DOI: 00.00000/000-000-0000-00-0.CH2

Maksym Naumenko, Iryna Hrashchenko,
Tetiana Tsalko, Svitlana Nevmerzhytska, Svitlana Krasniuk, Yurii Kulynych

CHAPTER 2

INNOVATIVE TECHNOLOGICAL MODES OF DATA MINING AND MODELLING FOR ADAPTIVE PROJECT MANAGEMENT OF FOOD INDUSTRY COMPETITIVE ENTERPRISES IN CRISIS CONDITIONS

ABSTRACT

Developed in this research scientific and practical applied project solutions regarding Data Mining for enterprises and companies (on the example of food industry) involve the application of advanced cybernetic computing methods/algorithms, technological modes and scenarios (for integration, pre-processing, machine learning, testing and in-depth comprehensive interpretation of the results) of analysis and analytics of large structured and semi-structured data sets for training high-quality descriptive, predictive and even prescriptive models. The proposed by authors multi-mode adaptive Data Mining synergistically combines in parallel and sequential scenarios:

– methods of preliminary EDA,
– statistical analysis methods,
– business intelligence methods,
– classical machine learning algorithms and architectures,
– advanced methods of testing and verification of the obtained results,
– methods of interdisciplinary empirical expert interpretation of results,
– knowledge engineering formats/techniques – for discovery/detection previously unknown, hidden and potentially useful patterns, relationships and trends (for innovative project management).

The main methodological and technological goal of this developed methodology of multi-mode adaptive Data Mining for food industry enterprises is to increase the completeness (support) and accuracy of business and technical-technological modeling on all levels of project management of food industry enterprises: strategic, tactical and operational.

By optimally configuring hyperparameters, parameters, algorithms/methods and architecture of multi-target and multidimensional explicit and implicit descriptive and predicative models, using high-performance hybrid parallel soft computing for machine learning – the improved methodology of multimode Data Mining (proposed by the authors) allows to find/detect/mine for new, useful, hidden corporate knowledge from previously collected, extracted, integrated Data Lakes, stimulating the overall efficiency, sustainability, and therefore competitiveness, of food industry enterprises at various organizational scales (from individual, craft productions to integrated international holdings) and in various food product groups and niches.

In more detail, the purposes of this research are revealed in two meaningful modules:

1. The first part of the detailed goals and objectives of this research relate to the effective use of Data Mining (and modeling) in the competitive management of enterprises and companies in modern economy, namely:

– research and verification of the effectiveness of the basic/main three types of Data Mining in the management of a competitive enterprise;

– detection of basic/main difficulties and challenges of Data Mining technology in the management of a competitive enterprise;

– research and generation of a list of basic/main expedient functional applied corporate tasks for the application of the improved concept of Data Mining;

– determination of the list of basic/main results of using the proposed Data Mining concept and methodology for an effective and competitive enterprise in dynamic and crisis conditions;

– finding the basic/main advantages of using the proposed Data Mining concept and methodology for an effective and competitive enterprise in dynamic and crisis conditions;

– research of the basic/main technological problems of using the proposed concept and methodology of Data Mining for an effective and competitive enterprise in dynamic and crisis conditions;

– detection of the basic/main ethical problems of using the proposed Data Mining concept and methodology for an effective and competitive enterprise in dynamic and crisis conditions;

– research and search for basic/main perspectives of intelligent data analysis in the management of a competitive enterprise or company.

2. The second and main part of the detailed goals and objectives of this publication relate to the effective use of Data Mining (and modeling) in the competitive management of enterprises and companies in the food industry, namely:

– determination of features and methods of analysis and analytics of High Dimensional big data of at enterprises of the food industry;

– research of features and development of methodological and technological techniques for effective mode of OnLine Data Mining at food industry enterprises;

– research of specifics and development of recommendations regarding the effective mode of Ad-Hoc Data Mining at food industry enterprises;

– research of the specifics and development of applied recommendations regarding the effective mode of Anomaly & Fraud Detection of technological data of food industry enterprises;

– identification of directions and development of recommendations for effective use of Hybrid Data Mining at food industry enterprises;

– detection of features and development of a complex of scientific and practical recommendations regarding the effective regime of Crisis Data Mining at food industry enterprises in dynamic and unstable external conditions;

– identification of directions and development of recommendations for future trends in the effective use of Data Mining at food industry enterprises.

It can not be argued that in modern conditions (pre-crisis, crisis and post-crisis conditions of both regional food industries and the global world; globalization and simultaneous very narrow specialization of the food industry sectors; the need to take into account a huge amount of stream and packet information from various sources and various formats; the need for a quick adaptive optimal management response/adaptation in response to rapid changes in the global or regional market situation; unstable and difficult to predict dynamics of external influences: international, national, sectoral, local direct regulatory and indirect public regulation of the food industry) – deployment of the multi-mode adaptive Data Mining methodology proposed by the authors – will result in enterprises, companies and organizations/institutions of the food industry gaining additional competitive advantages at the state, regional, branch and corporate management levels.

KEYWORDS

Food industry enterprise, data mining, machine learning, big data, food industry project management, efficiency and competitiveness.

Stable, effective and competitive management involves a cycle of: strategic/tactical/operational planning; then high-quality, productive and timely implementation; monitoring/controlling/auditing results and adaptive feedback. In a business environment that rapidly and deterministically (and sometimes stochastically) changes in different directions with different rates of dynamics (the food industry in developing countries during the period of multimodal global and regional crises is characterized by this) [1], the above-mentioned stable and effective management is of crucial importance for enterprises seeking to achieve and maintain competitive advantage.

Optimal management of competition involves a set of innovative strategies, practices and algorithms aimed at increasing the total efficiency of the enterprise/company compared to its competitors. In other words, effective competitiveness management can be defined as a systematic approach to identifying, developing and implementing strategies, operational measures and tactics that increase the potential/ability of an enterprise to outperform its competitors [2]. This also involves constant monitoring of competitive factors and the environment, strategic planning and effective operational-tactical use of all types of resources (in particular, corporate knowledge) to achieve the goals of enterprises/companies/organizations, especially in difficult crisis conditions (in particular, the post-covid 19 consequences for the global and regional economies) [3].

It should be noted that in the current era of digital technologies Industry 4.0 and Industry 5.0 – ALL enterprises and companies generate huge amounts of structured, semi-structured and unstructured data from various sources (such as: internet social networks/media and network activity of users; sensors, controllers and robots; corporate systems ERP, CRM, MES, WMS, EAM, HRM) at all levels of detail and time horizons of management [4, 5]. The emergence of large structured, semi-structured and unstructured Big Data has radically changed operational, tactical and strategic management, efficiency, and therefore the development trends of modern competitive enterprises [6]. It is the intelligent use of all big data (in different formats, in different modes, qualitative and not yet qualitative) that significantly increases the integrated efficiency, stability and robustness of enterprises, which is especially relevant in multimodal crisis periods. That is, the analysis and analytics of big data will support the adoption of optimal and timely management decisions at all levels of enterprise management, will contribute to the timely reengineering of technological and business processes at the enterprise, will be an important element of the mechanism for detecting incomprehensible anomalies and detecting potential threats in the internal or external environment of the company [7].

Taking into account the above, it is necessary to emphasize that in the context of a stable, competitive market position of an enterprise or company, it is worth emphasizing adaptive innovative management – this is a dynamic approach to decision-making and resource management that combines training in monitoring and evaluation to adjust strategies in response to changing conditions [8]. Adaptive management involves iterative cycles of planning, implementation, monitoring and evaluation, where decisions are constantly adjusted based on new information and feedback. In adaptive management, these ideas allow enterprises, companies and organizations to improve resource allocation, optimize interventions and increase resilience to uncertainty and change [9]. It is Data Mining that plays a crucial role in adaptive management, using machine learning of all types on large accumulated data sets of all formats to obtain new, useful and implicit information, identify hidden patterns and patterns, and therefore to more effectively support data-driven decision making based on empirical precedents, heuristics.

The immediate classical concept of Data Mining is a process project of discovery/detection of hidden patterns/regularities, i.e. formalized new knowledge from accumulated data sets/heuristics [10]. However, the modern, proposed concept/paradigm of Data Mining (as a subset of Data Science) provides powerful intellectual techniques, methods/algorithms and scenarios for searching and formalizing valuable new, previously unknown, useful information (insights) from multidimensional large sets of structured batch and stream data [11]. That is, modern Data Mining (taking into account: the problem/curse of data dimensionality, on-line data mining, ad-hoc data mining, hybrid data mining, anomaly & fraud detection, crisis data mining) is a relevant and mandatory technology/tool in modern innovative adaptive management of the enterprise (food industry in particular [12]).

Crisis management involves systematic planning, coordination, and implementation of strategies for prevention, preparation, response, and recovery after emergencies or disasters. It must be noted that Data Mining plays an important, often key role in accurate monitoring, effective prevention, thorough preparation and optimal response to crises of all types and levels. That is, the specially adapted and configured Data Mining technology is end-to-end for the settlement of pre-crisis, crisis and post-crisis situations of the enterprise (especially in food industry).

As part of crisis management, adaptive Data Mining effectively scenario-configured and parametrically configured within special technological modes (online Data Mining, ad-hoc Data Mining, hybrid Data Mining, crisis Data Mining) is a relevant and important factor in ensuring not only the stability/sustainability of the enterprise (food industry in particular [13]), but also, even, ensuring its competitiveness in multimodal crisis conditions.

In crisis management, this knowledge can inform decision-makers, improve proactive awareness of crisis factors/phenomena, optimally allocate limited resources, and effectively reduce risks. Such data mining tasks as hierarchical and non-hierarchical clustering, binary and non-binary classification, search for association rules (in particular, unexpected rules), construction/training/learning (**Fig. 2.1** shows the result of such training progress) regression predicative EXPLICIT models ("white explicit models" – for example, regression trees or "gray explicit models" – for example, logit regression equations) or even IMPLICIT models – in the form of configured and trained ANN (**Fig. 2.2** shows the example of configured and trained ANN for classification of employees of an enterprise of food industry) – will contribute to enterprises and food industry companies to increase their readiness, response efficiency and effectiveness of proactive measures/influences for recovery/stabilization in various scenarios and stages of crisis phenomena.

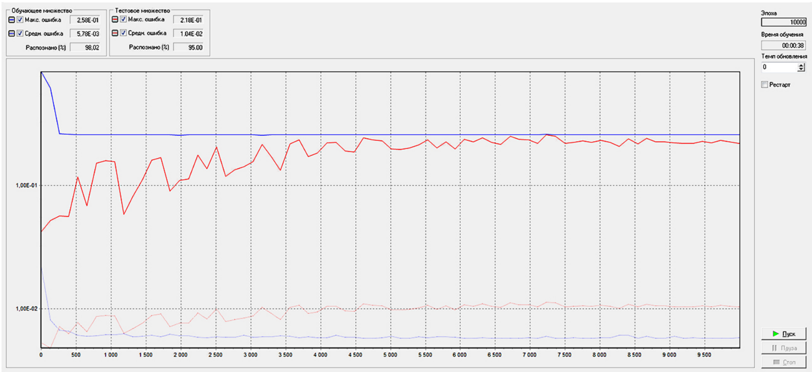


Fig. 2.1 The result of the progress of ANN training for the classification model of employees of an enterprise of food industry
Note: developed by the authors

Taking into account the above, the task of researching an effective and optimal concept of using innovative Data Mining modes in the effective adaptive management of enterprises/companies, with the aim of improving their competitiveness (especially in multimodal crisis conditions), becomes particularly relevant. Therefore, this publication aims to present the author's achievements and scientific and practical results (supported by theoretical studies, thematic industry research, accumulated industry heuristics and the author's empirical experience) regarding a specialized paradigm, concept, methodology and a set of operational and tactical measures for the optimal use of innovative Data modes Mining (and their proposed options/settings) for more effective and adaptive, anti-crisis management of the food industry enterprise.

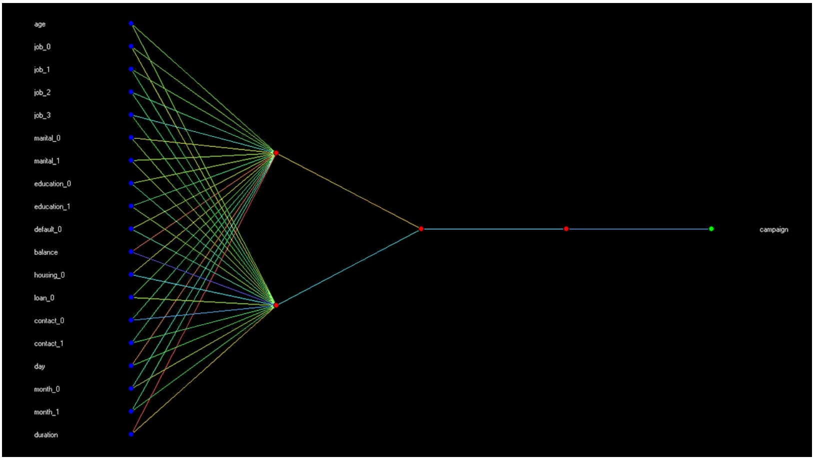


Fig. 2.2 The graph of configured and trained ANN Multilayer Perceptron for the classification model of employees of an enterprise of food industry (using Multilayer Perceptron architecture (3 hidden layers) with back-propagation, Sigmoid activation function, machine learning speed=0.1 and moment of inertia 0.9)
Note: developed by the authors

The provisions of the MODERN Data Mining (and Machine Learning) theory in unstable crisis conditions are revealed in the publications of such scientists as: V. Derbentsev [14], A. Matviychuk [15, 16], H. Velykoivanenko [17], etc.

