

UDC 004.67:336.71(045)

DOI:10.18372/1990-5548.89.21481

¹Volodymyr Shymkovych,
²Pavlo Minin

INFORMATION SYSTEM FOR BANK PORTFOLIO CREDIT RISK ANALYSIS BASED ON A THREE-TIER DATA WAREHOUSE ARCHITECTURE

Department of Information Systems and Technologies,
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Kyiv, Ukraine
E-mails: ¹v.shymkovych@kpi.ua ORCID 0000-0003-4014-2786
²mininpa2001@gmail.com ORCID 0009-0007-3682-940X

Abstract—This paper proposes an information system for bank portfolio credit risk analysis based on a three-tier data warehouse architecture. The proposed solution integrates a centralized data warehouse, online analytical processing, a modular data import subsystem, and a Business Intelligence platform to support automated data integration, analytical processing, visualization, and reporting. A three-tier data warehouse model with subject-oriented data marts is developed to improve data organization and analytical performance. The architecture of the data import module is designed using a clear separation of responsibilities between software components, providing modularity, scalability, and maintainability. The analytical capabilities of the developed system are demonstrated through regression analysis of non-performing loans, confirming its applicability for automated portfolio credit risk assessment and decision support.

Keywords—Business intelligence, data warehouse, portfolio credit risk, banking information system, , dashboard, OLAP, ELT.

I. INTRODUCTION

Over the past decade, the banking sector has been actively implementing digital technologies, including Customer Relationship Management (CRM) systems, data analysis platforms, and intelligent information systems to automate business processes and improve customer service quality [1]. The development of artificial intelligence [2], machine learning [3], and interactive communication channel contributes to the personalization of banking services, automation of customer interaction, and data-driven decision-making [4],[5].

At the same time, the issue of developing analytical systems for assessing and predicting banking risks remains understudied. Modern banking institutions face the problem of effective analytical processing of large volumes of financial data, which include both market information and internal operational indicators [6]. The constant growth of data flows, the limited capabilities of traditional analysis methods, and the insufficient level of automation of risk management processes complicate timely management decision-making [7]. Periodic changes in reporting templates from the National Bank of Ukraine add complexity to this process. The integration of traditional approaches to banking risk management with artificial intelligence, machine learning, and big data analysis technologies allows for increased risk assessment accuracy, responsiveness to changes in the economic

environment, and compliance with regulatory requirements [6]–[8].

Credit risk is particularly important in this context because lending constitutes one of the major areas of banking activity and directly affects both regulatory capital requirements and the financial performance of banks. According to the supervisory statistics of the National Bank of Ukraine, a significant share of the regulatory capital burden is associated with the loan portfolio. The structure of the corresponding risk components is illustrated in Fig. 1.

In addition to its impact on regulatory capital, lending represents a substantial source of banking income. As of November 1, 2023, interest income associated with lending to individuals and legal entities accounted for approximately 48% of the total income of the Ukrainian banking system [13]. The corresponding income structure is presented in Table I.

Thus, the high share of credit operations in the structure of bank income and their significant contribution to the regulatory capital burden confirm the importance of effective tools for comprehensive portfolio credit risk analysis.

The use of modern Business Intelligence (BI) systems, data warehouses, and specialized analytical platforms enables banking institutions to effectively process large amounts of financial information, conduct comprehensive analysis of credit portfolios, and ensure timely identification of potential risks [9], [10]. The integration of such technologies contributes

to increasing the efficiency of risk management processes, improving the quality of management decision-making, and ensuring compliance with modern regulatory requirements [10], [11].

The aim of the work is to increase the efficiency of credit risk analysis by developing a bank portfolio credit risk analysis system based on three-tier data warehouse architecture.

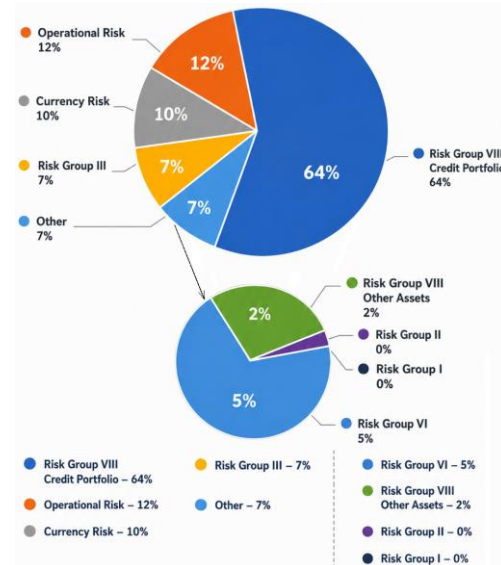


Fig. 1. Structure of risk components of regulatory capital of Ukrainian banks based on statistical data of the National Bank of Ukraine

TABLE I. INCOME STRUCTURE OF UKRAINIAN BANKS

Indicator	Sum, thous. UAH	Share
Interest income from corporate income tax of legal entities	75, 024, 498.13	30.87%
Interest income from personal income tax	41, 733, 182.15	17.17%
Interest income from KP	116, 757, 680.28	48.03%
Other income	126, 315, 204.51	51.97%
Total income	243, 072, 884.79	100.00%

II. LITERATURE REVIEW

BI and data warehouse technologies are widely used to support data-driven decision-making in financial and banking institutions. BI systems combine data integration, centralized storage, analytical processing, reporting, and visualization tools, thereby providing decision-makers with consolidated information obtained from heterogeneous operational sources [9]. In the banking domain, such capabilities are particularly important because analytical processes involve large volumes of transactional, customer, financial, regulatory, and external economic data that must be continuously integrated and processed.

The effectiveness of data warehouse technologies in the banking sector was investigated in [10]. Their empirical study demonstrated that data quality, system quality, user satisfaction, and organizational benefits significantly influence the effectiveness of data warehouse systems in banks. Related studies have also examined the adoption and utilization of Business Intelligence and Analytics technologies in

banking institutions. The study presented in [14] analyzed technological, organizational, and environmental factors affecting BIA usage in the banking sector and emphasized that the effectiveness of such systems depends not only on technological infrastructure but also on organizational readiness and human resources. A subsequent study [15] confirmed the importance of technological and organizational factors for BI utilization in commercial banks. Qatawneh [16] further demonstrated that BI adoption has a significant positive influence on decision-making performance in banking institutions.

