updated and requires more and more new solutions.

1. Literature Review

To study the current state of issues related to insider cyber fraud in banks, an analysis of existing scientific works in this field was carried out, which allows to get an idea of the history of the origin of cyber fraud, their types, the most popular technological vulnerabilities, the regulatory landscape, the impact of cyber threats on interested parties facilities, cyber security measures, etc. As of 2023, a search using the keywords "cyber" and "frauds" in the authoritative international scientific database Scopus yielded 1,262 documents, including 547 conference materials, 451 articles, 119 book chapters, 58 abstracts of conference presentations and other scientific works (Scopus, 2023). In order to understand the relevance of scientific works that are related to the topic of "cyber fraud" and form a generalized picture of the theoretical background, a map of concepts on the topic of "cyber fraud" was constructed using the VOSviewer software product (Figure 1).

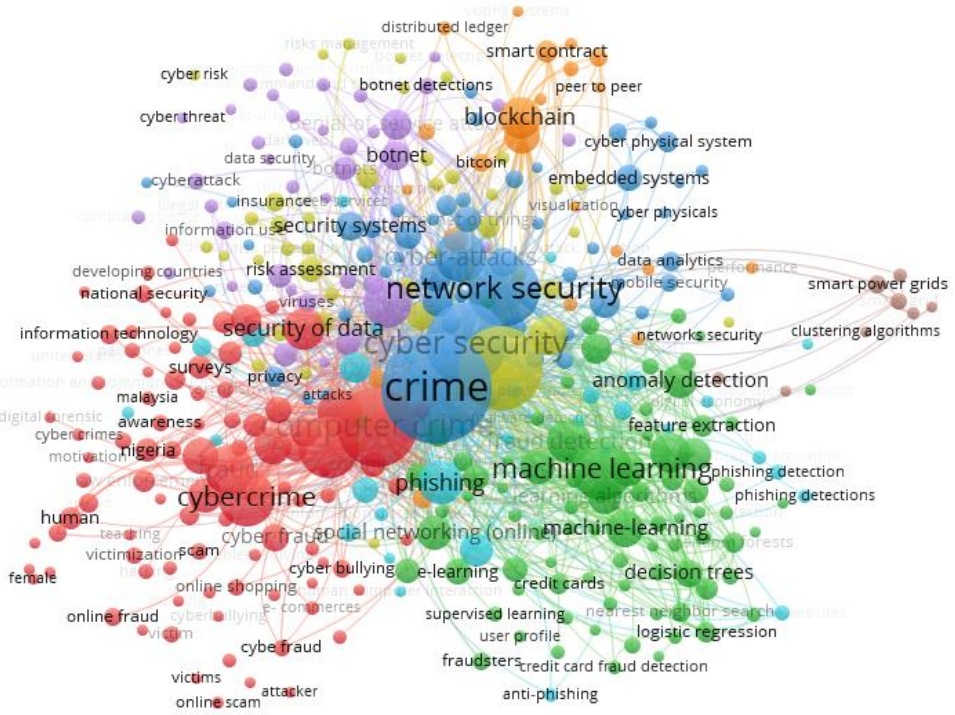

Figure 1. Distribution of scientific publications on the subject of "cyber fraud" by clusters in the international Scopus database

Source: compiled by the authors based on (Scopus, 2023).

The map of publications (Figure 1) makes it possible to distinguish eight clusters. The red cluster includes concepts more closely related to the criminal component of cyber fraud, since the most common terms are "computerized crime", "cybercrime", "law and legalization", "criminal activity", "digital forensics", etc. The connection with cyber fraud in the financial sphere in this cluster is represented by the terms "electronic commerce", "online shopping", "financial crime", "money laundering", "online banking". The second largest cluster (green) contains concepts such as "machine learning", "learning algorithms", "neural networks", "learning systems", "fraud detection", "social networking", "behavioral aspects". This shows that when researching the topic of cyber fraud, machine learning technologies are actively used and it is necessary to take into account the behavioral aspects of subjects who can commit cyber fraud. From the sphere of the financial sector, such concepts as "credit cards", "phishing", "financial transactions" appear in this cluster.

The blue cluster contains the concepts most closely related to cyber security: "cyber security", "control systems", "network security", "security systems", "cloud computing", "security threats". From the sphere of the financial sector, the following concepts appear in the blue cluster: "financial information", "electronic money". The rest of the selected clusters are significantly smaller than the three previous ones considered. However, it is worth noting that in the yellow cluster, concepts from the financial sphere that are directly related to banking activities are most often found – "banking system", "credit card theft", "electronic banking", "mobile banking", "fintech", "financial inclusion". This confirms the close connection between cyber fraud and banking. Purple, blue, orange, brown clusters are dedicated to examples of specific cyber fraud.