But, taking into account the significant and systemic specificities/peculiarities of the food industry, the authors studied and analyzed the following relevant scientific articles in detail. Applying data mining techniques and analytic hierarchy process to the food industry was researched in [18]. Data mining and optimization issues in the food industry were analyzed in [19]. In [20] was paid attention to Data mining application for customer segmentation based on loyalty. Global food production and distribution analysis using data mining and unsupervised learning were developed in [21]. Application of Data mining in food trade network was analyzed in [22]. Mining logistics data to assure the quality in a sustainable food supply chain was developed in [23]. A framework for modeling efficient demand forecasting using data mining in supply chain of food products export industry was proposed in [24]. Predicting consumer preference for fast-food franchises: a data mining approach was described in [25]. A case study of customers grouping using data mining techniques in the food distribution industry was described in [26]. Data mining on time series: an illustration using fast-food restaurant franchise data was investigated in [27]. Consumers' behavior in the food and beverage industry through data mining was researched in [28]. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity were reviewed in [29]. However, relevant and unresolved issues are the features/specificities of the paradigm, concept, methodology, and set of operational-tactical measures for optimal setting and deployment of special Data Mining Modes when making effective decisions regarding competitive and sustainable management, especially in conditions of multimodal crisis phenomena (in particular, for food industry enterprises).

The aim of the research: research and development of an effective and optimal concept, methodology, technological techniques and modes, scenarios for the use of Data Mining (and modeling) in the adaptive management of food industry enterprises, with the aim of improving their competitiveness (especially in dynamic and unstable external conditions, and even, in crisis conditions).

In more detail, the purpose of this study is revealed in two meaningful modules:

1. The first part of the detailed goals and objectives of this publication relate to the effective use of Data Mining (and modeling) in the competitive management of enterprises and companies in various sectors of the economy, namely:

– research and verification of the effectiveness of the basic/main three types of Data Mining in the management of a competitive enterprise;

– detection of basic/main difficulties and challenges of Data Mining technology in the management of a competitive enterprise;

– research and generation of a list of basic/main expedient functional applied corporate tasks for the application of the improved concept of Data Mining;

– determination of the list of basic/main results of using the proposed Data Mining concept and methodology for an effective and competitive enterprise in dynamic and crisis conditions;

– finding the basic/main advantages of using the proposed Data Mining concept and methodology for an effective and competitive enterprise in dynamic and crisis conditions;

– study of the basic/main technological problems of using the proposed concept and methodology of Data Mining for an effective and competitive enterprise in dynamic and crisis conditions;

– detection of the basic/main ethical problems of using the proposed DATA MINING concept and methodology for an effective and competitive enterprise in dynamic and crisis conditions;

– research and search for basic/main perspectives of intelligent data analysis in the management of a competitive enterprise or company.

2. The second and main part of the detailed goals and objectives of this publication relate to the effective use of Data Mining (and modeling) in the competitive management of enterprises and companies in the food industry, namely:

– determination of features and methods of analysis and analytics of big data of high Dimensionality at enterprises of the food industry;

– study of features and development of methodological and technological techniques for effective mode of OnLine Data Mining at food industry enterprises;

– study of specifics and development of recommendations regarding the effective mode of Ad-Hoc Data Mining at food industry enterprises;

– study of the specifics and development of applied recommendations regarding the effective mode of Anomaly & Fraud Detection of technological data of food industry enterprises;

– identification of directions and development of recommendations for effective use of Hybrid Data Mining at food industry enterprises;

– detection of features and development of a complex of scientific and practical measures regarding the effective regime of Crisis Data Mining at food industry enterprises in dynamic and unstable external conditions;

– identification of directions and development of recommendations for future development, trends in the effective use of Data Mining at food industry enterprises.

In this research, a thorough and systematic review of specialized specialized scientific literature, industry reports and practical examples and author's experience of Data Mining in enterprise management (food industry in particular) was applied. Classical methods of analysis and synthesis, deduction and induction are used in combination with the author's heuristics and empirical insights. This methodological approach provides a detailed, systematic and effective study of the role, functionality, technological mechanisms, methods/algorithms, modes, advantages and caveats of successful Data Mining projects of technological and business data in the effective management of a competitive enterprise in the food industry.

2.1 INNOVATIVE AND EFFECTIVE APPLICATIONS OF DATA MINING IN COMPETITIVE ENTERPRISE MANAGEMENT

The results of the author's empirical observations, accumulated heuristics and conducted research indicate that it is intelligent data analysis that is a transformational tool for effective enterprise management in modern dynamic (and often crisis) conditions, which potentially provides significant advantages in terms of efficiency, adoption of optimal business and technological solutions and increasing competitive advantages. However, successful implementation of data mining requires addressing issues related to data quality, complexity, scalability, and ethical concerns. By responsibly using intelligent data analysis and taking into account the future achievements of intelligent technologies and appropriate specialized hardware, enterprises and companies can use the full potential of Data Mining not only to ensure stability and sustainability, but also to stimulate qualitative transformation and quantitative growth.

Proven Data Mining technologies in the management of a competitive enterprise provide for three applied areas [30]:

1. Descriptive Analytics: involves transformation, generalization of accumulated, historical data – to understand the situation and current state.

2. Predictive Analytics: involves using accumulated historical data to predict future events and trends.

3. Prescriptive Analytics: generates recommendations for decisions and actions, based on prebuilt descriptive and predicative models.

The main difficulties and challenges of Data Mining technology in the management of a competitive enterprise are identified [31]:

1. Data quality and availability (inconsistent and incomplete data can hinder quality intelligence; data privacy and security issues limit access to sensitive information, etc.).

2. Computational complexity (machine learning on big data requires significant computing resources; real-time analysis requires efficient algorithms and infrastructure, etc.).

3. Interdisciplinary integration (combining expertise from different industries, levels of management, regions are critical, but difficult).

The following functional applied corporate sectors are proposed for the application of the improved *Data Mining* concept [32]:

1. Customer Relationship Management (CRM).

Data mining in CRM helps businesses understand and predict customer behavior to improve customer satisfaction and retention. Key techniques include: clustering to group customers based on behavioral or demographic similarities to enable targeted marketing strategies; classification to predict customer churn or propensity to buy to inform retention strategies; defining association rules to define relationships between products to facilitate market basket analysis and cross-selling.

2. Financial Analysis and Fraud Detection. Data mining is critical in financial analysis for risk assessment and fraud detection, including anomaly detection (identifying unusual patterns that indicate fraudulent activity), predictive modeling (assessing credit risk by analyzing historical data to predict the likelihood of credit default).

3. Supply Chain Management (SCM). Data mining optimizes supply chain processes, including demand forecasting, inventory management and supplier evaluation, including through: time series analysis (forecasting future demand to improve inventory planning); clustering (classification of suppliers based on performance indicators to assist in supplier selection and management), etc.

4. Human resource management (HRM). Data mining supports HRM in talent acquisition, performance evaluation and employee retention (predictive analytics to identify potential candidates likely to succeed based on historical hiring data; text mining to analyze employee feedback to assess workplace satisfaction and identify areas, which need to be improved, etc.).

The following list of possible results of using the proposed Data Mining concept and methodology for an effective and competitive enterprise in dynamic and crisis conditions is defined:

1. Making informed decisions. Data mining provides a strong foundation for strategic decisions by uncovering hidden patterns and correlations in data, leading to more accurate and effective decision making.

2. Efficiency and cost reduction. Automated data analysis reduces time and resource consumption compared to traditional methods. Improved forecasting and inventory management reduce operational costs.

3. Improved and enhanced customer experience. Highly personalized marketing and improved adaptive customer service (retail and wholesale) based on data analysis contribute to increasing the level of customer satisfaction and loyalty.

The following basic advantages of using the proposed concept and methodology of Data Mining for an efficient and competitive enterprise in dynamic and crisis conditions are clarified:

1. Additional competitive advantage. Businesses that use data mining can gain a competitive advantage by identifying market trends and customer preferences over their competitors.

2. Improved scalability. Data mining techniques can be scaled to handle large volumes of data, making them suitable for businesses of all sizes.

The following basic technological problems of using the proposed Data Mining concept and methodology for an efficient and competitive enterprise in dynamic and crisis conditions were detected:

1. Quality of input data. The effectiveness of data mining is highly dependent on the quality of the data. Incomplete, noisy, or biased data can lead to inaccurate models and misleading insights. Suggested: Implementation of thorough data cleaning and pre-processing techniques results in higher data quality and more reliable results.

2. Complexity and experience. Data mining requires specialized knowledge in statistics, machine learning, and domain expertise. The complexity of algorithms and the need for qualified specialists can be an obstacle for some enterprises. It is suggested that additional investment in training and hiring skilled data professionals is critical to successful data analytics initiatives.

3. Scalability issues. As data volumes grow, ensuring that data mining algorithms scale effectively becomes a challenge. High computing demands may require advanced infrastructure and resources. It is suggested that the use of cloud computing and distributed processing systems such as Hadoop and Spark can help solve scalability problems.

The following basic ethical problems of using the proposed Data Mining concept and methodology for an effective and competitive enterprise in dynamic and crisis conditions were also detected:

1. Privacy & security. Data mining, as a rule, involves the most detailed and comprehensive analysis of all collected personal data, which has long been a concern of civil society regarding the privacy & security of personal data. Data-driven companies must ensure compliance with all regulations for the protection of personal and sensitive data, at a minimum, such as GDPR, and implement proactive and more comprehensive security measures against the leakage of this data. It should be emphasized that even more clear and transparent methodologies for the collection, integration, pre-processing, analysis and analytics, interpretation of the above-mentioned data, in parallel with obtaining the interactive informed consent of users, simultaneously with the implementation of more reliable security protocols regarding the leakage of this data – are already vital for maintaining credibility and relevance in the Data Mining industry.

2. Bias/fairness. Methods and technologies of data extraction, integration and pre-processing (for further ML) can, for example, inadvertently form input biases, biases/distortions in training samples, which will lead to unfair or discriminatory or incorrect ML results. It is proposed to use a complex of methods and technologies of mathematical statistics and comprehensive human expert analysis to identify and mitigate potential bias, regular audit and descriptive and predicative ML models, in particular, systematic testing of the reliability and completeness of predictive models.

3. Transparency/Accountability. The use of Data Mining (and Data Science) in the process of making management decisions at all levels should be unambiguous, clear and transparent, with detailed documentation of prerequisites, methodologies and justified responsibility for results. It is believed that the regulatory establishment/enforcement/audit of the use of normative ethical rules (rather than just principles) and systematic management of personal data will help to ensure the responsible, non-harmful use of such big data.

Below are the results of research into the future trends & prospects of data mining in the management of a competitive enterprise or company:

1. Integration of artificial intelligence (intelligent, knowledge-oriented technologies) within the framework of modern management. The integration of AI with Data Mining (and Data Science) will lead to significant qualitative and quantitative changes in the management of a modern enterprise. On the one hand, modern DM modes will ensure that AI knowledge bases are filled with new, relevant, hidden regularities/patterns; on the other hand, AI will help improve the accuracy and completeness of semi-supervised ML. Synergies between AI and DM will drive hybrid interdisciplinary innovation and thus further efficiency.

2. Data Mining (and even Data Science) in real time, 24/7/365, not only of batch, but also of streaming data. With the increasing prevalence of streaming structured, semi-structured and unstructured data in real time, companies are moving to Data Mining (and later to Data Science) in the mode of integration, processing, analysis, analytics, testing and interpreting the results – in real time. This approach allows competitive companies to instantly make optimal decisions based on current data, increasing responsiveness and flexibility. Real-time Data Mining (and even Data Science) is especially valuable in areas such as detecting errors/malfunctions in process equipment of conveyors or process flow production lines (for example, in the food industry), etc.

3. Data mining in IoT (especially in the Internet of Industrial Things). The proliferation of devices, equipment, equipment with IoT technology generates additional huge volumes of streaming data. Data Mining will manage to extract new, hidden, valuable additional useful information (regularities/patterns) from this data, contributing to improvements in food industry areas such as optimal predictive maintenance, intelligent adaptive continuous manufacturing, adaptive dispatching, etc.

4. AutoML and AutoDM. There is only one way to know which algorithm/method or ensemble of algorithms, which settings of their hyperparameters and local parameters will help to obtain the best model in each particular case, is to try all algorithms, their combinations, and all combinations of their parameters. But, if the ML engineer also takes into account all possible variants of data normalization and variants of objective functions, a combinatorial «explosion», «curse of dimensionality» will definitely occur. So, trying to try all of this by hand is impractical, so ML tool vendors have put a lot of effort into developing support systems of the AutoML class (so-called automated ML).

The best such tools combine settings and selection of parameter variants, selection of algorithms/methods and data normalization variants. Hyperparametric fitting of a better model (or models) is often performed in the later stages of an ML project. Hence, autoML is used to reduce human interaction and automate all tasks to solve real-world ML (and therefore DM) problems/problems. This functionality involves automating the entire iterative process from raw data to ML model testing.

This will significantly reduce the time and effort required to analyze and analyze big data and make it more accessible to companies of all sizes. Remember, even though AutoML doesn't require human interaction, that doesn't mean it's completely superior to it.

5. Expanded and improved methods of exploratory intelligent visualization of multidimensional structured and semi-structured data in EDA mode. Expanded and improved methods of visualization of the results of analysis and data analytics will help and make it easier for decision-makers to interpret the results of data mining in depth and interdisciplinary, and then act on them in a reasoned way. Moreover, pictographs, parallel coordinates and other methods of intelligence visualization of multi-dimensional data will help to find/notice hints, trends, ideas that will become the basis of further data research, their analysis and subsequent analytics.

