МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ОДЕСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ ХАРЧОВИХ ТЕХНОЛОГІЙ

XI МІЖНАРОДНА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І АВТОМАТИЗАЦІЯ – 2018

Збірник доповідей

Частина I

Одеса,
4-5 жовтня 2018

ЗМІСТ

УДК 519.24:007.51:004.58

ZHYGAILO A.M., DETS D.V.
ONAFT (UKRAINE)

DATA MINING TECHNOLOGIES IN AUTOMATED SUPPORTING SYSTEMS FOR DECISION MAKING

Abstract. *To manage complex objects, various automated systems with support of making decisions are used, which accumulate large volumes of information in their databases. The development and implementation of new algorithms based on the methods of data mining, which, taking into account management objectives and object management information, could provide the choice of the most effective management solution, which is a very topical task.*

Formulation of the problem. The rapid growth of humanity's potential is due to the deep penetration of information technology, which contributes to improving the efficiency of management of various technological and business processes. Taking into account modern ideas, any production environment can be considered as two main components: cybernetic and physical. The stable development of cyber-physics systems is an organizational and technical concept that is based on the integration of the physical and informational space and defines the revolutionary aspect of Industry 4.0.

Within these two realities, products are manufactured which corresponds to certain descriptions, specifications, recipes, regulations, procedures, and structural schemes. All of them together can be considered as a model of the process of creating one or another product. Data of this model are used in the information (cybernetic) environment for planning production, monitoring, programming of controllers, etc. In the course of production in the cyber environment, information actions (operations) with different objects are used. Since the cyber environment is identical to the physical environment, physical actions (operations) are performed in parallel with the same objects. This explains the fact that it is necessary to change the methods of organization of production in the context of new technological changes. At the same time, an adequate verification of the consequences of these changes can take place in a virtual environment. And this, in turn, can provide a competitive advantage over the time of execution of works and spent resources.

Tasks to solve. Improvement of cyber-physics systems provides the opportunity to solve problems of modeling production based on reliable data, obtained in real time. Achieving the results will help to make effective managerial decisions at all levels and avoid critical human-factor mistakes. Therefore, it is necessary to develop and implement topical methods of data analysis and machine learning that will allow the creation and implementation of higher level automation systems that can include descriptive, diagnostic and predictive analytics.

The essence of the study. As the first example, the main section of the bakery industry, where the dough process is taking place, was considered. From this point the quality of bakery products begins. The main unit of this process is a kneading machine. Many programs can be used to dip a dough on such a machine. They take into account different properties of raw materials, among which the flour is the main indicators: gauge deformation meter, amount of raw gluten, number of falls, humidity, whiteness of flour. The chosen program should help to achieve the highest quality of the structural properties of the dough to ensure a stable high quality bakery products and minimize the cost of raw materials. When deciding on the choice of the required program, a human factor is very necessary. And this affects its accuracy and creates dependence of economic results of the enterprise on the actions of one employee. So there is a need for intellectual support or full automation of the process of choosing a dough machine program. For the conduct of intellectual analysis, namely the cluster, flour quality indicators were provided from the laboratory of the bakery enterprise. The analysis of the flour should allow the selection of optimally constructed technological steps for the production of quality products.

Information that can be obtained as a result of cluster analysis will greatly help improve the quality of bakery products by developing for each group of indicators of quality, a certain technology of kneading dough. Such information can not be obtained without the help of a software tool that performs intellectual analysis.

A new software module, Zhy & Bor [1], was developed to provide a stable choice of the most effective options for management solutions in the operation of complex objects. He implements the advanced data clustering procedure based on the k-means method, and his distinctive feature is the algo-

rithms for automatically calculating the number of clusters and determining the positions of the initial centers of clusters (centroids) (Fig. 1,2).

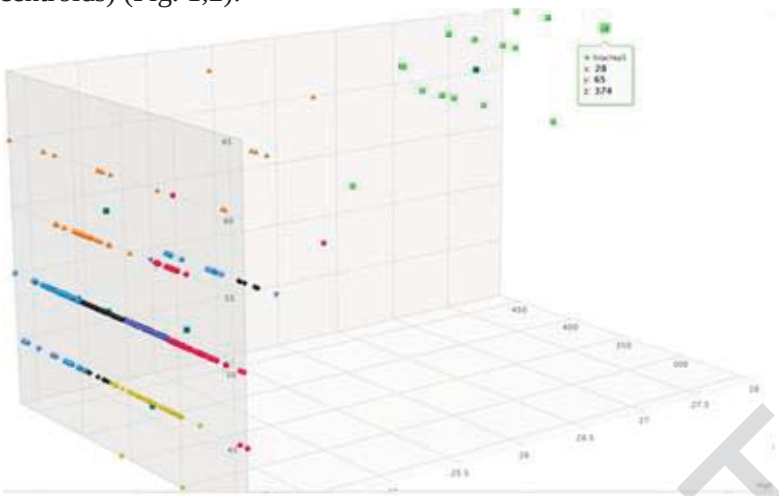

Fig. 1. - 3D model of clustering result.