As banks adopt increasingly sophisticated technology, they become more vulnerable to cyber attacks (Kuzior et al., 2022a). Therefore, some scientists highlight the issue of vulnerabilities related to online banking platforms (Wahab et al., 2023), mobile applications (Siano et al., 2020), the use of blockchain technologies (Ugochukwu et al., 2022) and cloud services (Setyaji et al., 2020). Studying these technological vulnerabilities is imperative to developing robust cybersecurity strategies. In addition, behavioral aspects of financial market entities play an important role in the cyber vulnerability of financial systems (Al, 2019).

Scholars reveal a variety of cyber fraud techniques targeting banks, including phishing (Yoro et al., 2023), malware attacks (Shamsi et al., 2023), identity and credit card theft (Dewi et al., 2023; Kuzior & Kuzior, 2018), and apps - extortionists (Rawat et al., 2023). Financial loss, reputational damage and the erosion of trust are common outcomes of cyber fraud. Therefore, researchers often examine the cyber security measures

that banks use to protect against cyber fraud (Kuzior et al., 2022b). For example, some scholars analyze the effectiveness of encryption and multifactor authentication (Almaiah et al., 2023), artificial intelligence (Rithani et al., 2023), biometrics (Malik et al., 2023), and other technological solutions.

The systematization of existing theoretical approaches to the consideration of the subject of cyber fraud in the financial sphere in general and in banks in particular allows to create a basis for conducting further research and solving new scientific problems. Since the subject of insider cyber threats is little researched, the issues of this article are relevant for the scientific basis.

2. Research Methodology and Data

In the context of the purpose of this work, it is necessary to model the probable behavior of insiders-cyber fraudsters in the bank. Since everything related to the assessment of behavioral aspects of human activity is largely subjective, the main difficulty in conducting such studies is the selection of input parameters for this. It is not possible to predict one hundred percent how a person will behave in a particular situation, in particular, an insider-cyber fraudster of a bank, as his behavior is determined by a number of endogenous and exogenous quantitative and qualitative factors, the influence of which is very difficult to analyze. Taking into account the nature of the potential fraudulent actions of a cyber-fraudulent bank insider, it is proposed to use possible combinations of search queries in the Google search system during the last five years from 2018 to 2023 as an array of input variables that will allow us to evaluate his possible behavior. basis of the formation of the input array of data for the presented study, two lists of search queries were formed: a list of queries describing the characteristics of cyber attacks and a list of queries characterizing the level of decreased trust in financial institutions (Table 1).

Table 1. Input array of data - requests

Legend	Search queries for characteristics of cyber attacks	Legend	Search queries characterizing the level of decreased trust in financial institutions
var1	Cyber police number	var11	How to block a transaction
var2	Police number	var12	How to block a bank card
var3	What to do when you are hacked	var13	How to change password of bank card
var4	How to respond to a cyber attack	var14	How to reduce the credit limit
var5	How to protect your computer	var15	Management of online payments
var6	How to prevent hacking	var16	Which bank is the most secure online
var7	The most common cyber attacks	var17	The most reliable banks
var8	Detection of a cyber attack	var18	Bank call center number
var9	How to protect yourself from cyber attacks	var19	How to change the bank

Source: compiled by the authors based on the Google Trends search engine, 2023.

Modeling of the likely behavior of insider cyber fraudsters in banks will be carried out in three stages.

At the first stage, using the method of principal components, an array of the most relevant variables for further research will be formed, obtained from the list of 20 key queries presented in Table 1. The main goal of this method is to transform data of a large dimension into a representation of a smaller dimension, fixing as many deviations as possible in data The principal component method algorithm has the following sequence:

1. Data centering (subtracting the average value of each variable from each value of the corresponding indicator).

2. Calculation of the covariance matrix (the covariance matrix describes the relationships between all pairs of variables in the data).

3. Decomposition of the covariance matrix into vectors of eigenvalues representing the directions of maximum dispersion in the data, and the corresponding eigenvalues indicate the amount of dispersion along these directions.

4. The selection of the main components is accompanied by the ranking of the vectors of the eigenvalues of the components in descending order. The eigenvector with the highest eigenvalue is the first principal component, the second largest is the second principal component, and so on.

5. Projecting data onto principal components (initial data are projected onto selected principal components, creating a new set of variables (principal components) that are uncorrelated and capture the most important information in the data.

6. Evaluation of the factor loadings of the input indicators within the selected components.

At the second stage of modeling, it is necessary to carry out clustering using the k-means method. This clustering method was chosen for this study because of its popularity in grouping points in such a way as to minimize the sum of squared distances between data points and the centroid of the cluster to which they belong.

The k-means clustering method algorithm includes the following sequential steps:

1. Primary selection of the centers of previous k clusters (selection of k variables subject to determination of the maximum distance between them).

2. Primary redistribution of objects between clusters (the principle of redistribution is based on determining the minimum distance between objects).

3. Starting the iterative process, which continues until the optimal cluster structure is formed, and the total number of iterations is equal to the maximum number.