2.2 INNOVATINE TECHNOLOGICAL MODES OF DATA MINING AND MODELLING FOR FOOD INDUSTRY ENTERPRISES IN CRISIS CONDITIONS

Data Mining in conditions of high dimensionality of data of enterprises and corporations of the food industry

Definition: High dimensionality refers to data sets with a large number of features (variables). Such data sets are common in very complex, interdisciplinary industries/productions/technologies of the food industry: baby food production, organic food production, diet (e.g. diabetic) and special (e.g. lactose-free, gluten-free) food production, and other products of deep complex technological processing.

Difficulties and problems associated with the high dimensionality of the company's input data:

1. Curse of Dimensionality (increased sparsity of data points in high-dimensional space; difficulty in recognizing meaningful patterns due to noise and irrelevant features).

2. Computational complexity (higher requirements for memory and processing power; longer computing time for model analysis and training).

3. Overfitting (increased risk of models capturing informational noise along with basic, deterministic patterns; reduced ability to generalize to new data in new external conditions).

4. Complexities of intelligence visualization within the framework of the EDA mode (below in **Fig. 2.3**, there is the attempt to perform discovery visualization (without prior dimensionality reduction) of the high dimensional financial characteristics of customers of network retail (aimed for the sale of food products).

As is clear from **Fig. 2.3**, high dimensionality creates significant analytical challenges, but it must be effectively managed using dimensionality reduction methods (PCA, t-SNE, autoencoders), deterministic feature selection (Filter Methods, Wrapper Methods, Embedded Methods) and regularization (L1 Regularization (Lasso) or L2 Regularization (Ridge)). These approaches improve model performance, reduce computational burden, and improve interpretation.

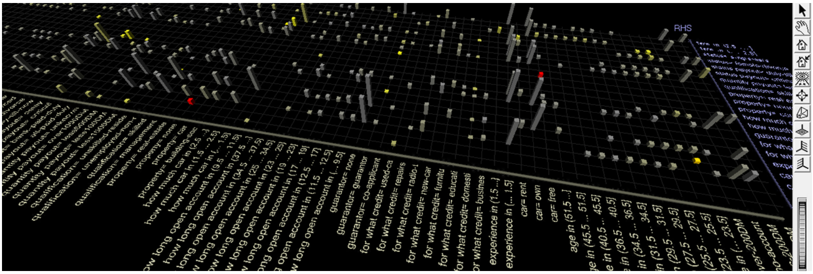


Fig. 2.3 The example of high dimensional EDA – discovery visualization (without prior dimension reduction) of the high dimensional financial characteristics of customers of chain retail (aimed for the sale of food products)
Note: developed by the authors

Online Data Mining (and later – Data Science) for food industry enterprises means the concept, methodology and technologies of analysis and analytics of big data, which are constantly generated, processed and updated in real time. This approach is different from traditional methods of batch processing and analysis, so it is crucial for those tasks and management functions of the food industry enterprise that require immediate analysis of streaming data and operational response. That is why, below, the actual results of the author's research on the principles, methodology, effective application, challenges and future directions of online Data Mining at food industry enterprises will be presented, emphasizing its role in dynamic and time-sensitive applications in the food industry.

In today's data-driven food industry, the need for real-time data analysis has become increasingly important. Online Data Mining addresses this requirement by enabling the continuous extraction of insights from all types and formats of streaming data. Let's consider the basic principles of Online Data Mining:

1. Incremental Processing: Data is processed incrementally (using data partitioning), ensuring that the system can handle high-velocity data streams without the need for complete retraining.

2. Continuous Learning: Online Data Mining involves algorithms that can learn and adapt continuously as new data arrives.

3. Adaptability: Algorithms must be adaptable to changing data clustering, general patterns, distributions, etc.

4. 24/7/365 Real-Time Online Analysis and Analytics: The primary goal is to provide real-time ideas & insights and appropriate decision-making capabilities.

Let's outline the basic methodologies of online Data Mining, because online Data Mining employs various methodologies, including:

– Online Streaming Algorithms: such algorithms designed to process data in a single pass or limited number of passes, such as the Hoeffding Tree, online K-means, and incremental PCA;

– Sliding Windows: techniques that use a fixed-size window of the most recent data points to update models continuously;

– Concept Drift Detection: methods for detecting and adapting to changes in data distribution over time, ensuring model relevance and accuracy;

– Incremental MLearning: machine learning methods that update models incrementally, such as online versions of support vector machines (or even RVM) and shallow neural networks.

The following are the main advantages of online data mining:

– Adaptability: capable of adapting to evolving data patterns and trends;

– Scalability: capable of handling big volumes of multidimensional data generated at high velocity;

– Resource efficiency: requires less computational resources compared to reprocessing entire datasets;

– Timeliness: provides immediate insights/ideas, enabling prompt decision-making and action.

Challenges and cautions of Online Data Mining:

1. Complexity: improvement, adaptation, support of effective algorithms/methods of online pre-processing of data, online analysis and online analytics is a rather non-trivial and complex task, which must also take into account the specifics of the country, region, specific enterprise and the specifics of the current managerial functional task/problem.

2. Data Quality: Ensuring the quality and consistency of streaming data will typically be a non-trivial and resource-intensive task.

3. Scalability: effective and timely management of the scalability of computing systems for online integration, data preprocessing, their analysis and analytics (especially predictive analytics) of high-speed data streams is a complex task that requires a highly experienced interdisciplinary team of human experts.

4. Latency: the need to check and test the results of online Data Mining – can cause a delay in making informed management decisions in real time.

5. Drift and evolution of the concept/paradigm encoded in the input data: the complexity of continuous adaptation and retraining due to temporal dynamics in: the distribution of data, their cluster structure, the number and values of local extremes and other changes in the fitness function or production function of the food industry enterprise.

Prospects and trends of online Data Mining at food industry enterprises:

1. Integration with Edge Computing: Using ancillary/peripheral computing to integrate and pre-process data as close as possible to its source, reducing latency and/or resource lag, and therefore maximizing throughput.

2. Advanced Machine Learning: the use of advanced machine learning techniques, such as deep machine learning using Deep Neural Networks, to enhance the capabilities of online unstructured data science [33].

3. Distributed Processing: use of innovative distributed computing infrastructures for effective script scaling of online Data Mining processes.

4. Explainable AI: development of alternative scenario methods and hybrid technologies for providing in-depth and extended interpretation and transparency of constructed models in the online Data Mining process.

The very clear example of this trend is the decision tree configured and built by the authors (**Fig. 2.4**), for the task "Estimation of Obesity Levels Based on Eating Habits and Physical Condition", which as the result helps to interpret and understand the relationship between the predicted target variable (the weight category of the consumer/patient – divided into 7 degrees: from underweight – to obesity of the 3rd degree) and other input categorical and quantitative factors (gender, age, height, weight, family history of obesity, indicators: Do you usually eat vegetables in your meals? How many main meals do you have daily? Do you eat any food between meals? Do you smoke? Do you monitor the calories you eat daily? How often do you have physical activity? How much time do you use technological devices such as cell phone, videogames, computer and others? How often do you drink alcohol? Which transportation do you usually use?).

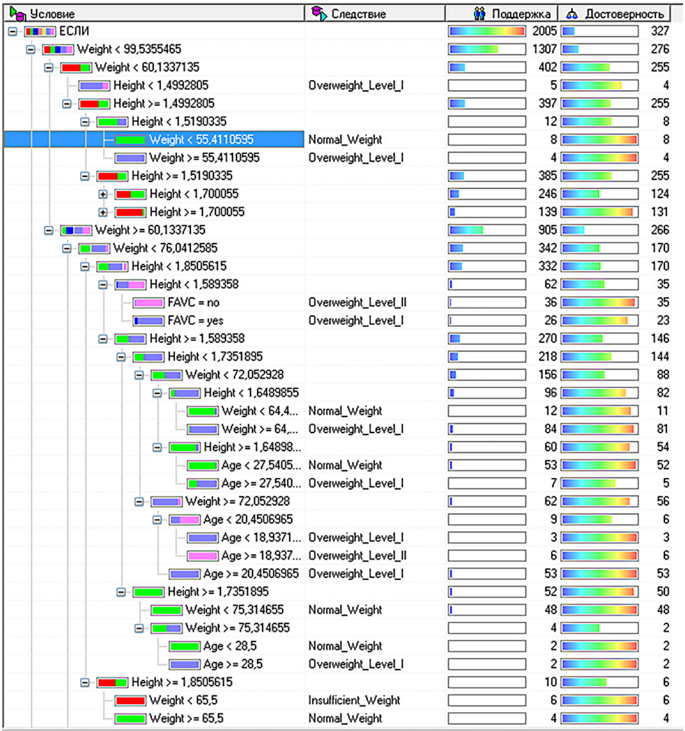


Fig. 2.4. The example of configured and constructed explainable decision tree using the C4.5 algorithm for the «Estimation of Obesity Levels Based On Eating Habits and Physical Condition» task, which helps to interpret and visually understand the relationship between the target variable and other input categorical and quantitative factors
Note: developed by the authors

Another example of the verbalization of such an implicit model (black box) obtained from a shallow ANN configured and trained by the authors with three hidden layers is given in **Fig. 2.5**.

```
Input database fields (features) (input symptoms):
          REGION
          FRESH
          MILK
          GROCERY
          FROZEN
          DETERGENTS
          DELICASSEN

Output database fields (output syndrome):
          CHANNEL

Preprocessing of database field values for transmission to network training
REGION=(REGION-2)/1
          FRESH=(FRESH-56077)/56074
          MILK=(MILK-36776,5)/36721,5
          GROCERY=(GROCERY-46391,5)/46388,5
          FROZEN=(FROZEN-30447)/30422
          DETERGENTS=(DETERGENTS-20415)/20412
          DELICASSEN=(DELICASSEN-23973)/23970

Functional converters
          Sigmoid1(A)=A/(0,1+|A|)
          Sigmoid2(A)=A/(0,1+|A|)
          Sigmoid3(A)=A/(0,1+|A|)

Syndromes of 1 level:
          Syndrome1_1=Sigmoid1( -0,07807371*REGION-0,2318511*-
FRESH-0,1125404*MILK+0,1231333*GROCERY+0,04119471*FROZEN-0,3295803*DETERGENTS+0,21418
85*DELICASSEN-0,1244661 )
----------------------------
          Syndrome1_10=Sigmoid1( 0,04964642*REGION+0,006197082*-
FRESH-0,01756757*MILK-0,05744172*GROCERY-0,09775436*FROZEN-0,03407726*DETERGENTS+0,01
003292*DELICASSEN+0,05423961 )
.....................................................................
Syndromes of 3 level:
          Syndrome3_1=Sigmoid3( -0,05175032*Syndrome2_1-0,08382219*Syn-
drome2_2-0,06745245*Syn-
drome2_3-0,08310731*Syndrome2_4+0,01224828*Syndrome2_5+0,01331895*Syndrome2_6-0,05745
047*Syndrome2_7-0,0655295*Syndrome2_8+0,03557626*Syndrome2_9+0,003902948*Syndrome2_10
-0,09182979 )
----------------------------
          Syndrome3_10=Sigmoid3( -0,05422281*Syndrome2_1-0,252343*Syn-
drome2_2-0,02300238*Syn-
drome2_3+0,0822472*Syndrome2_4-0,660238*Syndrome2_5+0,1182159*Syndrome2_6+0,04464602*
Syndrome2_7+0,03296235*Syndrome2_8-0,006042616*Syndrome2_9-0,2184449*Syndrome2_10+0,0
1921556 )

Output syndromes:
          CHANNEL_1=0,005829105*Syndrome3_1+0,3342025*Syn-
drome3_2-0,007643878*Syndrome3_3-0,06917655*Syndrome3_4-0,09772637*Syndrome3_5+0,2204
48*Syndrome3_6+0,2064538*Syndrome3_7-0,1307883*Syndrome3_8-0,1251397*Syndrome3_9-0,41
09597*Syndrome3_10-0,02931657
          CHANNEL_2=-0,04215042*Syndrome3_1-0,2311292*Syn-
drome3_2-0,07857881*Syndrome3_3-0,05260604*Syndrome3_4-0,05044578*Syndrome3_5-0,20043
5*Syndrome3_6-0,1850976*Syndrome3_7+0,289685*Syndrome3_8+0,1089799*Syndrome3_9+0,3452
721*Syndrome3_10-0,06909672

Final post-processing of output syndromes:
          CHANNEL belongs to the range corresponding to the syndrome with the maximum
value of the binary reliability score CHANNEL_1,...,CHANNEL_2
```

Fig. 2.5. Fragment of the experimental attempt of verbalization of the IMPLICIT model («black box»), generated from the configured and trained shallow ANN with three hidden layers and sigmoid activation function (this model is trained for binary classification of retail distribution channel of 6 main groups of food products for a holding/concern from food industry)
Note: developed by the authors

5. Ethical Considerations: further improvement and expansion of regulation/formalization of regulation (as at the interstate level – GDPR, national, industry and even corporate levels) of possible ethical issues related to data privacy and potential bias when using the results of online Data Mining.

Online Data Mining is primarily recommended for use in solving such problems of food industry enterprises as:

– detection of staff abuses or errors in real time;

– adaptive algorithmic recipe management;

– adaptive algorithmic management of the schedule and modes of operation of technological equipment (especially relevant – considering the current state of energy supply in Ukraine);

– medical and biological monitoring;

– detection of anomalies in raw materials in real time;

– monitoring of the company's information and communication network;

– proactive forecasting and timely detection of malfunctions of technological equipment;

– adaptive planning and dispatching of technical maintenance and all types of repairs (ad-hoc, current, major repairs, modernization) of equipment;

– total comprehensive quality control of the food industry enterprise in real time.