Another important research direction concerns the application of analytical and machine-learning methods to credit risk assessment. The study in [11] performed an extensive benchmarking of classification algorithms for credit scoring and demonstrated the potential of advanced computational approaches compared with traditional statistical techniques. The systematic review presented in [17] showed that modern machine learning and deep learning models can provide high

predictive performance, while identifying data imbalance, dataset inconsistency, and model transparency as major remaining challenges. The study in [18] subsequently systematized consumer credit risk assessment approaches in terms of classification algorithms, data characteristics, and learning methods, emphasizing the importance of appropriate data preparation and scenario-specific modeling.

Most studies in this research direction, however, concentrate primarily on the predictive accuracy of individual credit scoring or credit risk models. Such approaches typically assume that the data required for modeling have already been collected and prepared and therefore pay less attention to the architecture of the information infrastructure responsible for heterogeneous data acquisition, consolidation, storage, multidimensional analytical processing, and delivery of results to risk analysts and bank management.

A more integrated approach was proposed in [19], where the authors developed an automated credit assessment framework combining data warehouse technologies, automated ETL processes, and machine learning based credit risk estimation. Their solution demonstrates the advantages of integrating operational data ingestion with predictive analytics and represents an important step toward end-to-end automation of credit assessment. However, the framework is primarily oriented toward borrower-level automated credit assessment and machine-learning-based calculation of expected credit losses rather than portfolio-level analytical support based on subject-oriented data marts, multidimensional Online Analytical Processing (OLAP), interactive BI tools, and integration of heterogeneous internal and external banking information sources.

The reviewed studies therefore demonstrate substantial progress in three complementary areas: implementation and adoption of BI technologies in banking [10], [14] – [16], development of advanced credit risk assessment models [11], [17], [18], and integration of data warehouses with automated analytical processes [19]. At the same time, the reviewed approaches do not provide an integrated architecture specifically oriented toward bank portfolio credit risk analysis that simultaneously combines heterogeneous data acquisition, centralized storage, subject-oriented analytical data marts, OLAP processing, BI-based visualization and reporting, and modular integration with external information sources. This limitation determines the need for an information system architecture that provides a complete data processing cycle from data acquisition

and consolidation to portfolio risk analysis and decision support.

III. PROBLEM STATEMENT

Credit operations constitute a significant component of banking activities and simultaneously represent one of the major sources of financial risk. Therefore, effective portfolio credit risk management requires timely consolidation and analysis of information on loans, borrowers, reserves, regulatory indicators, and external economic factors. The National Bank of Ukraine establishes regulatory requirements for capital adequacy and continuously publishes supervisory information that can be used for banking risk analysis [12], [13]. Consequently, portfolio credit risk assessment requires the joint processing of both internal banking information and data obtained from external regulatory and statistical sources.

In existing analytical workflows, part of this information may be processed using separate office files and independent analytical tools. A typical organization of portfolio credit risk analysis based on MS Office Professional and data obtained from external information sources is illustrated in Fig. 2. Such an approach becomes inefficient when the number and size of datasets increase. In particular, working with multiple large MS Excel files complicates data synchronization, increases the amount of manual data preparation, creates risks of duplication and inconsistency, and limits scalability. A file containing approximately 250–300 thousand records can reach approximately 250 MB, while its loading may require 2.5–3 minutes depending on workstation performance. Consequently, the use of distributed working files becomes a bottleneck for repeated portfolio analysis and preparation of management reports.

Therefore, the problem considered in this study is the development of an information system capable of integrating heterogeneous banking and external data sources into a unified analytical environment and providing automated support for bank portfolio credit risk analysis.

Let the set of input information sources be

$$D = D_{\text{ABS}}, D_{\text{files}}, D_{\text{API}}, D_{\text{ext}}, \quad (1)$$

where D_{ABS} denotes data from the automated banking system, D_{files} represents reporting and analytical files, D_{API} denotes information obtained through software interfaces, and D_{ext} contains external regulatory and macroeconomic data.

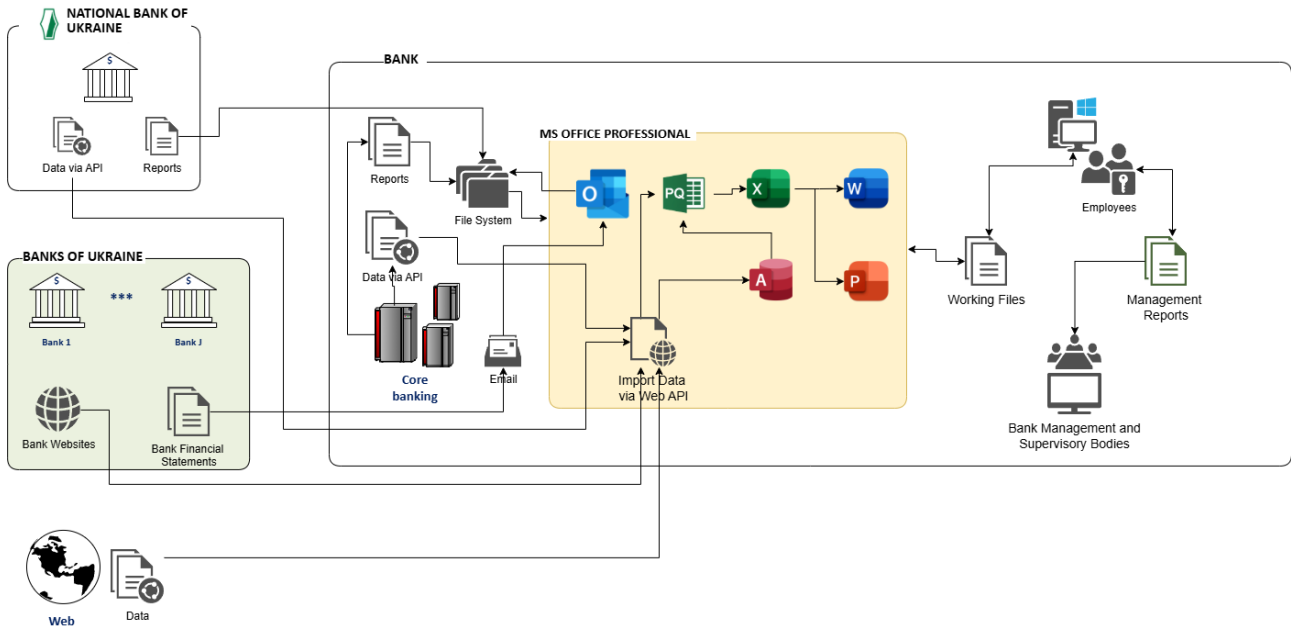


Fig. 2. Existing workflow for bank portfolio credit risk analysis using MS Office Professional and external data sources

The task is to construct an information system

$$S = \langle M_{\text{import}}, DW, OLAP, BI, M_{\text{API}} \rangle, \quad (2)$$

where M_{import} is the data import and transformation subsystem, DW is the centralized data warehouse, $OLAP$ is the multidimensional analytical processing layer, BI represents analytical, visualization, and reporting tools, and M_{API} provides integration with external information resources.