At the third stage of the study, the construction of potential portraits of insiders-cyber fraudsters in banks based on the selected variables by the method of principal components and clustering using the method of associative learning is envisaged. The construction of associative rules is the basis of affinity analysis, the essence of which is to identify the relationship between certain events that may have a fundamental condition (Kovalenko et al., 2019). The general algorithm of modeling using associative rules includes the following steps:

Formation of a set of events (transactions), which will form the basis of modeling.

A study of the structure of an associative rule that should include antecedent та consequent (X=>Y).

Definition of the main characteristics of the associative rule: support, confidence, interest lift, leverage, conviction and Zhang metric.

n association rule mining, support is a measure that indicates the frequency with which a particular set of items appears together in a dataset. It helps to identify the strength of the relationship between items in a transactional database (formula 1):

$$Support(X) = \frac{Transactions\ containing\ X}{Total\ transactions\ in\ the\ dataset}. \quad (1)$$

Confidence in the context of an associative rule, it is a measure of the accuracy of the rule and is equal to the ratio of the total number of transactions with the condition and the consequence to the number of transactions (formula 2):

$$Confidence(X \to Y) = \frac{Support(X \cup Y)}{Support(X)}, \quad (2)$$

where $Support(X \cup Y)$ is the support of the combined itemset $X \cup Y$, representing the transactions where both X and Y co-occur; $Support(X)$ is the support of the antecedent itemset X, representing the transactions where X occurs.

The higher the support and probability values, the higher the probability that a given transaction that contains the condition will also include the consequence. Interest lift is the ratio of the frequency of the condition and the consequence of the transaction to the frequency of the occurrence of the consequence (the larger the value, the more often the condition determines the occurrence of the consequence) (formula 3):

$$Lift(X \to Y) = \frac{Confidence(X \to Y)}{Support(Y)}, \quad (3)$$

where $Confidence(X \to Y)$ is the confidence of the association rule $X \to Y$; $Support(Y)$ is the support of the consequent itemset Y, representing the transactions where Y occurs. If the lift is equal to 1, then there

is no connection between the condition and the consequence. If the value is close to 0, then there is a strong inverse relationship.

Leverage, is equal to the difference of the observed frequency when the condition and the consequence are identified together, and the product of the frequency of detection of the condition and the consequence (formula 4):

$$Leverage(X \to Y) = Support(X \cup Y) - Support(X) * Support(Y), \quad (4)$$

where $Support(X \cup Y)$ is the support of the combined itemset $X \cup Y$, representing the transactions where both X and Y co-occur; $Support(X)$ of the antecedent itemset X, representing the transactions where X occurs; $Support(Y)$ of the antecedent itemset Y, representing the transactions where Y occurs. Conviction is a measure used in association rule mining to evaluate the degree of dependency between the antecedent and consequent of a rule. It focuses on the ratio of the expected frequency of incorrect predictions to the observed frequency. The formula (5) for conviction is given by:

$$Conviction(X \to Y) = \frac{1 - Support(Y)}{1 - Confidence(X \to Y)}, \quad (5)$$

where $Support(Y)$ is the support of the consequent itemset Y, representing the transactions where Y occurs; $Confidence(X \to Y)$ is the confidence of the association rule $X \to Y$.

Zhang's metric (6) allows to determine both association and dissociation. The value ranges from -1 to 1. A positive value indicates association and a negative value indicates dissociation.

$$Zhang(X \to Y) = \frac{Confidence(X \to Y) - Confidence(\grave{X} \to Y)}{Max[Confidence(X \to Y), Confidence(\grave{X} \to Y)]}. \quad (6)$$

4. Formulation of conclusions based on the obtained associative rules.

Thus, methodological support for modeling the probable behavior of insider-cyber fraudsters in banks will be implemented on the basis of a combination of three methods of statistical research: the method of principal components for the identification of relevant variables, the method of k-means clustering for the formation of research clusters, and the method of associative rules for building potential portraits of insiders - cyber fraudsters in banks. All necessary calculations in the work will be carried out using the Python 3 programming language.

3. Results and Discussions

According to the defined sequence of stages of modeling the probable behavior of the actions of insiders-cyber fraudsters in banks, it is first necessary to select the most relevant variables for further research using the method of principal components. Let's analyze the eigenvalues of the components obtained for 20 input variables (Table 2) and the graph of the stony scree (Figure 2). This will reveal the optimal number of components for further analysis.