Thus, as a conclusion on this matter, it can be stated that online Data Mining is a key approach to discover new, hidden patterns/knowledge in real-time 24/7/365 from streams of structured, semi-structured and unstructured big data that are constantly generated at food industry enterprises. Using stream (apach kafka & apach sparkl etc.) and parallel (map reduce) technologies of machine learning organization, ensemble machine learning (bagging, stacking and boosting etc.), specialized RE-learning strategies (Learning rate decay, Transfer learning, Training from scratch, Dropout etc.), methods of detecting drift/trend of the concept (for example, the displacement vector of the center of the cluster) – same online Data Mining provides timely, scalable and operational analytical support for the management of food industry enterprises. As the development and simultaneous cost reduction of peripheral distributed computing technologies, AutoML, distributed preprocessing of big data (not only ETL or ELT, but also Change Data Capture, Data replication, Data virtualization, Stream Data Integration) continues, the future of intelligent data analysis is online contains significant potential for further improvement of the process of making effective and competitive management decisions in real time by food industry enterprises.

Ad-hoc Data Mining refers to the exploratory and flexible application of data mining techniques to address specific, often unplanned, analytical queries/problems. Unlike traditional, classical approaches to Data Mining, which adhere to a predefined, deterministic set of algorithmic procedures, Ad-hoc Data Mining is characterized by its high adaptability, flexibility and ability to quickly respond to immediate analytical needs.

Ad-hoc Data Mining involves the spontaneous and situational use of those Data Mining methods/algorithms and technologies that allow data exploration and the search for patterns/patterns WITHOUT a pre-set, fixed, a priori analytical structure/prerequisites/parameters/constraints. This approach is especially valuable in dynamic environments/tasks where the types of analytical tasks, modes of data processing and their analytics, types of patterns, forecasting time horizons, quantification of classification tasks, etc., are rapidly changing. That is, ad-hoc Data Mining allows enterprises to adapt to new information conditions and requirements (internal and external) very quickly and in a timely manner, which is especially important in times of crisis. The example of performed ad-hoc binary classification (using the kNN method) of promising or non-promising places of commercial fishing of crabs is shown in **Fig. 2.6**, and the fragment of generated PMML code of this model is presented in **Fig. 2.7**.

Let's detail the principles of Ad-Hoc Data Mining:

– User-Driven: typically initiated by end-users or analysts who require immediate answers to specific questions;

– Flexibility: Ad-hoc Data Mining allows for the adjustment of analytical approaches based on the specific context and detailed requirements of the current query;

– Iterative Process: the process is iterative, involving repeated cycles of data exploration, analysis, refinement and deployment;

– Exploratory Nature: it emphasizes exploration and discovery, often used to generate unexpected hypotheses or uncover hidden patterns in data.

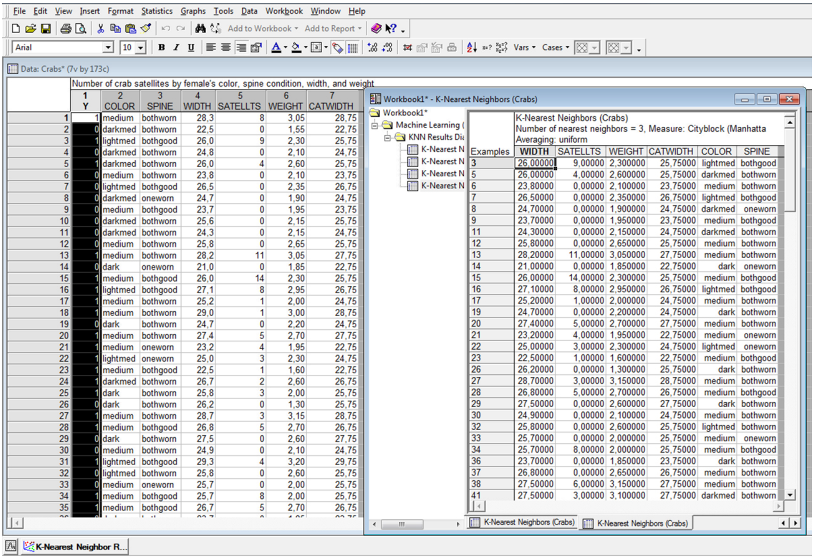


Fig. 2.6 The example of the ad-hoc binary classification (using the kNN method) of promising/unpromising places for commercial fishing of crabs (after a series of experiments, the following optimal settings of this ad-hoc binary classification model were determined: kNN classification algorithm, $k=3$, distance measure = Cityblock, averaging = uniform, sampling = 0.75)

Note: developed by the authors

```xml
<?xml version="1.0" encoding="windows-1251" ?>
<PMML version="2.0">
<DataDictionary numberOfFields="7">
          <DataField name="Y" optype="continuous">
          </DataField>
          <DataField name="WIDTH" optype="continuous"/>
          <DataField name="SATELLTS" optype="continuous"/>
          <DataField name="WEIGHT" optype="continuous"/>
          <DataField name="CATWIDTH" optype="continuous"/>
          <DataField name="COLOR" optype="categorical">
                    <Category label="dark" value="5"/>
                    <Category label="darkmed" value="4"/>
                    <Category label="lightmed" value="2"/>
                    <Category label="medium" value="3"/>
</DataField>
          <DataField name="SPINE" optype="categorical">
                    <Category label="bothgood" value="1"/>
                    <Category label="bothworn" value="3"/>
                    <Category label="oneworn" value="2"/>
</DataField>
</DataDictionary>
<KNearestNeighborModel modelName="STATISTICA K-Nearest Neighbor" knn_type="Regres-
sion" noOfNearestNeighbors="3" metric="cityblock" weighted="no">
          <KNNSchema>
                    <KNNField name="Y" usageType="predicted">
                    </KNNField>                   <KNNField name="WIDTH"
shift="-1,68000000000000e+000" scale="8,00000000000000e-002" />
                    <KNNField name="SATELLTS" shift="0,00000000000000e+000"
scale="6,66666666666667e-002" />
                    <KNNField name="WEIGHT" shift="-3,00000000000000e-001"
scale="2,50000000000000e-001" />
                    <KNNField name="CATWIDTH" shift="-3,25000000000000e+000"
scale="1,42857142857143e-001" />
                    <KNNField name="COLOR" shift="0,00000000000000e+000"
scale="1,00000000000000e+000" >
                              <Category label="dark" value="5"/>
<Category label="darkmed" value="4"/>                       <Category
label="lightmed" value="2"/>
<Category label="medium" value="3"/>              </KNNField>
<KNNField name="SPINE" shift="0,00000000000000e+000" scale="1,00000000000000e+000" >
                              <Category label="bothgood" value="1"/>
<Category label="bothworn" value="3"/>                      <Category label="oneworn"
value="2"/>
</KNNField>         </KNNSchema>
<KNearestNeighborExamples noOfExamples="129">
<Example id="1">
<AttributeInstance name="Y" value="1,00000000000000e+000"/>
<AttributeInstance name="WIDTH" value="2,60000000000000e+001"/>
<AttributeInstance name="SATELLTS" value="9,00000000000000e+000"/>
<AttributeInstance name="WEIGHT" value="2,30000000000000e+000"/>
<AttributeInstance name="CATWIDTH" value="2,57500000000000e+001"/>
<AttributeInstance name="COLOR" value="2,00000000000000e+000"/>
<AttributeInstance name="SPINE" value="1,00000000000000e+000"/>
</Example>
..................................................................
<Example id="129">
<AttributeInstance name="Y" value="0,00000000000000e+000"/>
<AttributeInstance name="WIDTH" value="2,45000000000000e+001"/>
<AttributeInstance name="SATELLTS" value="0,00000000000000e+000"/>
<AttributeInstance name="WEIGHT" value="2,00000000000000e+000"/>
<AttributeInstance name="CATWIDTH" value="2,47500000000000e+001"/>
<AttributeInstance name="COLOR" value="3,00000000000000e+000"/>
<AttributeInstance name="SPINE" value="2,00000000000000e+000"/>
</Example>
</KNearestNeighborExamples>
</KNearestNeighborModel>
</PMML>
```


Fig. 2.7 Fragment of generated PMML code of configured and trained classification model by using kNN method (k=3) (this model is built for binary classification of promising/unpromising places for industrial fishing/harvesting of crabs from data set with 129 records)

Note: developed by the authors

Ad-hoc Data Mining employs a variety of methodologies, including but not limited to: Query-Based Analysis, Visual Data Exploration, Descriptive Statistics, classic models of Machine Learning, Text Mining, Web Mining, Process Mining, SNA etc.

Ad-hoc Data Mining is applied across various levels and functions of food enterprise management, including:

– Business Intelligence (quickly addressing business queries, such as sales trends, customer behavior, and market analysis, etc.);

– Biological safety and Healthcare staff (investigating patient data for immediate insights into treatment outcomes, disease patterns, and healthcare utilization, etc.);

– Finance and Risk Management (analyzing financial transactions and market data to identify fraud, assess risks, and optimize investment strategies, etc.);

– BtB and BtG E-Commerce (exploring customer purchase data to uncover buying patterns, product preferences, and inventory needs, etc.);

– Information Security and Cyber Security of food enterprise telecommunications (examining call records and network data to detect anomalies, optimize network performance, and improve customer service, etc.).

Advantages of ad-hoc Data Mining: Timeliness (ensures faster understanding of the problem and context, which allows the management of food industry enterprises to respond more quickly to new, unexpected problems and obstacles); Customization (adjusts and adapts Data Mining to specific current urgent needs, providing more targeted, relevant and narrowly focused results); sometimes Resource Efficiency (in some simple cases and situations – reduces the need for careful preliminary planning and long-term allocation of resources – which is typical for traditional corporate Data Mining projects).

Challenges and cautions of Ad-hoc Data Mining for food enterprise management:

– Data Quality: ensuring the accuracy and reliability of data used in ad-hoc analyzes can be challenging, especially with unstructured, incomplete or noisy data;

– Scalability: Ad-hoc approaches may struggle to scale efficiently when dealing with very large semi-structured and unstructured datasets or complex queries from Data Lakes;

– Consistency: maintaining consistency and reproducibility of results can be difficult without standardized procedures;

– Skill Requirements: qualified and experienced interdisciplinary engineers and analysts with the ability to quickly adapt are needed to effectively apply and deploy various methods/algorithms of big data preprocessing, analysis and analytics (including hybrid and soft computation methods [34]).

Prospects and trends of Ad-hoc Data Mining at food industry enterprises:

– Enhanced Tool Integration: development of more sophisticated and customized tools that integrate ad-hoc Data Mining capabilities with user-friendly interfaces and advanced analytical function;

– Automated Ad-hoc Analysis: leveraging artificial intelligence to automate parts of the ad-hoc analysis process, improving speed and reducing the manual effort required;

– Collaborative Platforms: creating platforms that facilitate collaboration among analysts, allowing for shared insights and more comprehensive ad-hoc analyses;

– Real-Time Data Processing and Analysis: advancing real-time data processing and Analysis technologies to support immediate ad-hoc queries and analysis on streaming data.

Ad-hoc Data Mining is a crucial approach in the modern data-driven landscape, offering the flexibility and speed needed to address immediate analytical needs (for example, during unexpected situations in the logistics activities of the enterprise, complications in matters of currency import of important raw materials and equipment, in possible export VAT refund issues, unexpected changes in state and industry standards/norms, tax legislation, etc.). While ad-hoc Data Mining presents unique challenges, its advantages in providing timely, customized insights make it an invaluable tool across various domains of the food industry. Continued advancements in technology and methodologies with relevance of crisis management will further enhance the effectiveness and applicability of ad-hoc Data Mining for enterprises and companies in the food industry.

Mode of detection of anomalies in the intellectual analysis of technological data of food industry enterprises

Anomaly detection in structured, semi-structured, and even unstructured data mining is the process of identifying rare elements, events, or observations that are significantly different from the majority of data. Hence, the detection of anomalies, also known as the detection of outliers, is an important aspect of intelligent data analysis, and focuses on the detection of cases that differ significantly from the accepted/defined norm. These anomalies can detect and even predict critical consequences, such as: financial or logistical fraud, insider abuse in corporate management, intrusion into the company's information network; malfunction of the technological conveyor/equipment/equipment, etc. – which makes their detection (and better proactive forecasting) very important in the conditions of the specifics of the food industry. Thus, the main objective of the author's study of the Anomaly and Fraud Detection regime is to provide a comprehensive understanding of how anomaly detection affects different areas of activity and management levels of food industry enterprises, by identifying unusual data sets, their patterns/regularities and even, rare models (which may indicate relatively regular, but NOT massive events, facts, scenarios (with low statistical support)).

Components of the proposed concept of detecting anomalies in data:

1. Identifying Anomalies: Anomalies are data points that do not conform to expected patterns. They can be classified into point anomalies, contextual anomalies, and collective anomalies.

2. Types of data: Anomaly detection can be applied to different types of data, according to different classifications (numeric, categorical, time series and spatial data; structured, semi-structured and unstructured data; batch and streaming data, etc.).

3. The context of detecting anomalies in data: the context and specifics of the country, region, sector and segment of the food industry, macroeconomic conditions and the conditions of a specific enterprise/factory – greatly influences the choice of the Data Mining mode, the Data Mining method/algorithm, the optimal configuration of its hyperparameters and local parameters of the previously selected algorithm/method of Data Mining.