The developed system must provide: automated acquisition and consolidation of heterogeneous data; normalization and transformation of input information; centralized storage using a three-tier data warehouse architecture and subject-oriented data marts; multidimensional processing of portfolio indicators; integration of internal banking data with external regulatory and macroeconomic information; calculation and analysis of credit risk indicators; interactive visualization and reporting; and modularity and scalability of data import and analytical components.

As a result of processing the input data D , the system should form a consistent analytical representation D_A and a set of portfolio credit risk indicators R :

$$\Phi_S(D) \rightarrow \langle D_A, R, V \rangle, \quad (3)$$

where Φ_S is the data processing process by system S , D_A denotes integrated and prepared analytical data, R is a set of calculated portfolio credit risk indicators, and V represents analytical reports, dashboards, and visualization results used to support management decision-making.

Thus, the objective of the study is to design and implement an integrated information system that transforms heterogeneous banking data into a centralized analytical representation and provides tools for efficient portfolio credit risk analysis and decision support. The proposed solution should reduce dependence on distributed office files, decrease the amount of manual data processing, and create a scalable information infrastructure for further analytical and predictive procedures.

IV. BANK PORTFOLIO CREDIT RISK ANALYSIS SYSTEM

A. System Architecture

The proposed structure of the bank risk analysis automation system (Fig. 3) demonstrates the organization of the processes of data collection, integration, storage and analytical processing based on three-level data warehouse architecture. The system provides information from various sources, in particular the automated banking system, external information resources, reporting files and API interfaces, with subsequent loading into a centralized data warehouse. At the analytical processing level, the use of OLAP technologies is envisaged for the formation of aggregated indicators and multidimensional analysis, and the client level provides user access to visualization, reporting and data mining tools. Such architecture allows for comprehensive support of the bank's portfolio credit risk analysis processes and to increase the efficiency of management decision-making.

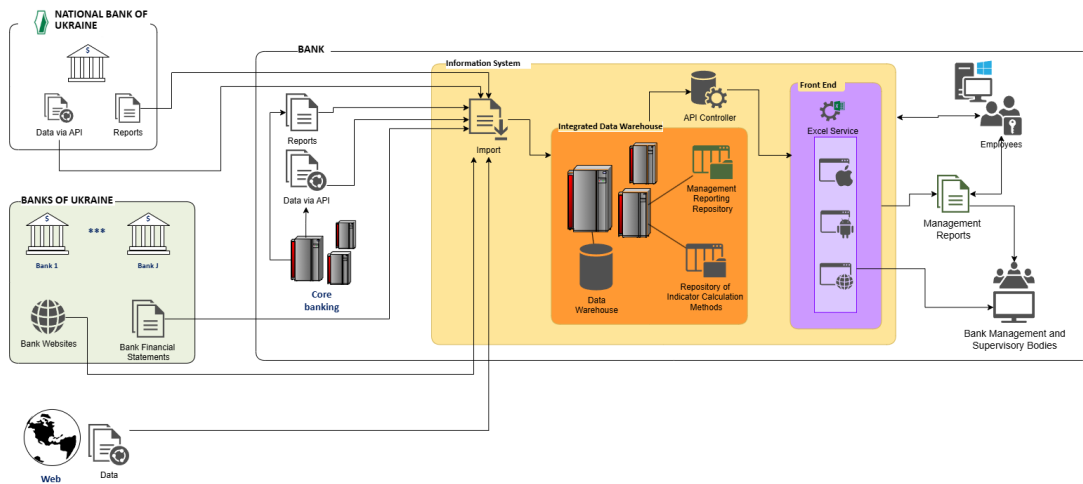


Fig. 3. A general framework for automating bank risk analysis is proposed

The bottom layer contains the database server, which is used to import data from many different sources. In the proposed system, these sources are transactional databases (automated banking system), separate reporting files, data obtained using the API interface.

The middle layer houses the OLAP server, which will transform the data into a structure better suited for analyzing and executing complex queries.

The top layer is the client layer. In the system, this level contains tools used for high-level data analysis, query building, and data mining.

The three-tier architecture demonstrates a number of advantages that are key to optimizing data processing and analysis in high-tech environments.

First of all, the distribution of functionality across different levels allows for high scalability of the system. Each level can be optimized separately, without affecting the rest of the components, which allows for an effective response to growing data processing needs.

Second, by separating the functionality and optimizing each level, the system can quickly and efficiently process large amounts of data and perform complex analytical queries.

In addition, the flexibility and modularity of this architecture allow the system to be easily adapted to the changing requirements and needs of banking institutions, which makes it an ideal choice for the proposed bank risk analysis automation system.

B. Three-tier Data Warehouse Model

A model for implementing a data warehouse based on interconnected subject data marts is proposed (Fig. 4). Such a structure is based on a logical representation of the relationships between the main entities of banking activities and provides optimization of the processes of storing, processing and analyzing information. The proposed model consists of three

interconnected tiers: the data integration and storage tier, the OLAP analytical processing tier, and the subject-oriented data mart tier. At the lower level, a central data warehouse is formed by integrating information from various sources, in particular transactional databases of an automated banking system, reporting files and external information resources accessible via API interfaces. Additionally, this level provides for the use of a metadata repository, which provides a description of the data structure, and a management reporting repository for storing prepared analytical materials. The structural organization of the proposed model is illustrated in Fig. 4. To formally describe the relationships between its tiers, the three-tier data warehouse model is represented as follows.

The proposed three-tier data warehouse model can be formally represented as

$$\mathcal{W} = \langle L_D, L_O, L_M, \Phi_I, \Phi_{DO}, \Phi_{OM} \rangle, \quad (4)$$

where L_D is the data integration and storage tier, L_O is the OLAP processing tier, and L_M is the subject-oriented data mart tier, Φ_I integration of sources into a centralized repository. The mappings Φ_{DO} and Φ_{OM} define the transformation of integrated warehouse data into multidimensional analytical structures and their subsequent organization into subject-oriented data marts, respectively.

At the first tier, heterogeneous data from the input sources defined in (1) are consolidated and normalized using the ETL procedures. The resulting integrated representation stored in the centralized data warehouse is defined as

$$D_W = \Phi_I(D), \quad (5)$$

where Φ_I denotes the integration, loading, and transformation operations applied to the heterogeneous input data D , and D_W is the consistent representation of the integrated banking data in the

centralized warehouse. Here, D_W in (2) denotes the data warehouse subsystem, whereas D_W denotes the integrated data representation stored within this subsystem.