Table 2. Eigenvalues, variance and cumulative variance of components

Component	Value	Dispersion	Cumulative dispersion
Component 1	3.322	1.900	0.166
Component 2	1.423	0.082	0.237
Component 3	1.341	0.144	0.304
Component 4	1.197	0.066	0.364
Component 5	1.131	0.033	0.421
Component 6	1.098	0.021	0.476
Component 7	1.078	0.022	0.530
Component 8	1.055	0.047	0.582
Component 9	1.008	0.045	0.633
Component 10	0.964	0.056	0.681
Component 11	0.907	0.053	0.726
Component 12	0.854	0.079	0.769
Component 13	0.775	0.017	0.808

Table 2 (cont.). Eigenvalues, variance and cumulative variance of components

Component	Value	Dispersion	Cumulative dispersion
Component 14	0.757	0.030	0.846
Component 15	0.727	0.052	0.882
Component 16	0.676	0.097	0.916
Component 17	0.579	0.101	0.945
Component 18	0.478	0.135	0.969
Component 19	0.343	0.056	0.986
Component 20	0.287	0.010	1.000

Source: compiled by authors.

The total number of obtained components corresponds to the total number of input variables. Considering the results of the eigenvalues of the obtained components presented in Table 2, the first nine components have an eigenvalue greater than 1. At the same time, the value of the cumulative variance for the data of the nine components is equal to 0.633, which means that more than 63% of the studied the phenomenon is explained by these components.

The stony scree graph (Figure 2) allows you to visualize the results of the first stage of the principal component method, as it shows the intrinsic values of each component. The dotted line on the graph indicates the place corresponding to the optimal number of components. In this case, it is 9.

In order to understand the degree of influence of each variable within each component, it is necessary to examine their factor loadings (Table 3). In essence, the factor loading is the correlation coefficient of the corresponding variable with the component to which it entered. In the context of the topic of this study, each component represents a certain portrait of a potential insider-cyber fraudster in the bank, and the largest values of the factor loadings of the variables indicate which features determine this portrait.

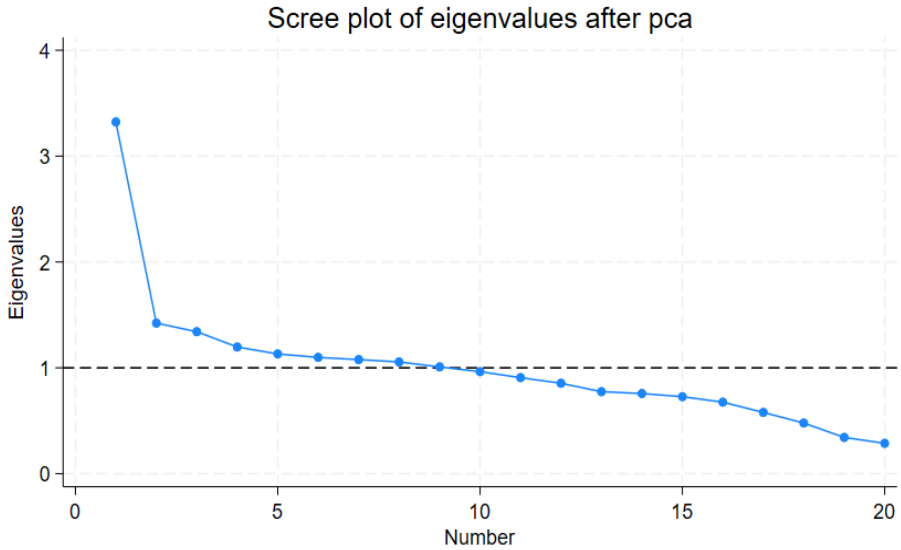

Figure 2. Graph of stony scree

Source: compiled by the authors.

Table 3. Factor loadings of indicators

Variable	C1*	C2	C3	C4	C5	C6	C7	C8	C9
var1	**0.451**	-0.019	-0.013	0.016	0.019	-0.098	-0.150	-0.091	-0.008
var2	0.271	-0.023	**0.361**	**0.378**	0.034	-0.130	-0.284	-0.122	0.076
var3	0.129	0.104	0.131	-0.168	-0.131	-0.135	**0.422**	0.030	**0.473**
var4	-0.046	0.195	-0.253	**0.593**	0.013	-0.017	0.082	0.048	-0.086
var5	**0.421**	0.095	-0.152	0.020	0.058	0.029	0.119	0.032	-0.033
var6	0.052	**-0.597**	0.002	0.101	0.056	0.125	0.137	0.283	0.039
var7	0.213	0.236	-0.173	-0.070	**-0.394**	0.166	-0.041	-0.006	-0.165
var8	0.150	-0.091	-0.134	0.112	0.091	0.206	0.164	0.052	**0.600**