Accurate and complete detection of anomalies in the data of food industry enterprises involves the iterative, multi-stage use in synchronous and asynchronous modes of the proposed subset of methods/algorithms with different paradigms:

1. Exploratory Data Analysis (EDA), namely discovery visualization: use these methods at the first stage of the task of detecting anomalies in the data of food industry enterprises. To summarize the main characteristics of a dataset, often using visual methods, to understand its structure, detect patterns, anomalies, and test hypotheses. EDA is a crucial step in the data analysis process that provides a thorough understanding of the dataset, guiding further statistical analysis and modeling efforts. The authors recommend using parallel coordinates, pictographs (but NOT the Chernov method). Below, in **Fig. 2.8** given the performed example of the above-mentioned intelligence discovery visualization of data about emissions of harmful substances (CO_2, CH_4, N_2O) within the environmental management of a food industry enterprise – from this visualization, it is possible to notice the abnormal, jump-like change in emissions.

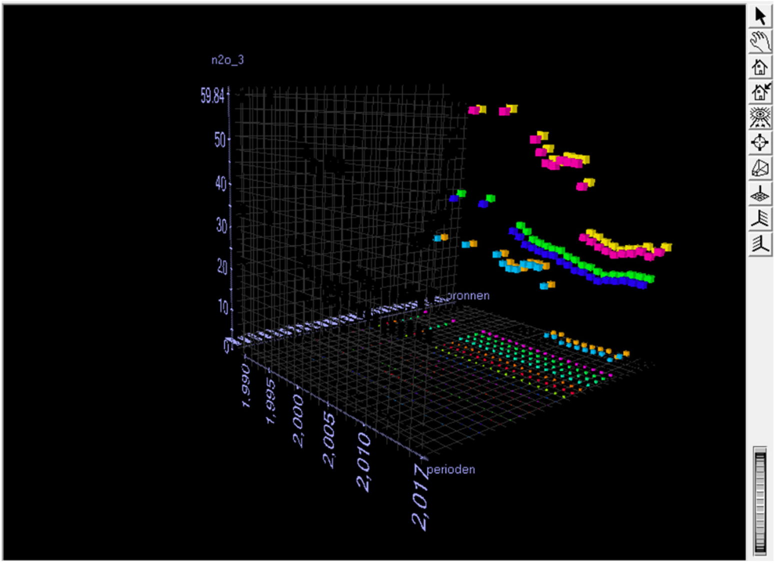


Fig. 2.8. The example of intelligence discovery visualization of data about emissions of harmful substances (CO_2, CH_4, N_2O) within the environmental management of a food industry enterprise
Note: developed by the authors

2. Statistical methods: at the second stage of anomaly detection – use several statistical algorithms simultaneously using the probability of specific data points (categories), in particular: Z-score, Grubbs test, Rosner test and the use of distribution-based threshold values.

3. Machine learning: the authors recommend combining different types of machine learning to detect anomalies in the data of food industry enterprises:

3.1. First you should use unsupervised machine learning: it predicts and detects anomalies in unfamiliar accumulated data (in a new subject area for the analyst/manager, a new food industry market, in a completely new task/situation) – that is why these data are usually NOT pre-labeled, and for this, it is recommended to use (specially configured):

3.1.1. Clustering methods: it is recommended by the authors to apply at the first stage – Hierarchical cluster analysis on a small random sample of big data set – to find the optimal parameter *k*. Using this already known optimal *k* value on second stage – for K-means or K-medians algorithm for the entire Big Data Set. Please, see **Fig. 2.9** with the illustration of the performed such preliminary hierarchical cluster analysis. It shows different variants of a priori *k* parameter: depending on the desired level of management: either *k*=2 for a higher/strategic level of management, or *k*=4 for middle/tactical level of management.

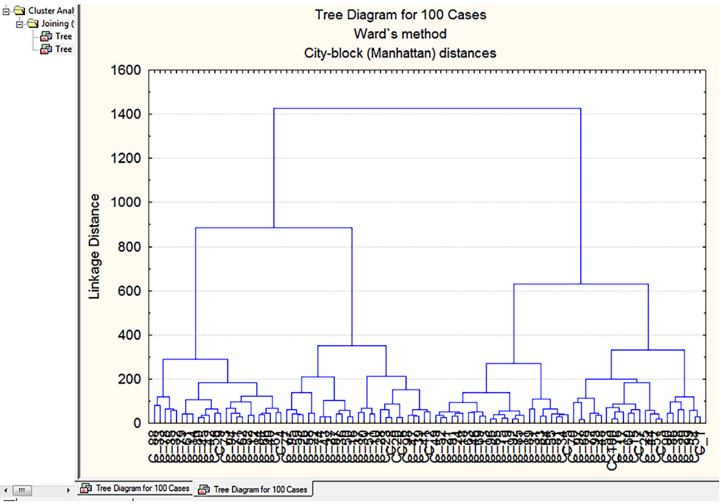


Fig. 2.9 The preliminary hierarchical cluster analysis trained on the random small sample (100 data vectors) from BIG input data set. A data set with the results of consumer blind testing of four different samples of products of a food industry enterprise was used
Note: developed by the authors

3.1.2. Algorithms for finding patterns in the form of rules – with the option of finding UNexpected rules – see author's example below, in **Fig. 2.10**.

3.1.3. Special architectures of Artificial Neural Networks in particular SOM, Autoencoders (as an example – see **Fig. 2.11** with the results of the author's application of the algorithm of unsupervised machine learning of artificial neural network SOM Kohhonen for detecting anomalies in multidimensional data).

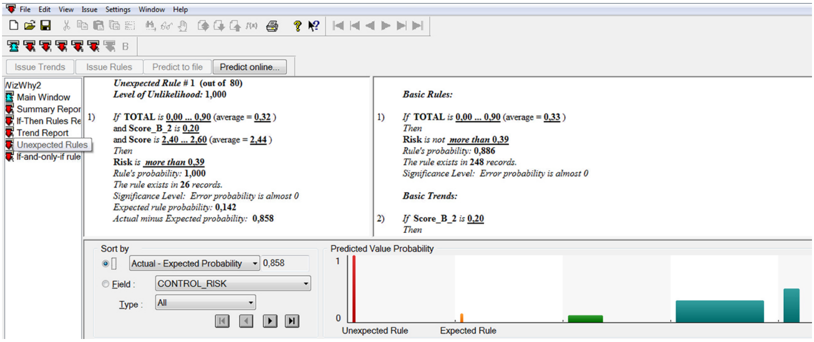


Fig. 2.10 The example of finding UNEXPECTED patterns in the form of rules – using modified Apriori method. The results of the financial audit of enterprises (food industry in particular) were used as input data
Note: developed by the authors

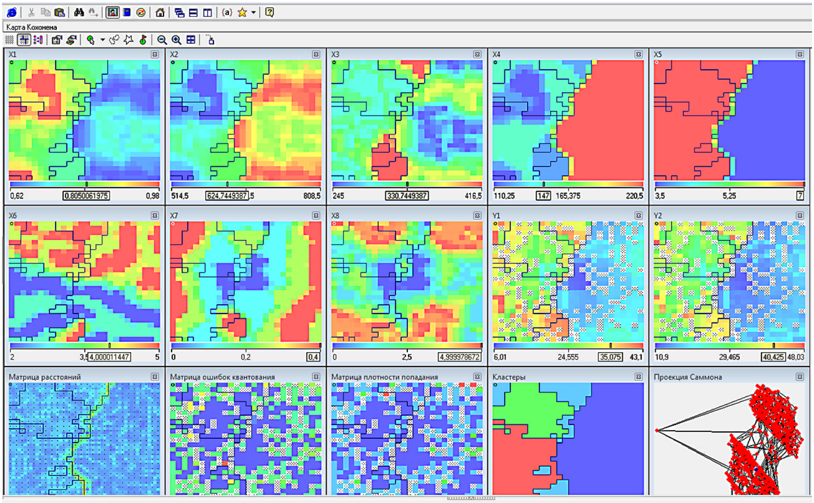


Fig. 2.11 The visualizations of the results of the Anomaly Detection Analysis (for energy audit purposes) using the dataset with projects of energy efficiency for food industry enterprises (mode and purpose of analysis: Anomaly and Fraud Detection visualization; algorithm: SOM Kohhonen; rectangular cells; neighborhood function: step function)
Note: developed by the authors

3.2. In the future, with the accumulation of initial experience and expertise regarding the current management task/situation, it is advisable to apply supervised machine learning: this type of ML uses labeled data to train models that distinguish between normal and abnormal cases. It is recommended that the authors use classification methods/algorithms. At the same time, it is recommended to first apply binary classification settings (typical/template operation/transaction/item or atypical/suspicious), and on the next iteration – ternary, and then use soft computational approaches).

3.3. With the deployment of the project/production of a food industry enterprise, Big Data will be accumulated, the analysis and analytics of which require the use of semi-supervised machine learning. semi-supervised machine learning combines marked/labeled data (of course, not a large amount) and large volumes of unlabeled data, followed by the application of Data Mining methods/algorithms to solve classification and forecasting problems. It is semi-supervised machine learning that will allow enterprises and companies in the food industry to increase the accuracy and efficiency of their Big Data analysis and analytics.

4. Ensemble methods: Combine multiple models to improve reliability and accuracy, such as a combination of statistical and machine learning methods.

It should be noted that effective anomaly detection methods/algorithms are based on two basic approaches to calculating anomalies in data:

– an approach based on distance in a multidimensional feature space: identifies anomalies based on distance measurements between data points, for example, Euclidean distance, Mahalanobis distance and NN, kNN, CBR methods, etc.;

– density-based approach: detects anomalies by examining the density of data points using methods such as local outlier factor (LOF) and isolation forest, etc.

Some practical applications of the anomaly detection mode in the practice of the food industry:

1) detection of fraud, abuses and errors in commercial activity: detection of intentional fraudulent and unintentional (erroneous) transactions, operations and influences in the sphere of procurement, in the sphere of sales activities of the food industry enterprise/holding;

2) cyber security (and, even, physical security monitoring): detection of network intrusions, malicious software and unauthorized access in cyber security, on the fly recognition of unusual actions, movements and movements of employees, third parties, etc.;

3) health care: monitoring of all types and formats of data about employees (and visitors) for early detection of diseases and abnormal conditions, infections;

4) controlling production processes and operations: detection of defects and violations in production processes, technological operations to ensure proactive control of quality and efficiency;

5) detection of fraud, abuse and errors in financial activities: detection of unusual trading patterns and market anomalies, fraudulent or erroneous financial transactions/operations in the field of financial management of the food industry enterprise/holding;

6) environmental monitoring: detection of unusual environmental conditions and/or detection of unusual conditions, processes and manifestations in the internal technological environment of a food industry enterprise.

The authors highlight the following challenges and problems when applying the anomaly detection regime in the practice of all levels of management of food industry enterprises:

1. High dimensionality of all types of data. High dimensionality refers to datasets with a large number of features (variables). The data anomaly detection mode should be activated secondarily, i.e. only AFTER performing the normal mode of detecting widespread, accurate and complete patterns/patterns in the data. But in the conditions of modern big data, even this primary mode of Data Mining is complicated due to "the curse of data dimensionality".

2. Imbalanced and underrepresentative input data: Anomalies occur quite rarely, which is why unbalanced and underrepresentative input data sets make them difficult to detect.

3. Evolution of anomaly patterns: anomalies can evolve and change over time, requiring adaptive methods to detect them in real time.

4. Noise and outliers: distinguishing real anomalies and threats from stochastic informational noise and outliers in the form of mechanical or technical errors in the data can be quite difficult, requiring reliable detection algorithms and, even, a wider selection of human expertise (which makes it difficult to detect anomalies on large data in AutoML mode).

5. Time-consuming and labor-intensive, availability and cost of a comprehensive interpretation of identified potential anomalies in the data of a food industry enterprise. In the above-mentioned conditions of large data of high dimensions, sometimes unbalanced and not sufficiently representative, in the conditions of the dynamics of changing patterns of anomalies, the presence of information noise – the evaluation, interpretation and formalization of detected anomalies requires more and more time and experience of human expertise of preliminary results of detecting anomalies by Data Mining methods. The experience of the authors of this study shows that among the previously detected anomalies in the data, only 3–5 % pass this stage of human interdisciplinary examination and will be included in the enterprise's corporate knowledge base.

Future directions and prospects for improving the anomaly detection regime for food industry enterprises:

1. Detection of anomalies in real time: deployment of real-time streaming data processing and analysis architectures (Apach Kafka and Apache Spark etc.) to detect anomalies as they occur, "on the fly".

2. Deep learning: extensive use of architectures and methods of deep machine learning (Deep ANN) on large unstructured data – to increase the accuracy and speed of pre-processing (feature extraction). Analysis and analysis of complex patterns of unstructured data. It provides for automatic recognition and subsequent classification of big data video, audio, and photo formats for the purpose of detecting and identifying shortages/defects of food products, monitoring the functioning of equipment and facilities, and safety purposes of the food industry enterprise.

3. Explained artificial intelligence: development of methods to improve the interpretation and transparency of detected anomaly models (even further attempts to explain/verbalize the "black boxes" of trained shallow artificial neural networks (Example of verbalization of such an implicit model (black box), obtained from configured and trained by the authors of shallow An ANN with three hidden layers is given above, in **Fig. 2.5**).

4. Adaptive algorithms: Comprehensive use of adaptive algorithms that can learn and evolve with changing data patterns.

5. Integration with big data: integration of anomaly detection methods with big data platforms for efficient processing of large and diverse data sets on the fly.

Summarizing the results of the author's research on effective methods and tactical techniques for detecting anomalies in business and technological data of food enterprises, it should be emphasized that the mode of detecting anomalies in Data Mining is a critical area that allows identifying rare but important events in all functional areas of management of food enterprises industry. Using a range of statistical, machine learning and ensemble ML techniques, it is the anomaly detection mode that uncovers unexpected new, hidden and valuable information/knowledge that can help prevent fraud, abuse, errors, negligence, improve security, improve production results and ensure more thorough controls product quality of food industry enterprises. Constant progress in the field of real-time analysis and analytics, the involvement of deep learning architectures and adaptive algorithms with elements of soft-computations – will further expand the possibilities and effect of the application of the anomaly detection mode in the future in the food industry.