At the second tier, the integrated data are transformed into a set of multidimensional OLAP areas:

$$O = \Phi_{DO}(D_W) = \{O_j | O_j = \Phi_j^{OLAP}(D_W), j = 1, \dots, p\}, \quad (6)$$

where O_j denotes the j -th analytical area, $\Phi_j^{OLAP}(D_W)$ is the corresponding aggregation and multidimensional transformation operator, and p is the number of analytical areas. In the proposed system, these areas include operational, client, credit, financial, and risk-related data.

At the third tier, subject-oriented data marts are formed according to particular analytical and management tasks:

$$M = \Phi_{OM}(O) = \{M_k\}_{k=1}^q, \quad (7)$$

where

$$M_k = \Psi_k(\{O_j | j \in J_k\}), \quad k = 1, \dots, q,$$

M_k denotes the k -th subject-oriented data mart, J_k defines the set of OLAP areas required for its formation, Ψ_k is the corresponding selection, aggregation, and transformation operator, and q is the number of subject-oriented data marts.

Thus, Fig. 4 provides the structural representation of the model, whereas (4)–(7) formalize the data transformations and relationships between its three tiers. Consequently, the proposed model implements a sequential transformation

$$D \xrightarrow{\Phi_I} D_W \xrightarrow{\Phi_{DO}} O \xrightarrow{\Phi_{OM}} M,$$

providing traceable processing of heterogeneous banking information from its acquisition and consolidation to multidimensional analysis and task-oriented representation for BI-based decision support. Within the notation introduced in (3), D_W , O , and M represent successive forms of the analytical data representation D_A , which is subsequently used by the analytical and BI components to obtain the portfolio credit risk indicators R and visualization results V .

In the implemented model, the OLAP tier comprises operational, client, credit, financial, and risk analytical areas, while the upper tier contains subject-oriented data marts for clients, the loan portfolio, risks, products, operational activities, financial results, provisions, and bank reporting. This organization enables analytical structures to be adapted to the information requirements of risk analysts, bank management, and other decision-makers.

One of the key advantages of the proposed approach is the centralized integration of data from various sources, in particular transactional databases of the automated banking system, reporting files and information received via API interfaces. This allows increasing the completeness, consistency and reliability of data used for the analysis of portfolio credit risks and the formation of management decisions. An important component of the proposed model is the support of analytical data processing based on OLAP technology, which ensures the rapid execution of complex analytical queries, the formation of aggregated indicators and the conduct of multidimensional analysis of large volumes of information. The use of OLAP structures allows increasing the efficiency of risk analysts and ensuring the prompt receipt of the necessary information for assessing the bank's credit risks.

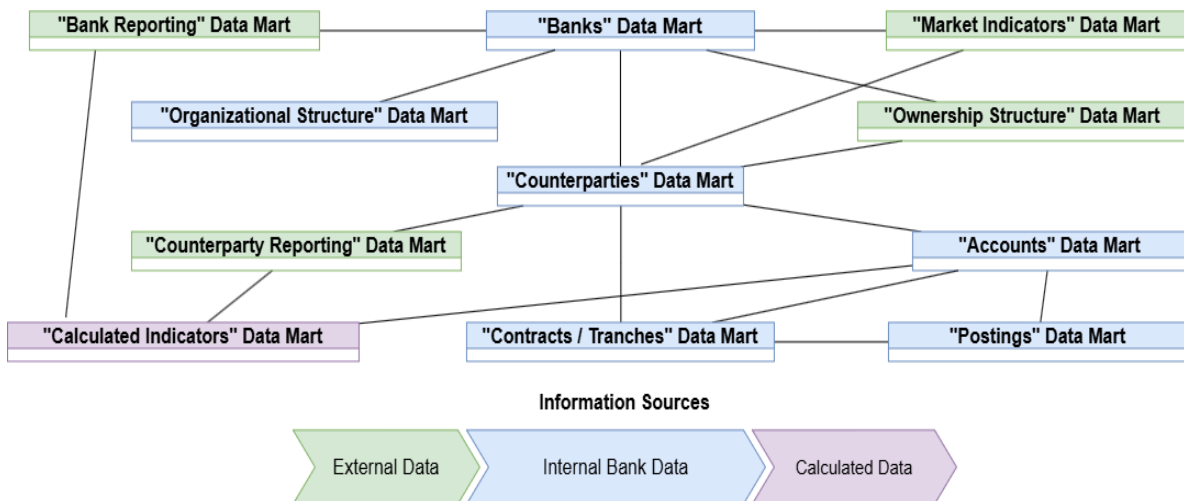


Fig. 4 A model of three-tier data warehouse architecture for analyzing bank portfolio credit risks

C. BI system integration

To determine the most appropriate software solution for implementing a bank's risk analysis automation system, a comparative analysis of three free open source BI platforms was conducted. The platforms were selected taking into account such criteria as operating system support, compatibility

with various database management systems, availability of analytical processing and data visualization tools, reporting capabilities, and functionality to support management decision-making. The results of the comparative analysis are presented in Table II.

TABLE II. COMPARATIVE ANALYSIS OF OPEN SOURCE BI SYSTEMS

Comparison criteria	Metabase	BIRT	KNOWAGE
Operating system	Windows, macOS, Linux	Windows, macOS, Linux	Windows, Linux, macOS, Docker
Database support	MySQL, PostgreSQL, SQLite, several others	Various, including MySQL, PostgreSQL, Oracle, SQLite	MySQL, PostgreSQL, Oracle, MS SQL Server, and others
Dashboard output support	+	+	+
KPI management and control	-	-	+
Analytical tools	Limited	Advanced	Advanced
Community and support	Active	Active	Active
Visual interface	Modern	Outdated	Modern

One of the platforms reviewed is Metabase, an open-source analytics solution designed for querying, analyzing, and visualizing data. The platform was developed by Metabase, headquartered in San Francisco (USA), and its first version was presented in 2014 by Samir Al-Sakran. Metabase is focused on simplifying user access to analytical information and supports work with a wide range of data sources.

Another platform considered is Business Intelligence and Reporting Tools (BIRT), an open source project that provides reporting and data visualization tools. BIRT is developed within the Eclipse Foundation. Initially, the project was initiated and supported by Actuate with the participation of IBM and Innovent Solutions. The platform is mainly focused on the formation of integrated reporting and the development of analytical components in corporate information systems.

The KNOWAGE platform was also considered – a comprehensive BI solution with open source code, which in previous versions was developed under the name SpagoBI. The project is supported by the European open software community OW2 and developed by Engineering. KNOWAGE provides a wide range of features for data analysis, reporting, dashboard creation and key performance indicator (KPI) management.