Table 3 (cont.). Factor loadings of indicators

Variable	C1*	C2	C3	C4	C5	C6	C7	C8	C9
var9	-0.038	0.234	**0.357**	0.036	0.214	0.207	**0.369**	-0.118	0.129
var10	-0.025	0.166	**-0.309**	0.175	0.057	**-0.391**	-0.158	**0.307**	0.196
var11	**0.378**	0.117	0.107	0.015	-0.034	-0.113	-0.091	0.029	-0.073
var12	0.125	-0.085	0.044	0.037	0.173	-0.228	**0.575**	-0.094	**-0.492**
var13	-0.095	-0.171	-0.281	**0.359**	-0.017	0.239	0.144	**-0.463**	-0.057
var14	0.130	**-0.313**	-0.241	-0.094	**0.339**	0.120	-0.232	**-0.347**	0.122
var15	0.161	-0.207	-0.057	0.037	-0.034	**0.343**	0.054	**0.614**	-0.187
var16	0.065	0.063	**0.305**	**0.361**	**-0.397**	**0.363**	-0.030	-0.031	0.038
var17	0.033	0.221	0.206	-0.176	**0.417**	**0.434**	-0.235	0.050	-0.126
var18	-0.170	-0.107	**0.366**	**0.320**	**0.310**	-0.245	-0.083	0.169	0.005
var19	**0.459**	-0.050	0.015	-0.007	0.159	-0.097	0.055	-0.055	-0.034
var20	-0.012	**0.425**	-0.267	0.117	**0.402**	0.175	0.085	0.157	0.017

**C is an abbreviation of the word component.*

Source: compiled by the authors.

For a better perception of the obtained results, we highlight in this table only those values of factor loadings that absolutely exceed the value of 0.3 (Table 3). This will identify the most relevant variables within the selected components. As you can see, each of the presented components is determined by a different combination of input variables. This once again confirms the possibility of identifying different potential portraits of cyber-fraud insiders in the bank.

To implement cluster analysis, the Silhouette criteria were determined, the maximum values of which correspond to the optimal number of clusters that can be formed during k-means clustering. The input data for the cluster analysis served as the selected 9 components. The largest value of the Silhouette criterion (0.357) corresponds to the optimal number of clusters of 2. However, this number of clusters is not effective. Experiments with a different set of clusters showed that the most effective distribution for this study will be achieved if 3 clusters are selected, which corresponds to the Silhouette criterion value of 0.27. A visual representation of the formed clusters is shown in Figure 3.

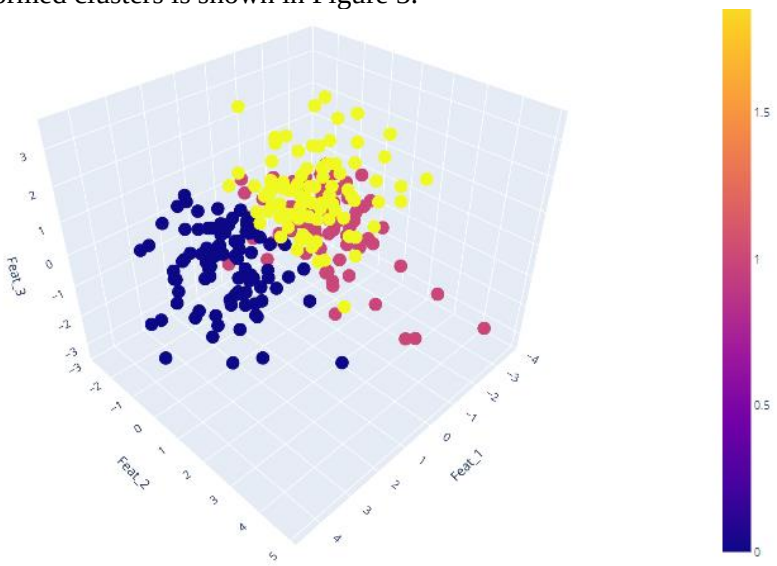

Figure 3. Results of k-means clustering

Source: compiled by the authors.

Thus, the clustering result made it possible to form three groups of variables taking into account the factor loadings of Table 3, which will be the most important for associative analysis. Their results are shown in Table 4.

Table 4. Selected groups of variables for associative analysis

Group	Variables
I	Cyber police number
	How to strengthen the protection of your computer
	How to block a transaction
	How to change the servicing bank (how to transfer salary payments from one bank to another)
II	How to prevent hacking of personal data (site, social networks)
	How to reduce the limit on a bank card
	Black list of users
III	Police number
	How to protect yourself from cyber attacks
	How to understand that the computer (phone) has been hacked
	Which bank is the most secure (on the Internet)
	Bank support number (or bank call center number)

Source: compiled by the authors.

The first group includes variables that are aimed at determining the security policy against potential cyber frauds, which, on the other hand, can be an opportunity to breach the bank's security by insiders-cyber fraudsters in particular. The second group combines variables that are directly related to the protection of personal data of customers, which can be the basis for obtaining the necessary information for cyber-fraudulent insiders. The third group includes variables that are also related to ensuring the security of users' personal data, identifying potential vulnerabilities of banks.

The third stage of modeling involves conducting an associative analysis using Python 3. The "apriori" algorithm was used to determine the associative rules. Its structure is presented in the form of formula (7):

$$frequent\_itemsets = apriori(one\_hot, min\_support=0.01, use\_colnames=True) \quad (7)$$

where one_hot – input data, in the format one-hot encoding;

min_support=0.01 – sets occurring in 1% of transactions;

use_colnames=True – uses variable values.