Hybrid Methods and Algorithms of Data Mining to support decision-making in complex, interdisciplinary, complex, multidimensional tasks of enterprise management (on the example of the food industry)

Hybrid Data Mining methods/algorithms combine numerous separate autonomous computing methods (from such areas as: complex multidimensional DB queries, Statistical Analysis, ML, Mathematical Programming, Soft Computations) to increase the accuracy, efficiency, completeness and stability of the results of intelligent in-depth data analysis.

The main tasks of Data Mining hybridization: hybrid clustering, hybrid classification and hybrid predictive modeling.

Hybrid Data Mining methods/algorithms involve the combination of two or more algorithms/methods to create a more efficient, adaptive analytical tool in order to overcome the limitations of individual stand-alone algorithms/methods by synergistically using their strengths.

Hybrid approaches are particularly useful in highly complex, interdisciplinary industries/productions/technologies of the food industry (baby food production, organic food production, diet food production, etc.).

Today, in the era of big data and in anticipation of multimodal crisis phenomena, the greatest potential and hopes are placed on a hybrid approach to processing, analytics and data analysis, which in these dynamic and unstable conditions will be a powerful tool for solving complex problems that cannot be overcome by "pure" methods. classic BI approaches and methods. The implementation and universality of such hybrid approaches is limited by the requirements for the uniformity of data types/formats and the specific variability of decision-making at different levels of management in different functional departments of the food industry enterprise. That is why, for the effective use of hybrid Data Mining scenarios, a unified/standardized presentation of data, metadata and knowledge is an important condition.

Below, the results of the analysis of the three main tasks of Hybrid Data Mining will be presented:

1. Hybrid Regression – most often implemented through technologies of ensemble ML. Ensemble methods combine multiple learning algorithms to achieve better predictive performance than any single algorithm alone. Key ensemble techniques include:

– Bagging (Bootstrap Aggregating) Combines the predictions of multiple models trained on different subsets of the data to reduce variance and improve accuracy;

– Boosting: sequentially applies weak classifiers to the data, adjusting their weights based on the accuracy of previous classifiers to reduce bias and variance;

– Random Forests: an ensemble of decision trees where each tree is trained on a random subset of features, enhancing model robustness and accuracy.

2. Hybrid clustering algorithms integrate different clustering techniques to enhance the quality and interpretability of clusters. Common approaches include:

– Hierarchical-K-means Clustering: combines the hierarchical and K-means clustering methods to first identify broad clusters and then refine them into more precise sub-clusters;

– Density-Based and Partitioning Methods: merges density-based clustering (e.g., DBSCAN) with partitioning methods (e.g., K-means) to identify clusters of varying shapes and densities.

3. Hybrid Classification algorithms integrate multiple classification techniques to improve prediction accuracy and generalization. Examples include:

– Neural Network and Decision Tree Hybrid: combines the high accuracy of neural networks with the interpretability of decision trees;

– Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) Hybrid: utilizes the robustness of SVM for boundary detection and the simplicity of KNN for local classification.

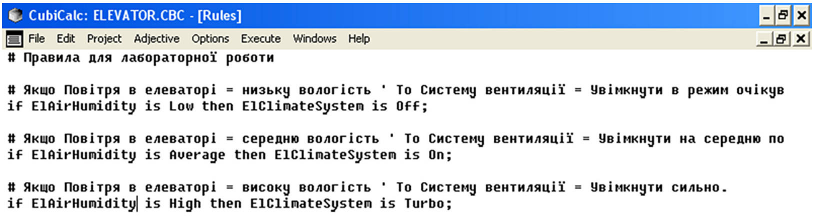
Currently, there is no precisely defined classification of hybrid information technologies in the scientific literature. Most often, options for the classification of hybrid systems are considered according to the following criteria: "homogeneity – heterogeneity", "hardness – softness", hybrid systems of the first order and hybrid systems of the second order, "equality – hierarchy".

It is worth noting that within Soft Computations, three aspects are improved (compared to Non-hybrid technologies): uncertainty management, retraining, adaptation.

Below, in **Fig. 2.12** – the given fragment of the base of fuzzy production rules for adaptive control of the auxiliary climate system for the small grain elevator/storage.

Using the above-mentioned fragment of the fuzzy production rules base for adaptive control of the grain elevator's auxiliary climate control system, the authors conducted a comparative simulation/modeling of the fuzzy technological control system and the classical one with hard technological rules – **Fig. 2.13**.

As it is possible to see from the figure, it can be claimed that the climate control system based on fuzzy logic adapts to changes more quickly and fluctuations in its operation are smaller. That is, the use of control systems based on fuzzy logic makes it possible not only to increase the level of flexibility and interpretability of control algorithms, but also to increase the resource of controlled systems and reduce their energy consumption.


CubiCalc: ELEVATOR.CBC - [Rules]
File Edit Project Adjective Options Execute Windows Help

```
# Правила для лабораторної роботи

# Якщо Повітря в елеваторі = низьку вологість ' То Систему вентиляції = Увімкнути в режим очікув
if ElAirHumidity is Low then ElClimateSystem is Off;

# Якщо Повітря в елеваторі = середню вологість ' То Систему вентиляції = Увімкнути на середню по
if ElAirHumidity is Average then ElClimateSystem is On;

# Якщо Повітря в елеваторі = високу вологість ' То Систему вентиляції = Увімкнути сильно.
if ElAirHumidity is High then ElClimateSystem is Turbo;
```

Fig. 2.12 The fragment of the fuzzy base of production rules for adaptive control of the auxiliary climate system for the small grain elevator
Note: developed by the authors

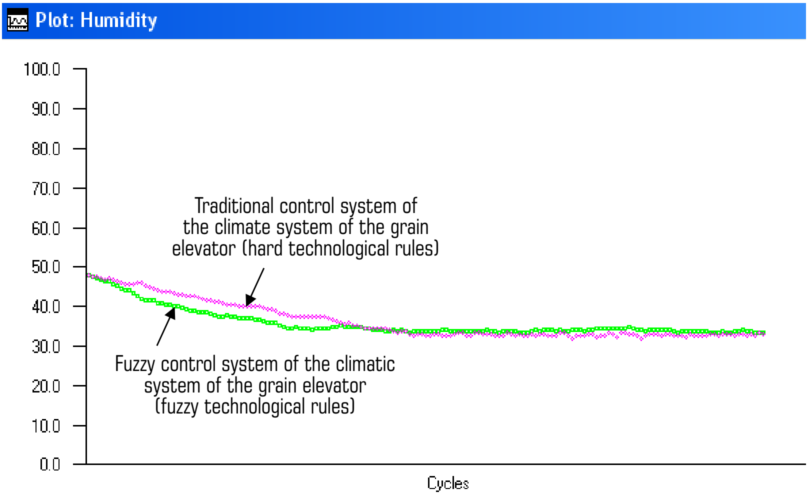


Fig. 2.13 Comparative simulation/modeling of the fuzzy technological climate control system and the classical one with hard technological rules (on the example of small grain elevator/storage)
Note: developed by the authors

Advantages and benefits of using Hybrid Data Mining:

– Robustness: hybrid approaches can handle a variety of data distributions and noise levels, making them more robust;

– Improved Accuracy: by combining different algorithms, hybrid methods often achieve higher predictive accuracy than individual methods;

– Scalability: many hybrid methods are designed to handle large datasets efficiently, making them suitable for big data applications;

– Flexibility: hybrid methods can be tailored to specific problems by selecting and combining appropriate algorithms.

Challenges and difficulties in using Hybrid Data Mining:

– Computational Cost: combining multiple algorithms may increase computational requirements;

– Complexity: hybrid methods can be more complex to implement and understand than single-method approaches;

– Integration: ensuring seamless integration of different algorithms and maintaining consistency can be difficult;

– Parameter Tuning: hybrid methods often require careful tuning of multiple parameters, which can be challenging.

Let's consider promising areas of Hybrid Data Mining:

– Automated Hybrid Systems: development of automated systems that can dynamically select and combine algorithms based on the data characteristics and problem requirements;

– Adaptive Hybrid Methods: creation of adaptive hybrid methods that can learn and evolve based on feedback and changing data patterns;

– Interdisciplinary Approaches: leveraging advances from other fields such as biology, physics, and social sciences to inspire novel hybrid Data Mining techniques;

– Interdisciplinary Approaches to hybridization: using algorithmic advances from other fields, such as biology, physics, and social sciences, to create new hybrid data analysis and modeling techniques (e.g., gravity search, ant colony method, etc.).

Hybrid intelligent systems, depending on the architecture, are divided into three main types:

– sequential hybrid systems (sequentially process data using different algorithms/methods). For example: the analytics subsystem of a food industry enterprise integrates and pre-processes input from users (for example, attracting posts and reactions on social networks with TextMining algorithms in Sentiment Analysis mode) before retraining a regression model for forecasting demand for new types of products;

– parallel hybrid systems (several intelligent methods/algorithms are simultaneously applied to the same data). For example: within the analytics subsystem of a food industry enterprise for optimal planning, dispatching and synchronization of loading and operation modes of logistics (transport and warehouse) equipment – can use parallel and competitive different (classical and alternative) methods of mathematical programming and different sets of optimization parameters of these methods with for the purpose of further comparison of the results of solving such a complex optimization problem (below, in **Fig. 2.14** – the author's example of solving such an optimization problem using a rather innovative (for the classic food industry management) method of genetic algorithms is shown);

– hierarchical hybrid intelligent systems (presuppose cybernetic organization of intelligent system components in a hierarchical structure, where higher-level components control lower-level components). For example: a distributed artificial intelligence system for 24/7/365 diagnostics of very different technological equipment and equipment of a food industry enterprise (to identify various types of current malfunctions and predict future ones).

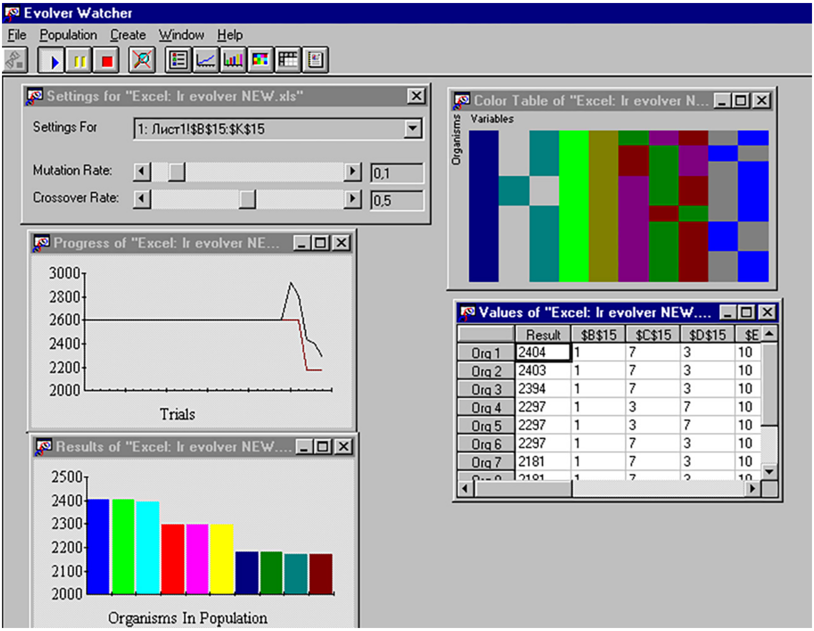


Fig. 2.14 The use of genetic algorithm to solve the complex problem of optimal planning, dispatching and synchronization of loading and operation modes of logistics (transport and warehouse) equipment on the example of a food industry enterprise (population size=10; mutation=0.1; crossover=0.5; stopping conditions=1000 trials)
Note: developed by the authors

Hybrid technologies, Data Mining algorithms/methods and their use scenarios represent an innovative approach to increasing the accuracy, completeness, robustness (i.e. integrated reliability and quality) of Data Mining (also and Data Science) of big data of food industry enterprises. By synergistically combining the strengths, advantages, or radically/absolutely different, or similar, related methods/algorithms of DM, overcoming/overcoming their certain shortcomings and limitations – it is hybrid technologies that will help solve complex problems of enterprises and companies (especially in the food industry) – much more effectively.

Future advances in automated and self-adaptive hybrid intelligent technologies and systems will provide additional potential for sustainability and competitiveness in various functional areas and management levels of food industry enterprises and companies (especially in the context of multimodal crises).

Data Mining for crisis management (in particular for enterprises and companies of the food industry) involves systematic analysis of all available internal and external data sets to identify early warning signs, data anomalies, monitor states, activities and events in real time 24/7/365 for further reactive generation and evaluation of the effectiveness of response strategies to identified crisis situations. In other words, crisis data mining is the process of finding, extracting and formalizing new, hidden and useful patterns/patterns from large data sets to aid in optimal crisis management and adaptive response. This process increases the resilience of the organization, helps reduce risks and supports informed decision-making during crises.

That is why, the urgent tasks of crisis data mining are a comprehensive study of the specifics of the methodology, effective tasks and challenges/cautions related to the intelligent analysis of crisis data, emphasizing the importance of crisis data mining for improving the decision-making process during various types of crises, including natural disasters, pandemics and socio-political disturbances, military actions.

The complexity and unpredictability of crises requires reliable and thorough approaches based on large (structured, semi-structured and even unstructured) data for effective detection of anomalies and threats.