Based on the results of the comparative analysis, it was established that the KNOWAGE platform meets the requirements of the proposed bank risk analysis automation system to the greatest extent. KNOWAGE benefits from support for a wide range

of operating environments, including Windows, Linux, macOS, and Docker, as well as compatibility with a variety of database management systems, including MySQL, PostgreSQL, Oracle, and Microsoft SQL Server. This provides the possibility of integrating the platform with various data sources in the banking information system.

In addition, KNOWAGE supports advanced capabilities of analytical processing, creation of interactive information panels, creation of reports and monitoring of KPIs. The presence of an active community of developers and an open development model contributes to the further improvement of the platform. Thus, based on a set of defined criteria, KNOWAGE was chosen as the basic BI platform for the implementation of the bank's portfolio credit risk analysis automation system.

D. Division of Responsibilities Between Import Components

The module architecture is built on the basis of a multi-level organization of software components and includes Data Transfer Classes (DTO), a controller, a service layer, repositories, configuration components and file source processing tools. This approach ensures the separation of responsibility between individual system components, increases its modularity, simplifies software maintenance and creates opportunities for further expansion of functionality. The structure of the data import module of the proposed system is presented in the form of a class diagram (Fig. 5).

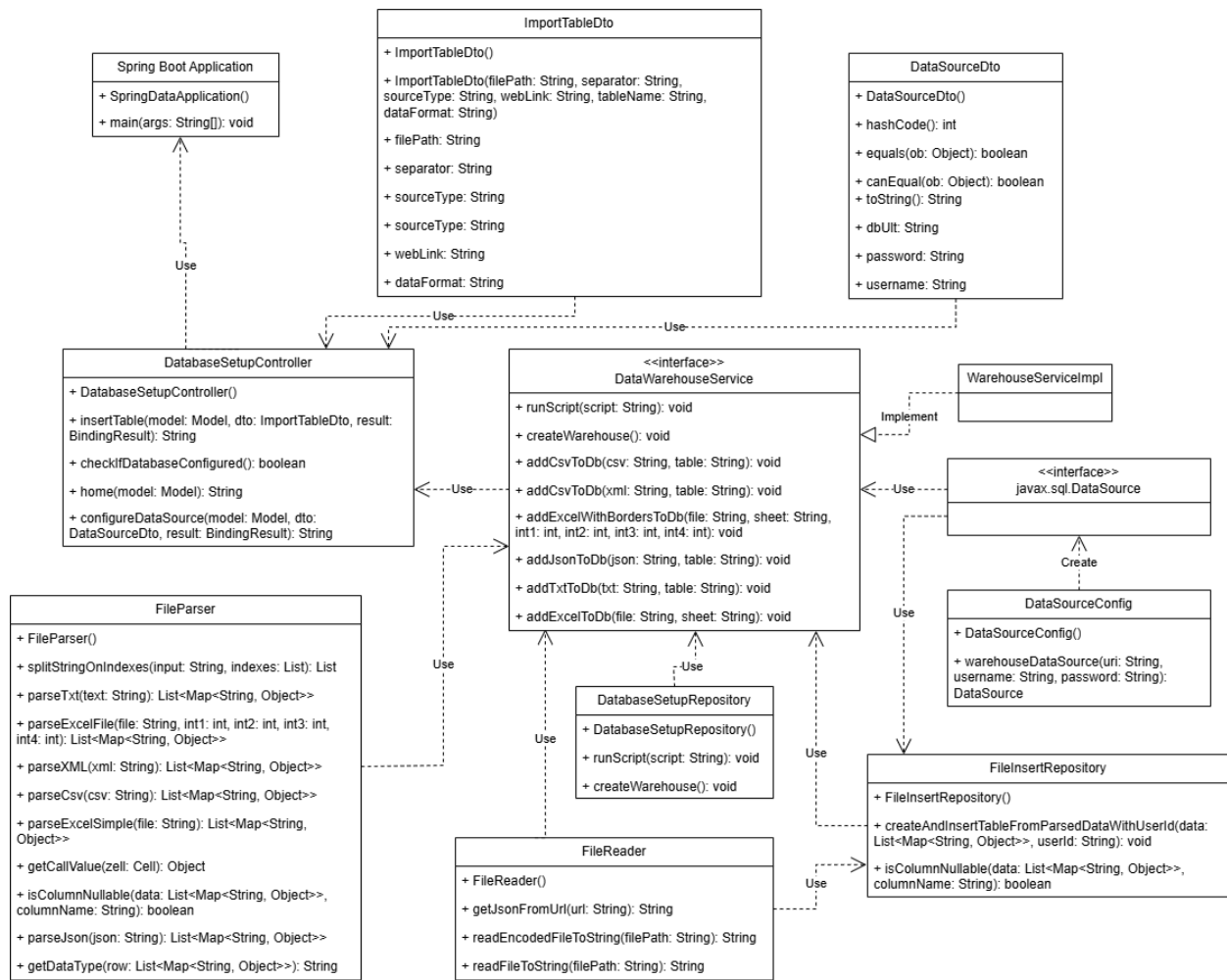


Fig. 5. Import module class diagram

At the top level of the module are the data transfer components ImportTableDTO and DataSourceDTO, which provide the exchange of parameters between individual parts of the system and describe the structure of input data and the parameters for connecting to information sources. The DatabaseSetupController controller is responsible for processing external requests and provides interaction between the presentation layer and the business logic of the system.

The central component of the module is the DataWarehouseService service and its implementation WarehouseServiceImpl, which provide the execution of import operations, conversion and loading of data into the warehouse. The lower level of the architecture is represented by the DatabaseSetupRepository and FileInsertRepository repositories, the DataSourceConfig configuration class, as well as the FileReader and FileParser components, which are responsible for working with file sources and converting the received information into the system's internal format.

The ImportTableDTO class is used to transfer parameters required for performing import operations. It contains attributes that define the characteristics of input data, in particular, the path to the file, data format, source type, table name and other import configuration parameters. Using this class provides a standardized mechanism for transferring information between system components and separates the structuring of input data from the logic of their processing.

The DataSourceDTO class is designed to describe data source parameters and store information required for establishing a connection. It contains source characteristics, in particular, the connection type, resource address and authentication parameters. Using this component provides flexibility for the import module and support for working with different types of information sources.

The DatabaseSetupController controller implements the level of interaction with external requests and is responsible for processing operations related to setting up and executing import processes. It provides acceptance of HTTP requests, transmission

of parameters to the service layer and formation of results of operations. This approach helps to separate the presentation and business logic levels, which increases the scalability and maintainability of the system.