The process of creating association rules has the form (8):

$$rules = association\_rules(frequent\_itemsets, metric="confidence", min\_threshold=0,2) \quad (8)$$

where metric="confidence" – rule reliability measure;

min_threshold=0.2 – rules with a reliability of at least 20%.

As a result of the associative analysis, three models of associative rules with corresponding quality criteria were obtained. Considering the results of the associative analysis for the first group of variables, only 20 associative rules were obtained. Their probability ranges from 0.214 to 0.75. After ranking the obtained set of associative rules by the level of probability corresponding to a value in the range from 0.6 to 0.75, the following results were obtained (Table 5).

Table 5. Results of associative analysis for the first group of variables, the probability of associative rules of which is in the range of 0.6-0.75

Cause	Effect	Support	Probability	Lift	Leverage	Evidence	Metric Zhang
How to change the bank_40	Cyber police number_24	0.023	0.600	26.000	0.011	2.442	0.980
How to protect your computer_54	How to block a transaction_31	0.027	0.600	22.286	0.011	2.433	0.974

Source: compiled by the authors.

As you can see, the rules How to change the bank => Cyber police number and How to protect your computer => How to block a transaction are fulfilled with a probability of 60%. At the same time, the support value is 2.3% and 2.7%, respectively, which means that the presented rules are found in more than 2% of all

transactions. A relatively low level of support value in the context of detecting potential fraudulent actions of insider cyber fraudsters in banks is normal, since the moment of detecting fraudulent actions is quite complex and may depend on a significant set of factors. The high elevator value for both rules, 26 and 22.286 respectively, suggests that the presented effects are often determined precisely by the given causes, compared to situations where the causes are absent. The significance of the obtained associations, which is described by leverage, is the same and is 1.1%. The positive value of Zhang's metric for both associative rules is positive, 0.98 and 0.974, respectively, which confirms the presence of an association between causes and effects.

Thus, if we transform the obtained results of the associative analysis for the first group of variables into a potential insider-cyber-fraudster in the bank, we can conclude that the reason for the change of the servicing bank is cyber-fraud itself, since most likely after that there is a need to search for a cyber-police number. Thus, an insider-cyber-fraudulent bank can gain access to the personal financial information of the affected client. The second associative rule from the table. 5 gives an insider-cyber-fraudster in the bank an opportunity to understand potential vulnerabilities in the protection of the computer of the user-client of the bank for a blocked transaction.

Similarly, for the preliminary analysis of the results, we will select those associative rules that have the highest value of associative probability for the second group. Many associative rules have a probability value of 100%, however, taking into account the low number of previously obtained results for the search queries "How to reduce the credit limit" and "Black list of customers", all the consequences of the generated associative rules correspond to zero number of relevant queries, which does not allow correctly investigate the relationship between the condition and the consequence. In addition, the lift value for all pairs of associative rules approaches unity, which also confirms the absence of a relationship between the condition and the outcome.

The number of obtained associative rules for the third group is large, so it is necessary to analyze their quality. Similarly to the previous results for this group of variables, there is also a value of associative probabilities at the level of 100%, but not for all constructed associative rules it really confirms the presence of a cause-and-effect relationship between the condition and the effect. Therefore, there is a need to analyze low values of associative probabilities, on the basis of which it is possible to prove the presence of a qualitative connection between the condition and the consequence. Table 6 presents the associative rules for a set of variables for which the associative probabilities are in the range from 0.6 to 1, and the search queries are not zero.

Table 6. Results of associative analysis for the third group of variables, the probability of associative rules of which is in the range of 0.6-1

Cause	Effect	Support	Probability	Lift	Leverage	Evidence	Metric Zhang
How to protect yourself from cyber attacks_13	Police number_63	0,065	1,000	15,294	0,011	inf*	0,946
How to find that phone is hacked_28	Bank call center numbe_52	0,038	0,750	19,500	0,011	3,846	0,964
Which bank is the most secure online_0, How to protect yourself from cyber attacks_13	Police number_63	0,065	1,000	15,294	0,011	inf*	0,946
How to protect yourself from cyber attacks_13	Which bank is the most secure online_0, Police number_63	0,042	1,000	23,636	0,011	inf*	0,969
How to protect yourself from cyber attacks_0, How to find that phone is hacked_28	Bank call center numbe_52	0,038	0,750	19,500	0,011	3,846	0,964
How to find that phone is hacked_28	Bank call center numbe_52, How to protect yourself from cyber attacks_0	0,035	0,750	21,667	0,011	3,862	0,969
Which bank is the most secure online_0, How to find that phone is hacked_28	Bank call center numbe_52	0,038	0,750	19,500	0,011	3,846	0,964

Table 6 (cont.). Results of associative analysis for the third group of variables, the probability of associative rules of which is in the range of 0.6-1