Let's outline below the applied functional tasks/directions of using crisis datamining:

1. Natural disasters (earthquake prediction through seismic data analysis to predict potential earthquakes; flood monitoring through satellite imagery and weather data analysis to predict and monitor floods, etc.).

2. Pandemics (detection of disease outbreaks by searching for early signs of disease outbreaks in social networks and medical records; distribution of medical resources, i.e. optimization of the distribution of medical goods based on predictive models of the development of pandemics, etc.).

3. Socio-political crises (prediction of the development of socio-cultural and/or political conflicts through monitoring of social networks and news to assess and forecast the zones, causes and epicenter of the conflict; displacement tracking, i.e. using Data Mining to track the spontaneously displaced population and provide them with preventive assistance, etc.).

As such, crisis intelligence is a critical tool in modern crisis management, offering the potential to significantly improve response times and outcomes, and reduce potential harm. Continuous progress in data collection, their pre-stream processing, machine learning (including advanced anti-crisis analytics in the formed Data Lakes) and interdisciplinary interpretation are essential to overcome current challenges and maximize the effectiveness of crisis data mining.

Below, the results of the author's case studies will be briefly presented regarding the identified directions of future development and trends of Data Mining in particular (and sometimes Data Science in general) in the dynamic, unstable external conditions of food industry enterprises

Data mining involves the use of advanced computing techniques to extract meaningful information from complex data sets. As the volume and complexity of data continues to grow, the need for more sophisticated analysis tools becomes increasingly critical. In this subsection, the future directions and potential consequences of intelligent data analysis are considered, so the author highlights the following new trends and perspectives of Data Mining:

1. Automated machine learning (AutoML). Automated machine learning aims to automate the end-to-end process of applying machine learning to real-world problems. Future developments

in AutoML are expected to: increase accessibility, i.e. lower the barrier for non-experts by simplifying the process of creating, configuring and deploying machine learning models; efficiency gains: Streamline model selection, hyperparameter tuning, and feature development processes, reducing time and computational resources required.

2. Exploratory and in-depth real-time analysis and analytics. Real-time analytics involves continuous analysis of streaming data to provide instant insights and facilitate rapid decision-making. Future advances are likely to include: real-time decision-making – integrating real-time analytics with business operations to facilitate immediate response to emerging trends and anomalies; improve predictive maintenance: Use real-time data to predict equipment failures and optimize maintenance schedules in industrial applications.

3. Expanded and improved cross-industry, cross-task interpretation of Data Mining results. As machine learning models become more complex, ensuring their interpretability is critical. Future developments in model interpretation are expected to: increase transparency: develop methods to make complex models more understandable and transparent to stakeholders; increasing trust: Increase trust in and acceptance of AI systems by providing clear explanations of model solutions.

4. Integration of diverse and multi-format data sources of varying quality with the formation of specialized Data Lakes. Combining data from different sources can provide more comprehensive information. Future directions for such NON-homogeneous data integration include: data fusion: developing advanced data fusion techniques to seamlessly integrate structured and unstructured data from different domains; interdisciplinary collaboration: fostering collaboration across disciplines to leverage diverse data sets and generate holistic understanding.

5. Ethical considerations regarding total Data Mining. The ethical implications of data mining are becoming increasingly important. Key future areas include: Bias Mitigation: Developing methods to detect and mitigate biases in data and models to ensure equity and fairness; data privacy: improving data privacy protection mechanisms to protect sensitive information while enabling data analysis.

6. Potential systemic and total impact of innovative Data Mining. Advances in data mining are expected to have a profound impact on all areas of the food industry: ERP, CRM, MES, WMS, EAM, HRM; monitoring and proactive actions to ensure sanitary, biological and food safety; in the field of financial management – improved fraud/abuse/error detection and scenario-based long-term risk management (especially insurance risk management for food industry [35, 36]); investment management of food industry enterprises using more accurate and adaptive predicative models; in the field of production management – more accurate and systematic technological forecasting and better optimization of complex production processes and their components (see **Fig. 2.15** the author's example of the trained shallow ANN for forecasting product output (under the influence of external stochastic factors) depending on the number/ volume of 4 components: HCl, NH_3, H_2O and the amount of chemical reaction catalyst); proactive TQM in real time; optimal management of technical maintenance, current and capital repairs of equipment and facilities of food industry enterprises; etc.

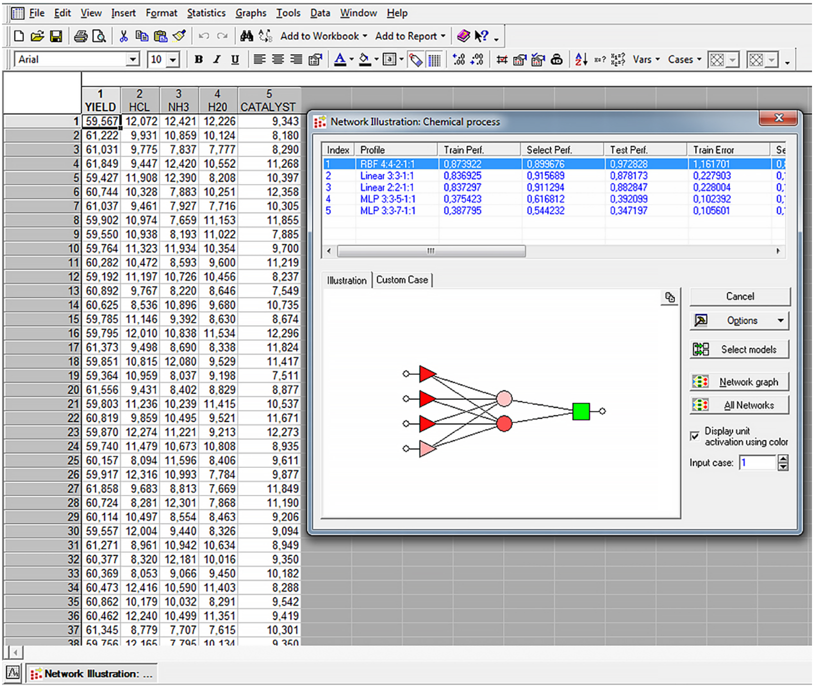


Fig. 2.15 The example of the configured and trained shallow ANN for forecasting the output of ultra-processed food product (under the conditions of the action of the complex of external stochastic factors) depending on the amount/volume of 4 components and the amount of catalyst for the chemical and technological reaction
Note: developed by the authors

Thus, the future prospects of Data Mining in particular (and sometimes Data Science in general) thanks to progress in all types of machine learning; technologies for effective organization of data integration, their processing, analysis and analytics in real time 24/7/365; complex in-depth expert interpretation of results; integration of data, knowledge and relevant ethical regulatory mechanisms – are relevant (for enterprises and food industry companies in particular). These areas of development and improvement of Data Mining in particular (and sometimes Data Science in general) should qualitatively improve the effectiveness of extracting valuable, new and hidden regularities/patterns/insights from large multidimensional low-quality data of all formats, will contribute to the adoption of operational, informed proactive decisions at all levels management, in all functional areas and all sectors of the food industry. Further continuation of research and interdisciplinary, inter-project cooperation will be necessary to realize the full potential of Data Mining in particular (and sometimes Data Science in general) to increase the integrated sustainability, efficiency, competitiveness of enterprises and companies in the food industry, especially in multimodal crisis external conditions.

CONCLUSIONS

Proposed in this publication scientific and practical applied solutions regarding Data Mining for enterprises and companies (on the example of food industry) involve the application of advanced cybernetic computing methods/algorithms, technological modes and scenarios (for integration, pre-processing, machine learning, testing and in-depth comprehensive interpretation of the results) of analysis and analytics of large structured and semi-structured data sets for training high-quality descriptive, predictive and even prescriptive models.

The proposed by authors multi-mode adaptive Data Mining synergistically combines in parallel and sequential scenarios: methods of preliminary EDA, statistical analysis methods, business intelligence methods, classical machine learning algorithms and architectures, advanced methods of testing and verification of the obtained results, methods of interdisciplinary empirical expert interpretation of results, knowledge engineering formats/techniques – for discovery/detection previously unknown, hidden and potentially useful patterns, relationships and trends.

The main methodological and technological goal of this developed methodology of multi-mode adaptive Data Mining for food industry enterprises is to increase the completeness (support) and accuracy of business and technical-technological modeling on all levels of management of food industry enterprises: strategic, tactical and operational.

By optimally configuring hyperparameters, parameters, algorithms/methods and architecture of multi-target and multidimensional explicit and implicit descriptive and predicative models, using high-performance hybrid parallel soft computing for machine learning – the improved methodology of multimode Data Mining (proposed by the authors) allows to find/detect/mine for new, useful, hidden corporate knowledge from previously collected, extracted, integrated Data Lakes, stimulating the overall efficiency, sustainability, and therefore competitiveness, of food industry enterprises at various organizational scales (from individual, craft productions to integrated international holdings) and in various food product groups and niches.

But, summarizing the above, it is worth emphasizing, that even a very experienced team of specialists/experts will not give an unequivocal effective recommendation on the first attempt, which Data Mining mode/algorithm/scenario, in which configuration of global hyperparameters and method-oriented parameters – will work most effectively (in a specific sector of the food industry, a specific region, for a specific enterprise at this time) without a set of additional studies, a set of experiments and a subsequent series of tests/trials.

What is why, the decision the problems (previously detected by the authors in BI practice of food industry) related to: multidimensionality of input big data; their verifiability and representativeness;

internal inconsistency, duplication and damage to data integrity; hidden presence of various types of data biases; informational noise and interferences; data outliers; missing and/or corrupted data) – are important determinant factors of the proposed concept/methodology of effective multi-mode Big Data Mining for food industry enterprises (for example, in production of vegetable proteins and synthetic meat, in production healthy snacking, in reducing harm of ultra-processed foods, in foods upcycling, for foods traceability, for farm-to-table and local sourcing initiatives, for circular economy practices in food waste reduction, for connected and transparent food supply chains, for net-zero & scope 3 emissions audit, for monitoring of crackdown on greenwashing etc.).

The Online Data Mining mode at food industry enterprises involves the concept, methodology and technology of online rapid data analysis and analytics in real time 24/7/365. This approach differs from traditional Data Mining modes, therefore it is relevant and crucial for those tasks and functions of food industry enterprise management that require maximum immediate (but often approximate) analysis of streaming data and operational response, emphasizing the relevant role of such mode in dynamic and time-sensitive applications in the food industry (in particular, in continuous production chemical and technological processes, production conveyors, etc.).

That is why the authors in this study especially paid considerable attention to the special mode of Anomaly & Fraud Detection Data Mining within the framework of proactive adaptive management in the food industry, which involves 24/7/365 monitoring/analysis/analytics of large data to identify anomalous/frauds/threatening patterns and trends, that help become the basis for proactive strategies, tactics and responses. That is this (Anomaly & Fraud Detection) Data Mining mode involves monitoring and detecting outliers, anomalies in streaming and batch, structured/semi-structured/unstructured big data regarding commercial, financial, technological, logistics, HR, marketing functions of food industry enterprise/corporation/holding in order to recognize maximum possible threats in real time mode. Using this mode, food industry enterprises can increase their robustness, sustainability by improving the efficiency of proactive decision-making in pre-crisis and crisis periods.

In addition, within the framework of this study, the methodology of effective Hybrid Data Mining is proposed (taking into account the specifics of enterprises and companies of the food industry), because it has been proven that it is the hybrid methods/algorithms of Data Mining indeed represent a powerful approach to increasing the accuracy, support, reliability and representativeness of Data Mining models trained from all types and formats of big data for a food industry enterprise. I.e. by combining the strengths of different methods, hybrid data mining methods can solve most complex enterprise/company problems more effectively than single-method/algorithm-oriented approaches. Future advances in automated and multilevel hybrid systems hold the potential for further expanding the capabilities and applied applications of Data Mining in complicated&complex functional areas and product sectors of food industry enterprises (in particular, especially in the modern high-tech food production: farm optimization with precision farming, regenerative & agro-ecological farming, automation in food storage warehouses and supermarkets etc.).

It is also worth highlighting the improved mode of Ad-hoc Data Mining for enterprises and companies of the food industry, which is a relevant and important mode in today's dynamic and stochastic (and therefore difficult to predict) macroeconomic and industry environment. It is this mode that offers the flexibility and speed needed to address the urgent and critical analytical needs of a food industry enterprise/company in today's environment.

Using the above-mentioned advances in Crisis Data Mining mode and, thanks to deep interdisciplinary expertise, enterprises and companies in the food industry will be able to use the full potential of Data Mining to achieve vertical and horizontal adaptive resilience and sustainability, even in times of crisis.

It is worth noting that by effectively using the fireproofed modes of Data Mining of a food industry enterprise, it is possible to reach a qualitatively top new level in Descriptive, Predictive and Prescriptive Analysis and Analytics of management system with help of such additional innovated technologies, as: automated machine learning (AutoML), streaming data analytics in real time, improved and in-depth cross-industry interpretation of results, innovative pre-processing and integration of various big data sources (in particular, using not only ETL or ELT, but also such technologies of integration and preliminary preparation of input data as: Change Data Capture, Data replication, Data virtualization, Stream Data Integration), new methods and scenarios of ensemble semi-supervised machine learning, multilevel and ternary hybridization of Data Science methods, distributed and multi-agent AI.