The `DataWarehouseService` service implements the main business logic of the import module and ensures the execution of operations related to the preparation and loading of data into the repository. It contains methods for creating the repository structure, adding data from various formats (CSV, XML, Excel, JSON), as well as performing information conversion and storage operations. The interaction of the service layer with repositories provides abstraction of business logic from the details of physical data storage.

The `DatabaseSetupRepository` and `FileInsertRepository` repositories are responsible for interaction with the data storage layer. They implement database access operations and perform the necessary operations of creating, reading, updating and deleting information. Using the repository approach provides a unified mechanism for working with data and simplifies further scaling of the software solution.

The `DatasourceConfig` configuration component is designed for centralized management of connection parameters to external and internal data sources. It contains the settings necessary for creating and managing connections, which simplifies the system's adaptation to changes in the configuration of the bank's information infrastructure.

The `FileReader` and `FileParser` components provide work with file data sources. The `FileReader` class is responsible for receiving data from files of various formats, while the `FileParser` implements mechanisms for parsing and converting the received information into the system's internal format. This approach provides the ability to work with different file types and simplifies adding support for new formats in the future.

Thus, the proposed structure of the import module provides modularity, scalability and flexibility of the bank's risk analysis automation system. A clear division of responsibility between components simplifies the process of developing, testing and maintaining software, and also provides the ability to expand the functionality of the system in accordance with the changing requirements of the bank's information infrastructure.

E. Applying the ELT approach in the process of data integration and processing

In the context of building an effective data management system, the use of the Extract, Load, Transform (ELT) approach is proposed instead of the traditional Extract, Transform, Load (ETL) model.

The main difference between these approaches is the sequence of data transformation operations. In the ETL model, data is first extracted from sources, transformed in an intermediate environment, and only then loaded into the repository. In contrast, ELT involves the initial loading of raw data into the repository, where further transformation is performed directly by the data storage system.

Using the ELT approach allows for more efficient use of the computing resources of the database server, which is optimized for performing operations on large volumes of information and complex analytical queries. This increases the speed of data processing and reduces the need to use additional components for intermediate transformation.

Performing transformation operations directly in the database environment also contributes to an increase in the level of information security. Reducing the number of intermediate components and data access points reduces the risk of unauthorized intervention, which is especially important for banking information systems.

In addition, the ELT approach simplifies the system architecture and increases its flexibility. Changing the data transformation logic does not require modifying complex ETL processes, since the corresponding operations are performed directly in the data warehouse. This facilitates the scaling of the system and its adaptation to new business requirements.

Thus, the use of ELT in the proposed system ensures effective integration and processing of large volumes of data, simplifies the architecture of the information system, increases the level of security and creates the basis for further scaling of the solution in accordance with the needs of the banking sector.

V. RESULTS OF ANALYSIS OF PORTFOLIO CREDIT RISK INDICATORS

One of the key tasks of the proposed bank risk analysis automation system is to ensure effective analysis of loan portfolio indicators and identification of factors that affect the level of credit risk. To demonstrate the analytical capabilities of the system, a regression analysis of the share of Non-Performing Loans (NPL) was conducted. This indicator is one of the main indicators for assessing the quality of the loan portfolio and is used by the National Bank of Ukraine to monitor the state of credit risks in the banking system.

To conduct the analysis using the developed system, statistical indicators were loaded on the bank's loan portfolio, in particular information on the distribution of loans to individuals and legal entities by

debtor classes in accordance with the regulatory requirements of the NBU, the volume of formed reserves, the structure of lending by types of economic activity and currencies. Additionally, the NBU reference information was imported, necessary for normalization and reconciliation of data in the repository.

After loading information into the data repository, its automated normalization was performed using the PostgreSQL database management system. Reference data were converted into the appropriate structure of the analytics schema tables, and the main indicators were formed in specialized tables for further use in analytical procedures. This approach allows providing a single data storage structure and preparing an information basis for statistical analysis.

To build a regression analysis sample, data on the bank's loan portfolio by debtor classes were used. Based on these data, the NPL indicator was calculated, which was defined as the ratio of the volume of problem debt to the total volume of the loan portfolio. The results obtained were used for further research into the dependence of the NPL indicator on external economic factors.

TABLE IV. RESULTS OF ESTIMATING THE COEFFICIENTS OF THE NPL REGRESSION MODEL

Variable	Coefficient	Standard error	t-statistic	p-value	95% confidence interval
Constant	0.5699	0.0397	14.307	<0.001	[0.491; 0.649]
USD	-0.00322	0.00079	-4.066	0.00011	[-0.0078; -0.0027]

The results confirmed the statistical significance of the factor under study; however, the relatively low value of the coefficient of determination indicates the need for additional sample preparation to improve the accuracy of the model. The analysis was carried out using a 95% confidence interval.

The initial calculation showed the statistical significance of the model; however, the coefficient of determination R^2 had a relatively low value. This is explained by the fact that during a certain period the US dollar exchange rate remained fixed, which reduced the variability of the factor attribute and affected the accuracy of the regression model. To improve the quality of forecasting, an adjusted sample was formed by removing periods with an unchanged exchange rate value. The results of the regression model evaluation after sample adjustment are given in Tables V and VI. After removing periods with a fixed exchange rate value, the coefficient of determination of the model increased to $R^2 = 0.227$, which indicates an increase in the accuracy of approximation of the dependence of the NPL indicator on the US dollar exchange rate.

Analysis of the obtained regression dependence showed that with the increase in the US dollar

exchange rate, the predicted NPL value tends to decrease. This may be due to the structural features of the banking system's loan portfolio and the influence of other economic factors that were not included in the single-factor model.

TABLE III. MAIN STATISTICAL CHARACTERISTICS OF THE NPL REGRESSION MODEL

Indicator	Value
Number of observations	83
Plural R	0.412
R^2	0.169
Adjusted R^2	0.159
Standard error	0.069
Significance F	0.00011

exchange rate, the predicted NPL value tends to decrease. This may be due to the structural features of the banking system's loan portfolio and the influence of other economic factors that were not included in the single-factor model.

TABLE V. MAIN STATISTICAL CHARACTERISTICS OF THE NPL REGRESSION MODEL BASED ON THE ADJUSTED SAMPLE

Indicator	Value
Number of observations	64
Plural R	0.4764
R^2	0.2269
Adjusted R^2	0.2144
Standard error	0.0583
Significance F	6.9248×10^{-5}

Additionally, a similar analysis was conducted for a separate bank – JSC "SENSE BANK". The results of regression modeling of the NPL indicator for JSC "SENSE BANK" are given in Tables VII and VIII. The obtained value of the coefficient of determination $R^2 = 0.679$ indicates a high explanatory power of the model. The US dollar exchange rate has a statistically significant positive effect on the NPL indicator, which

is confirmed by the p -value of 5.76×10^{-17} . Therefore, for the bank under study, an increase in the exchange rate is accompanied by an increase in the share of non-performing loans in the loan portfolio.