Cause	Effect	Support	Probability	Lift	Leverage	Evidence	Metric Zhang
How to find that phone is hacked_28	Which bank is the most secure online_0, Bank call center numbe_52	0,035	0,750	21,667	0,011	3,862	0,969
Which bank is the most secure online_0, How to protect yourself from cyber attacks_0, How to find that phone is hacked_28	Bank call center numbe_52	0,038	0,750	19,500	0,011	3,846	0,964
Which bank is the most secure online_0, How to find that phone is hacked_28	Bank call center numbe_52, How to protect yourself from cyber attacks_0	0,035	0,750	21,667	0,011	3,862	0,969
How to protect yourself from cyber attacks_0, How to find that phone is hacked_28	Which bank is the most secure online_0, Bank call center numbe_52	0,035	0,750	21,667	0,011	3,862	0,969
How to find that phone is hacked_28	Which bank is the most secure online_0, Bank call center numbe_52, How to protect yourself from cyber attacks_0	0,031	0,750	24,375	0,011	3,877	0,974

**inf means informative.*

Source: compiled by the authors.

As you can see, the search query "Police number" is a consequence of the cause “How to protect yourself from cyber attacks” with a probability of 100%. Within other associative rules, where "Police number" also occurs as a consequence or part of a consequence together with another search query, the cause remains unchanged. The search query “Bank call center number” is present in seven obtained associative rules and all seven times as a result. With a probability of 75%, this search query appears as a result of another search query – “How to find that phone is hacked”. At the same time, it is worth noting that this cause-and-effect relationship is present both directly between this pair of search queries and in combination with other search queries.

The rest of the resulting associative rules contain queries that have zero frequency of occurrence, so there is no need to interpret them. At the same time, the value of support for the considered associative rules is from 3.1% to 6.5%. This means that the presented rules are found in from 3.1% to 6.5% of all transactions. This result is absolutely normal when it comes to the analysis of potential fraudulent schemes. The high lift value for both types of associative rules, from 15.294 to 24.375, confirms that the presented effects are often determined precisely by the considered causes, compared to situations where the causes are absent. The significance of the obtained associations, which is described by leverage, is the same and is 1.1%. The positive value of Zhang’s metric for both associative rules is positive, from 0.964 to 0.974, which confirms the presence of an association between causes and effects.

Therefore, based on the results of the third associative analysis, a potential insider-cyber-fraudster in the bank is once again convinced that the vulnerability of users to cyber-attacks is accompanied by a search for a police number to eliminate negative consequences, which once again confirms the effectiveness of fraudulent actions on the condition of gaining access to the personal data of bank customers . The second associative rule from the table. 6 gives the insider-cyber-fraudster in the bank an understanding that the majority of banking transactions by the modern user of banking services today take place with the help of the phone, since the need to find out whether the phone has been hacked by cyber-fraudsters is most likely accompanied by a call to the bank's call center. Therefore, an insider-cyber-swindler of a bank can, having gained physical access to a bank client's phone, carry out a number of fraudulent actions with his bank account.

Conclusion

The modeling of the probable behaviour of insider cyber fraudsters in banks is one of the important problems for the financial system and cyber security as a whole. Its solution becomes a necessary component of the strategy for ensuring cyber security in the banking sector. In the process of researching this problem, this paper systematizes the existing approaches to cyber fraud in banks, developed by modern experts and scientists. As a result, a positive trend in the dynamics of the number of published materials of conferences and articles using the keywords "cyber" and "frauds" in the international database Scopus during the years 2000-2023 was revealed, which only indicates the growth of interest in this topic in scientific circles. The analysis of publications by keywords with the help of the analytical application VOSviewer made it possible to form the most important research clusters, among which the problem of insider cyber fraud is poorly studied.

In the research process, possible combinations of search queries in the Google search system were used, which made it possible to identify two sets of variables that are critically important for the topic of insider cyber fraud. The first included variables that directly characterize cyberattacks, the second - those that characterize the level of decreased trust in financial institutions. The obtained variables allow us to indirectly understand the behavior of insiders. Its simulation was implemented in three stages. At the first and second stages of the research, using the method of principal components and clustering by the k-means method, an array of the most relevant variables for further research was formed. The method of principal components made it possible to reduce the dimensionality of the data, and cluster analysis contributed to the formation of their groups. As a result, they included twelve of the twenty initial variables, which were grouped into three lists.

At the next stage of modeling, using associative analysis, three models of associative rules were built, on the basis of which the following conclusion was formulated - the most interesting for potential insiders-cyber fraudsters in banks is the personal financial information of the client, access to the personal account of the bank client, as well as gaining access to his phone Therefore, in order to minimize the consequences of the actions of insiders-cyber fraudsters in the bank, customers need to take appropriate preventive measures, namely: use multi-factor authentication for online banking transactions; communicate only with verified employees of the bank, who can accordingly confirm the fact that they really work in this bank, and do not share personal passwords and other confidential information directly with bank employees; regularly update security software, including anti-virus programs; monitor the activity of your own account in real time to detect unusual or suspicious activity; insure yourself against potential cyber fraud; etc.