Additionally, the authors emphasize that Big Data Mining in the food industry is currently very promising in the context of R&D in the field of genetics, which involves the use of advanced ML methods/algorithms and ML technologies for the extraction/discovery/search/mining of new, useful, hidden genetic patterns from large sets of data related to food genetics (for example, analysis and analytics of the genetic composition of: agricultural crops, in livestock industry, in fishing industry etc.). Using the above proposed advanced innovative computational AI modes, researchers and industry professionals can extract valuable insights from genetic data, leading to innovations in crop and livestock production, food safety, and personalized nutrition. It is the above-proposed improved modes, scenarios and Data Mining technologies that should be widely used for the food industry, in particular, for: improving crop production, livestock breeding, food safety and quality, nutritional personalized genomics, effective&productive organic products etc. Therefore, data mining in the food industry, especially in the field of genetics, has a huge potential to improve productivity of food production, safety and quality, and of course, to reduce industry risks [37, 38] and, eventually, to improve the financial stability of food industry enterprises/companies.

It can not be argued, that in modern global macroeconomic standings (pre-crisis, crisis and post-crisis conditions of both regional food industries and the global world food industry; globalization and simultaneous very narrow specialization, often personalization of the food industry selected sectors; the need to take into account a huge amount of stream and packet information of various formats from different sources; the need for a quick adaptive optimal management actions in response to rapid changes in the global or regional market situation; unstable and difficult to predict dynamics of external influences: international, national, sectoral, local direct regulatory and indirect (from civil society and public organizations) regulations of the food industry standards) – deployment of the multi-mode adaptive Data Mining methodology proposed by the authors – will result in enterprises, companies and organizations of the food industry shall gained additional competitive advantages at the national state, regional, branch and corporate management levels.

Moreover, the proposed scientific and practical applied solutions, approaches, technologies, modes, configurations, settings (regarding High Dimensionality of Input Data, online Data Mining, ad-hoc Data Mining, hybrid Data Mining, crisis Data Mining, Data Mining for Anomaly & Fraud Detection) will be also particularly effective for enterprises and companies in countries and regions, in that industries (i.e., not only in the food industry) – where managerial decision-making requires complex expert analysis, is associated with significant capital risky investments, there are many branches of scenarios and nodes of risky decision-making, difficult-to-forecast negative effects of stochastic and dynamic external factors are possible, there are open uncertainties situations (probabilities and fuzziness), specific industry/sectoral/product risks.

REFERENCES

1. Ostapenko, T., Onopriienko, D., Hrashchenko, I., Palyvoda, O., Krasniuk, S., Danilova, E. (2022). Research of impact of nanoeconomics on the national economic system development. Innovative development of national economies. Kharkiv: PC TECHNOLOGY CENTER, 46–70. doi: https://doi.org/10.15587/978-617-7319-64-0.ch2
2. Palyvoda, O., Karpenko, O., Bondarenko, O., Bonyar, S., Bikfalvi, A. (2018). Influence of network organizational structures on innovation activity of industrial enterprises. Problems and Perspectives in Management, 16 (3), 174–188. https://doi.org/10.21511/ppm.16(3).2018.14
3. Illiashenko, S., Bilovodska, O., Tsalko, T., Tomchuk, O., Nevmerzhytska, S., Buhas, N. (2022). Opportunities, Threats and Risks of Implementation the Innovative Business Management Technologies in the Post-Pandemic Period COVID-19. WSEAS transactions on business and economics, 19, 1215–1229. https://doi.org/10.37394/23207.2022.19.107
4. Krasnyuk, M., Nevmerzhytska, S., Tsalko, T. (2024). Processing, analysis & analytics of big data for the innovative management. Grail of Science, 38, 75–83. https://doi.org/10.36074/grail-of-science.12.04.2024.011
5. Krasnyuk, M., Elishys, D. (2024). Perspectives and problems of big data analysis analytics for effective marketing of tourism industry. Science and Technology Today, 4 (32), 833–857. https://doi.org/10.52058/2786-6025-2024-4(32)-833-857
6. Krasnyuk, M. T., Hrashchenko, I. S., Kustarovskiy, O. D., Krasniuk, S. O. (2018). Methodology of effective application of Big Data and Data Mining technologies as an important anti-crisis component of the complex policy of logistic business optimization. Economies' Horizons, 3 (6), 121–136. https://doi.org/10.31499/2616-5236.3(6).2018.156317
7. Tsalko, T., Nevmerzhytska, S., Didenko, Ye., Kharchenko, T., Bondarenko, S. (2020). Optimization of goods implementation on the basis of development of business process re-engineering. Journal of Management Information and Decision Sciences, 23 (2). Available at: https://

www.abacademies.org/articles/optimization-of-goods-implementation-on-the-basis-of-development-of-business-process-reengineering-9174.html

8. Krasnyuk, M., Hrashchenko, I., Goncharenko, S., Krasniuk, S., Kulynych, Y. (2023). Intelligent management of an innovative oil and gas producing company under conditions of the modern system crisis. Access Journal – Access to Science, Business, Innovation in the Digital Economy, 4 (3), 352–374. https://doi.org/10.46656/access.2023.4.3(2)
9. Turlakova, S., Bohdan, B. (2023). Artificial intelligence tools for managing the behavior of economic agents at micro level. Neuro-Fuzzy Modeling Techniques in Economics, 12, 3–39. https://doi.org/10.33111/nfmte.2023.003
10. Krasnyuk, M. T., Hrashchenko, I. S., Kustarovskiy, O. D., Krasniuk, S. O. (2018). Methodology of effective application of Big Data and Data Mining technologies as an important anti-crisis component of the complex policy of logistic business optimization. Economies' Horizons, 3 (6), 121–136. https://doi.org/10.31499/2616-5236.3(6).2018.156317
11. Y. Krasnyuk, M., Kulynych, Y., Krasniuk, S. (2022). Knowledge discovery and data mining of structured and unstructured business data: problems and prospects of implementation and adaptation in crisis conditions. Grail of Science, 12-13, 63–70. https://doi.org/10.36074/grail-of-science.29.04.2022.006
12. Tereshchenko, S., Kulinich, T., Matienko, V., Tymchyna, Y., Nevmerzhytska, S., Ievseitseva, O. (2023). Financial and marketing innovative management of ecological agribusiness. Financial and Credit Activity Problems of Theory and Practice, 6 (53), 487–500. https://doi.org/10.55643/fcaptp.6.53.2023.4255
13. Ishchejkin, T., Liulka, V., Dovbush, V., Zaritska, N., Puzyrova, P., Tsalko, T. et al. (2022). Information subsystem of agri-food enterprise management in the context of digitalization: the problem of digital maturity. Journal of Hygienic Engineering and Design, 38, 243–252. Available at: https://keypublishing.org/jhed/jhed-volumes/jhed-volume-38-fpp-18-tymur-ishchejkin-viktoriia-liulka-vita-dovbush-nadiia-zaritska-polina-puzyrova-tetiana-tsalko-svitlana-nevmerzhytska-yuliia-rusina-olena-nyshenko-svitlana-bebko/
14. Derbentsev, V. D., Bezkorovainyi, V. S., Matviychuk, A. V., Hrabariev, A. V., Hostryk, A. M. (2023). A comparative study of deep learning models for sentiment analysis of social media texts. CEUR Workshop Proceedings, 3465, 168–188.
15. Derbentsev, V., Kibalnyk, L., Radzihovska, Yu. (2019). Modelling multifractal properties of cryptocurrency market. Periodicals of Engineering and Natural Sciences (PEN), 7 (2), 690–701. https://doi.org/10.21533/pen.v7i2.559
16. Derbentsev, V., Bezkorovainyi, V., Akhmedov, R. (2020). Machine learning approach of analysis of emotional polarity of electronic social media. Neuro-Fuzzy Modeling Techniques in Economics, 9, 95–137. https://doi.org/10.33111/nfmte.2020.095
17. Derbentsev, V., Velykoivanenko, H., Datsenko, N. (2019). Machine learning approach for forecasting cryptocurrencies time series. Neuro-Fuzzy Modeling Techniques in Economics, 8, 65–93. https://doi.org/10.33111/nfmte.2019.065

18. Carneiro, F., Miguéis, V. (2021). Applying Data Mining Techniques and Analytic Hierarchy Process to the Food Industry: Estimating Customer Lifetime Value. Proceedings of the International Conference on Industrial Engineering and Operations Management. Kotler, 2000, 266–277. https://doi.org/10.46254/sa02.20210125
19. Vlontzos, G., Pardalos, P. M. (2017). Data mining and optimisation issues in the food industry. International Journal of Sustainable Agricultural Management and Informatics, 3 (1), 44–64. https://doi.org/10.1504/ijsami.2017.082921
20. Hajiha Hajiha, A., Radfar, R., Malayeri, S. S. (2011). Data mining application for customer segmentation based on loyalty: An iranian food industry case study. 2011 IEEE International Conference on Industrial Engineering and Engineering Management, 504–508. https://doi.org/10.1109/ieem.2011.6117968
21. Shekhar, H., Sharma, A. (2023). Global Food Production and Distribution Analysis using Data Mining and Unsupervised Learning. Recent Advances in Food, Nutrition & Agriculture, 14 (1), 57–70. https://doi.org/10.2174/2772574x14666230126095121
22. Massaro, A., Dipierro, G., Saponaro, A., Galiano, A. (2020). Data Mining Applied in Food Trade Network. International Journal of Artificial Intelligence & Applications, 11 (2), 15–35. https://doi.org/10.5121/ijaia.2020.11202
23. Ting, S. L., Tse, Y. K., Ho, G. T. S., Chung, S. H., Pang, G. (2014). Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry. International Journal of Production Economics, 152, 200–209. https://doi.org/10.1016/j.ijpe.2013.12.010
24. Holimchayachotikul, P., Phanruangrong, N. (2010). A Framework for Modeling Efficient Demand Forecasting Using Data Mining in Supply Chain of Food Products Export Industry. Proceedings of the 6th CIRP-Sponsored International Conference on Digital Enterprise Technology. Berlin Heidelberg: Springer, 1387–1397. https://doi.org/10.1007/978-3-642-10430-5_106
25. Hayashi, Y., Hsieh, M.-H., Setiono, R. (2009). Predicting consumer preference for fast-food franchises: a data mining approach. Journal of the Operational Research Society, 60 (9), 1221–1229. https://doi.org/10.1057/palgrave.jors.2602646
26. Taghi Livari, R., Zarrin Ghalam, N. (2021). Customers grouping using data mining techniques in the food distribution industry (a case study). SRPH Journal of Applied management and Agile Organisation, 3 (1), 1–8. https://doi.org/10.47176/sjamao.3.1.1
27. Liu, L.-M., Bhattacharyya, S., Sclove, S. L., Chen, R., Lattyak, W. J. (2001). Data mining on time series: an illustration using fast-food restaurant franchise data. Computational Statistics & Data Analysis, 37 (4), 455–476. https://doi.org/10.1016/s0167-9473(01)00014-7
28. Liao, C.-W. (2009). Consumers' behavior in the food and beverage industry through data mining. Journal of Information and Optimization Sciences, 30 (4), 855–868. https://doi.org/10.1080/02522667.2009.10699915
29. Jiménez-Carvelo, A. M., González-Casado, A., Bagur-González, M. G., Cuadros-Rodríguez, L. (2019). Alternative data mining/machine learning methods for the analytical evaluation of

food quality and authenticity – A review. Food Research International, 122, 25–39. https://doi.org/10.1016/j.foodres.2019.03.063

30. Krasnyuk, M., Krasniuk, S. (2020). Comparative characteristics of machine learning for predicative financial modelling. Scientific bulletin ΛΌΓΟΣ, 55–57. https://doi.org/10.36074/26.06.2020.v1.21
31. Krasnyuk, M., Tkalenko, A., Krasniuk, S. (2021). Results of analysis of machine learning practice for training effective model of bankruptcy forecasting in emerging markets. Scientific bulletin ΛΌΓΟΣ. https://doi.org/10.36074/logos-09.04.2021.v1.07
32. Krasnyuk, M., Krasniuk, S. (2021). Modern practice of machine learning in the aviation transport industry. Scientific bulletin ΛΌΓΟΣ. https://doi.org/10.36074/logos-30.04.2021.v1.63
33. Krasnyuk, M., Krasniuk, S., Goncharenko, S., Roienko, L., Denysenko, V., Liubymova, N. (2023). Features, problems and prospects of the application of deep machine learning in linguistics. Bulletin of Science and Education, 11 (17), 19–34. Available at: http://perspectives.pp.ua/index.php/vno/article/view/7746/7791
34. Krasnyuk, M., Hrashchenko, I., Goncharenko, S., Krasniuk, S. (2022). Hybrid application of decision trees, fuzzy logic and production rules for supporting investment decision making (on the example of an oil and gas producing company). Access Journal – Access to Science, Business, Innovation in the Digital Economy, 3 (3), 278–291. https://doi.org/10.46656/access.2022.3.3(7)
35. Shirinyan, L., Arych, M. (2019). Impact of the insurance costs on the competitiveness of food industry enterprises of Ukraine in the context of the food market security. Ukrainian Food Journal, 8 (2), 368–385. https://doi.org/10.24263/2304-974x-2019-8-2-15
36. Shirinyan, L., Arych, M., Rohanova, H. (2021). Influence of Insurance on Competitiveness of Food Enterprises in Ukraine. Management Theory and Studies for Rural Business and Infrastructure Development, 43(1), 5–12. https://doi.org/10.15544/mts.2021.01
37. Arych, M., Levon, M., Arych, H., Kulynych, Y. (2021). Genetic testing in insurance: implications for non-EU insurance markets as a part of the European integration process. On-line Journal Modelling the New Europe, 36, 25–52. https://doi.org/10.24193/ojmne.2021.36.02
38. Arych, M., Kuievda, I., Dvořák, M., Hinke, J. (2023). The farming costs (including insurance) of the agricultural holdings in the European Union. Journal of International Studies, 16 (1), 191–205. https://doi.org/10.14254/2071-8330.2023/16-1/13