TABLE VI. RESULTS OF ESTIMATING THE COEFFICIENTS OF THE NPL REGRESSION MODEL BASED ON THE ADJUSTED SAMPLE

Variable	Coefficient	Standard error	t -statistic	p -value	95% confidence interval
Constant (Y-intercept)	0.6121	0.0453	13.507	3.91×10^{-19}	[0.5216; 0.7026]
USD	-0.00667	0.00154	-4.327	6.92×10^{-5}	[-0.00975; -0.00359]

TABLE VII. MAIN STATISTICAL CHARACTERISTICS OF THE REGRESSION MODEL OF THE NPL INDICATOR FOR JSC "SENSE BANK"

Indicator	Value
Number of observations	64
Plural R	0.8244
R^2	0.6792
Adjusted R^2	0.6741
Standard error	0.0583
Significance F	5.76×10^{-17}

TABLE VIII. RESULTS OF ESTIMATING THE COEFFICIENTS OF THE NPL REGRESSION MODEL FOR JSC "SENSE BANK"

Variable	Coefficient	Standard error	t -statistic	p -value	95% confidence interval
Constant (Y-intercept)	-0.3515	0.0533	-6.593	7.86×10^{-10}	[-0.4579; -0.2451]
USD	0.01767	0.00154	11.468	5.76×10^{-17}	[0.01459; 0.02074]

In contrast to the general trend of the banking system, for JSC "SENSE BANK" dependence was established with the growth of the NPL indicator with an increase in the US dollar exchange rate. Such a result may indicate the need for additional analysis of the currency structure of the loan portfolio and improvement of approaches to managing problem debt.

The conducted study confirmed the possibility of using the proposed system for interactive analysis of the bank's portfolio credit risk indicators. The example of the NPL indicator demonstrated the full cycle of working with data: from their import and normalization in the data warehouse to performing statistical analysis and obtaining predictive dependencies.

The use of regression analysis made it possible to identify the presence of statistical dependencies between the loan portfolio quality indicator and an external economic factor – the exchange rate. The results obtained confirm the feasibility of using the proposed system as a tool to support management decision-making in the field of banking risk management.

VI. CONCLUSIONS

This paper proposed an information system for bank portfolio credit risk analysis based on a three-tier data warehouse architecture. The proposed architecture integrates a centralized data warehouse, OLAP processing, BI technologies, and a modular data import subsystem to support efficient storage,

processing, visualization, and analysis of banking data.

A three-tier data warehouse model with subject-oriented data marts was developed to organize analytical data processing and provide scalable support for credit risk assessment tasks. The architecture of the data import module was designed with a clear separation of responsibilities between controllers, services, repositories, configuration components, and data processing classes, improving the modularity, maintainability, and extensibility of the system. In addition, the advantages of the ELT approach over the traditional ETL process for banking applications were substantiated, demonstrating its benefits in terms of simplified architecture, improved scalability, and more efficient utilization of database resources.

A comparative analysis of open-source BI platforms showed that KNOWAGE best satisfies the functional requirements of the proposed system, particularly regarding dashboard support, KPI monitoring, analytical capabilities, and compatibility with multiple database management systems.

The analytical capabilities of the developed system were demonstrated through regression analysis of non-performing loans. The experimental results confirmed the feasibility of using the proposed architecture for automated processing, statistical analysis, and visualization of portfolio credit risk indicators, providing decision support for banking risk management. Overall, the proposed information system reduces dependence on large office files,

improves the efficiency of analytical workflows, and provides a scalable foundation for further development of banking risk analysis solutions.