References

1. Al, A. W. E. (2019). Employee Behavioural Factors and Information Security Standard Compliance in Nigeria Banks. *International Journal of Computing and Digital Systems*, *8*(4), 387–396. [CrossRef]

2. Almaiah, M. A., Al-Otaibi, S., Shishakly, R., Hassan, L. H., Lutfi, A., Alrawad, M., Qatawneh, M., & Alghanam, O. A. (2023). Investigating the role of perceived risk, perceived security and perceived trust on Smart M-Banking Application using SEM. *Sustainability*, *15*(13), 9908. [CrossRef]

3. Setyaji, P., Yin-Fah, B. C., & Chen, T. K. (2020). Cloud Based Intrusion Prevention System with Machine Learning Approach. *International Journal of Pharmaceutical Research*, *12*(2). [CrossRef]

4. Dewi, Y., Suharman, H., Koeswayo, P. S., & Tanzil, N. D. (2023). Factors influencing the effectiveness of credit card fraud prevention in Indonesian issuing banks. *Banks and Bank Systems*, *18*(4), 44–60. [CrossRef]

5. Google Trends. (2023). Query search. [Link]

6. Malik, A., Gehlot, S., & Vyas, S. (2022, October). Proposed Framework for Implementation of Biometrics in Banking KYC. In *International Conference on Computing, Communications, and Cyber-Security*, 193-202. Singapore: Springer Nature Singapore. [CrossRef]

7. Statista (2023). Number of cyber incidents in the financial industry worldwide from 2013 to 2022. [Link]

8. Rawat, R., Oki, O., Chakrawarti, R. K., Adekunle, T. S., Gonzáles, J. L., & Ajagbe, S. A. (2023). Autonomous Artificial Intelligence Systems for Fraud Detection and Forensics in Dark Web Environments. *Informatica, 47*(9). [CrossRef]

9. Rithani, M., Kumar, R. P., & Doss, S. (2023). A review on big data based on deep neural network approaches. *Artificial Intelligence Review, 56*(12), 14765–14801. [CrossRef]

10. Shamsi, M. A., Smith, D. D., & Gleason, K. C. (2023). Space transition and the vulnerabilities of the NFT market to financial crime. *Journal of Financial Crime, 30*(6), 1664–1673. [CrossRef]

11. Siano, A., Raimi, L., Palazzo, M., & Panait, M. (2020). Mobile Banking: An Innovative Solution for Increasing Financial Inclusion in Sub-Saharan African Countries: Evidence from Nigeria. *Sustainability, 12*(23), 10130. [CrossRef]

12. Ugochukwu, N. A., Goyal, S. B., & Sampathkumar, A. (2022). Blockchain-Based IoT-Enabled system for secure and efficient logistics management in the era of IR 4.0. *Journal of Nanomaterials*, 1–10. [CrossRef]

13. Wahab, F., Khan, I., Kamontip, Hussain, T., & Abbas, A. (2023). An investigation of cyber attack impact on consumers' intention to purchase online. *Decision Analytics Journal, 8*, 100297. [Link]

14. Yoro, R. E., Malasowe, B. O., Akazue, M. I., Ibor, A. E., & Ojugo, A. A. (2023). Evidence of personality traits on phishing attack menace among selected university undergraduates in Nigerian. *International Journal of Power Electronics and Drive Systems, 13*(2), 1943. [CrossRef]

15. Scopus (2023). Query result. [Link]

16. Kovalenko, I. I., Davydenko, Ye. O., & Shved, A. V. (2019). Metodyka poshuku asotsiatyvnykh pravyl. Visnyk Cherkaskoho derzhavnoho tekhnolohichnoho universytetu [Methods for finding associative rules. Bulletin of Cherkasy State Technological University], *3*, 50–55. (in Ukrainian) [CrossRef]

17. Kuzior, A.; Brożek, P.; Kuzmenko, O.; Yarovenko, H.; Vasilyeva, T. (2022a). Countering Cybercrime Risks in Financial Institutions: Forecasting Information Trends. *Risk Financial Management Journal*, 15, 613. [CrossRef]

18. Kuzior, A., Vasylieva, T., Kuzmenko, O., Koibichuk, V., Brożek, P. (2022b). Global Digital Convergence: Impact of Cybersecurity, Business Transparency, Economic Transformation, and AML Efficiency. *Open Innovation Technology, Market and Complexity Journal*, 8, 195. [CrossRef]

19. Kuzior, A., Kuzior, P. (2018). Identity Theft: The Escalation of the Problem – The Multidimensional Consequences. In *Von der Agorá zur Cyberworld. Soziale und kulturelle, digitale und nicht-digitale Dimensionen des öffentlichen Raumes*. Edited by Hg. Gerhard Banse and Xabier Insausti. Berlin: Trafo, 81-89. [Google Scholar]