REFERENCES

- [1] S. Kalyani and N. Gupta, "Is artificial intelligence and machine learning changing the ways of banking: a systematic literature review and meta analysis," *Discov Artif Intell*, vol. 3, no. 1, Dec. 2023, <https://doi.org/10.1007/s44163-023-00094-0>.
- [2] V. Shymkovych, G. Nowakowski and S. Telenyk, "Joint Punctuation Restoration and Text Capitalisation with a Hybrid XLM-RoBERTa—LSTM Model," *2025 IEEE 13th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Gliwice, Poland, 2025, pp. 990–997, <https://doi.org/10.1109/IDAACS68557.2025.11322226>
- [3] O. Steblianko, V. Shymkovych, P. Kravets, A. Novatskyi, and L. Shymkovych, "Scientific article summarization model with unbounded input length," *Inf. Comput. and Intell. syst. j.*, no. 5, pp. 150–158, Dec. 2024, <https://doi.org/10.20535/2786-8729.5.2024.314724>.
- [4] C. Ledro, A. Nosella, and A. Vinelli, "Artificial intelligence in customer relationship management: literature review and future research directions," *JBIM*, vol. 37, no. 13, pp. 48–63, Mar. 2022, <https://doi.org/10.1108/jbim-07-2021-0332>.
- [5] D. Ozay, M. Jahanbakht, P. J. Componation, and A. Shoomal, "State of the Art and Themes of the Research on Artificial intelligence (AI) Integrated Customer Relationship Management (CRM): Bibliometric Analysis and Topic Modelling," *2023 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD)*. IEEE, pp. 1–6, Nov. 22, 2023, <https://doi.org/10.1109/ictmod59086.2023.10438124>.
- [6] P. Kaushik, N. Girija, S. Malarvizhi, A. Hans, A. Kumar, and J. Singh, "Big Data Mining in Financial Risk Management," *2024 1st International Conference on Advances in Computing, Communication and Networking (ICAC2N)*. IEEE, pp. 1739–1744, Dec. 16, 2024, <https://doi.org/10.1109/icac2n63387.2024.10895187>.
- [7] M. Leo, S. Sharma, and K. Maddulety, "Machine Learning in Banking Risk Management: A Literature Review," *Risks*, vol. 7, no. 1, p. 29, Mar. 2019, doi: 10.3390/risks7010029.
- [8] Md. S. Rahman et al., "FinTech in sustainable banking: An integrated systematic literature review and future research agenda with a TCCM framework," *GF*, vol. 6, no. 1, pp. 92–116, 2024, <https://doi.org/10.3934/gf.2024005>.
- [9] P. Rikhardsson and O. Yigitbasioglu, "Business intelligence & analytics in management accounting research: Status and future focus," *International Journal of Accounting Information Systems*, vol. 29, pp. 37–58, Jun. 2018, <https://doi.org/10.1016/j.accinf.2018.03.001>.
- [10] A. Al-Okaily, M. Al-Okaily, A. P. Teoh, and M. M. Al-Debei, "An empirical study on data warehouse systems effectiveness: the case of Jordanian banks in the business intelligence era," *EMJB*, vol. 18, no. 4, pp. 489–510, May 2022, <https://doi.org/10.1108/emjb-01-2022-0011>.
- [11] S. Lessmann, B. Baesens, H.-V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, Nov. 2015, <https://doi.org/10.1016/j.ejor.2015.05.030>.
- [12] The Instruction on the procedure for regulating the activities of banks in Ukraine was approved by the Resolution of the NBU Board No. 638 of 28.08.2001 (as amended and supplemented). (In Ukrainian) URL: <https://zakon.rada.gov.ua/laws/show/z0841-01#Text> (accessed 07.08.2026)
- [13] Supervisory statistics of the National Bank of Ukraine. (In Ukrainian) URL: <https://bank.gov.ua/ua/statistic/supervision-statist> (accessed 07.08.2026)
- [14] A. Bany Mohammed, M. Al-Okaily, M. Al-Majali, and R. Masa'deh, "Business Intelligence and Analytics (BIA) Usage in the Banking Industry Sector: An Application of the TOE Framework," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 8, no. 4, p. 189, 2022, <https://doi.org/10.3390/joitmc8040189>.
- [15] A. Bany Mohammed, M. Al-Okaily, D. Qasim, and M. K. Al-Majali, "Towards an understanding of business intelligence and analytics usage: Evidence from the banking industry," *International Journal of Information Management Data Insights*, vol. 4, no. 1, p. 100215, 2024, <https://doi.org/10.1016/j.jjime.2024.100215>.
- [16] N. Qatawneh, "Empirical insights into business intelligence adoption and decision-making performance during the digital transformation era: Extending the TOE model in the Jordanian banking sector," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 4, p. 100401, 2024, <https://doi.org/10.1016/j.joitmc.2024.100401>.
- [17] S. Shi, R. Tse, W. Luo, S. D'Addona, and G. Pau, "Machine learning-driven credit risk: a systemic review," *Neural Computing and Applications*, vol. 34, no. 17, pp. 14327–14339, 2022, <https://doi.org/10.1007/s00521-022-07472-2>.
- [18] X. Zhang and L. Yu, "Consumer credit risk assessment: A review from the state-of-the-art classification algorithms, data traits, and learning methods," *Expert Systems with Applications*, vol. 237, p. 121484, 2024, <https://doi.org/10.1016/j.eswa.2023.121484>.

[19] N. Biswas, A. Sarkar Mondal, A. Kusumastuti, S. Saha, and K. C. Mondal, "Automated credit assessment framework using ETL process and machine learning," *Innovations in Systems and*

Software Engineering, vol. 21, no. 1, pp. 257–270, 2025, <https://doi.org/10.1007/s11334-022-00522-x>.

Received: June 19, 2026

Accepted: August 10, 2026

Published: September 18, 2026

Shymkovych Volodymyr. ORCID 0000-0003-4014-2786. Candidate of Technical Sciences. Associate Professor.

Department of Information Systems and Technologies, Faculty of Informatics and Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

Education: National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, (2010).

Research area: artificial intelligence, machine learning and deep learning; intelligent information systems; knowledge graphs; natural language processing; large language models and Retrieval-Augmented Generation; computer vision; neural network control systems; automated design of hardware components of neural network systems on FPGAs; embedded computer systems; intelligent data analysis.

Publications: over 70 publications.

E-mail: v.shymkovych@kpi.ua

Minin Pavlo. ORCID 0009-0007-3682-940X. Postgraduate Student.

Department of Information Systems and Technologies, Faculty of Informatics and Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine.

Education: Kyiv Polytechnic Institute National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine, (2024).

Research interests: information technologies in banking, machine learning and neural networks, algorithmic optimization and data analysis, automation in enterprise financial systems, software engineering and enterprise architecture.

Publications: 1.

E-mail: mininpa2001@gmail.com

В. М. Шимкович, П. А. Мінін. Інформаційна система для аналізу кредитних ризиків банківського портфеля на основі тривірневої архітектури сховища даних

У статті пропонується інформаційна система для аналізу кредитних ризиків банківського портфеля на основі тривірневої архітектури сховища даних. Запропоноване рішення інтегрує централізоване сховище даних, онлайн-аналітичну обробку, модульну підсистему імпорту даних та платформу бізнес-аналітики для підтримки автоматизованої інтеграції даних, аналітичної обробки, візуалізації та звітності. Для покращення організації даних та аналітичної продуктивності розроблено тривірневу модель сховища даних з предметно-орієнтованими вітринами даних. Архітектура модуля імпорту даних розроблена з використанням чіткого розподілу відповідальності між програмними компонентами, що забезпечує модульність, масштабованість та зручність обслуговування. Аналітичні можливості розробленої системи демонструються за допомогою регресійного аналізу непрацюючих кредитів, що підтверджує її застосовність для автоматизованої оцінки кредитного ризику портфеля та підтримки рішень.

Ключові слова: ВІ-система, аналітична система, автоматизація оцінки ризиків, дашборд, сховище даних, OLAR, ELT.

Шимкович Володимир Миколайович. ORCID 0000-0003-4014-2786. Кандидат технічних наук. Доцент.

Кафедра інформаційних систем та технологій. Факультет інформатики та обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Освіта: Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна, (2010).

Напрямок наукової діяльності: штучний інтелект, машинне навчання та глибоке навчання; інтелектуальні інформаційні системи; графи знань; обробка природної мови; великі мовні моделі та генерація з доповненим пошуком; комп'ютерний зір; нейромережеві системи керування; проектування апаратних компонентів нейромережевих систем на FPGA; вбудовані системи; інтелектуальний аналіз даних.

Кількість публікацій: понад 70 публікацій.

E-mail: v.shymkovych@kpi.ua

Мінін Павло Андрійович. ORCID 0009-0007-3682-940X. Аспірант.

Кафедра інформаційних систем та технологій. Факультет інформатики та обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.

Освіта: Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна, (2024).

Напрямок наукової діяльності: інформаційні технології в банківській справі, машинне навчання та нейронні мережі, алгоритмічна оптимізація та аналіз даних, автоматизація в корпоративних фінансових системах, інженерія програмного забезпечення та корпоративна архітектура.

Кількість публікацій: 1.

E-mail: mininpa2001@gmail.com