Кл. 6				Кл. 5				Кл. 7				Кл. 4			
Инф	Кол-во сыр. клейк. %	Кач-во сыр. клейк., ИДК, у.е.	Число падения, с	Инф	Кол-во сыр. клейк. %	Кач-во сыр. клейк., ИДК, у.е.	Число падения, с	Инф	Кол-во сыр. клейк. %	Кач-во сыр. клейк., ИДК, у.е.	Число падения, с	Инф	Кол-во сыр. клейк. %	Кач-во сыр. клейк., ИДК, у.е.	Число падения с
1	24	45	402	6	24	50	390	31	27	65	354	33	24	60	373
2	24	45	409	18	25	50	378	32	27	65	391	34	24	60	369
3	24	45	402	19	25	50	363	37	27	60	421	35	24	60	373
4	24	50	413	20	25	50	363	40	27	60	441	36	24	60	369
5	24	45	410	21	25	50	373	45	27	65	415	38	24	60	374
7	25	50	449	22	25	50	373	49	28	65	457	39	24	60	374
8	25	50	449	23	24	50	364	50	28	65	426	41	24	55	373

Fig.2 - The result of clustering in tabular form.

The second example of demonstrating the benefits of developing a software module was the implementation of a full-featured experiment (FFE) [2] to optimize the prescription composition of a probiotic cosmetic product for tonicity of oily skin; development of technologies of serum fermented beverage and serum jelly with the use of herbal extract of the aronia of an amphibious sanitary purpose; development of technologies of dairy products for pregnant women feeding, enriched with extracts of echinacea, cranberries, sea buckthorn, etc.

Since the description of the behavior and prediction of the state of objects is complicated by the lack of a sufficient number of mathematical models, in some cases, a simple way out of the situation is the use of regression patterns. The advantage is the ability to describe any object with full compliance with the algorithm of the active experiment. Experiments, as a rule, are multifactorial and related to the optimization of the quality of materials, the search for optimal conditions for conducting technological processes, the development of the most rational structures of equipment, etc. FFE is the most easily implemented among many methods of active experiment. When using FFE, the goal of a researcher is to get a linear mathematical model of the process and determine with the subsequent strategy of conducting the experiment. Finding a model of FFE consists of: a) planning and conducting an experiment; b) check reproducibility (homogeneity of sample dispersions); c) obtaining a mathematical model of an object verifying the statistical significance of the sample regression coefficients; e) checking the adequacy of the mathematical description.

FFE was implemented as a software module, which allows you to receive mathematical models (response functions), using from 2 to 5 factors and from 2 to 10 repetitions of the experiment. A greater number of factors is not feasible, because already at 5 factors and 10 repetitions it is necessary to conduct 320 experiments. In fig. 3-7 the results of the module work are presented on the example of optimization of the process of extraction of biologically active substances from the fruits of the aronia of black-and-white cottage cheese. The indicator of biological activity depends on two factors: the temperature of cheese whey and the duration of extraction.

Факторы	Уровень			Интервал варьирования	Размерность	Зависимость кодированной переменной от натуральной
	Верхний	Нижний	Основной (центр)			
Z_1	60	40	50	10	°C	$X_1=(Z_1-50)/10$
Z_2	50	30	40	10	мин.	$X_2=(Z_2-40)/10$

Fig. 3. Terms of conducting FFE

Номер эксперимента	Порядок варьирования факторов		Порядок варьирования взаимодействия	Результаты опытов		$\overline{Y_u}$
	X_1	X_2	X_1X_2	Y_{1u}	Y_{2u}	
1	-	-	+	80.1	79.9	80
2	+	-	-	71.6	71.5	71.55
3	-	+	-	51.6	51.8	51.7
4	+	+	+	49.9	50.1	50

Fig. 4. The matrix of the planning of the experiment and the results of experiments.

Дисперсия воспроизводимости: $S^2_{\{y\}} \approx 0.008$
Среднее квадратическое отклонение коэффициентов: $S_{коэф} \approx 0.04472$
Из таблиц распределения Стьюдента по числу степеней свободы = 4,
при уровне значимости $\alpha = 0.05$ находим $t_{кр} = 2.77645$
Сравниваем полученное значение $\varepsilon_{bi} = t_{кр} \cdot S_{коэф} \approx 0.12416$
с коэффициентами уравнения регрессии $|b_i| > \varepsilon_{bi}$
Все коэффициенты не удовлетворяющие данному условию выделяются .

Fig. 5. Verification of reproducibility and statistical significance.

$$Y(X_1,X_2) = 63.31 - 2.54 \cdot X_1 - 12.46 \cdot X_2 + 1.69 \cdot X_1 \cdot X_2$$

Адекватность модели вцелом будем определять по критерию Фишера $F_{расч} < F_{таб}$
Экспериментальное значение F-критерия (критерия Фишера) равно: $F_{расч} = S^2_{ад} / S^2_{\{y\}} \approx 0.0125$,
где дисперсия адекватности модели (остаточная дисперсия) $S^2_{ад} = 0.0001$
Из таблиц распределения Фишера по числу степеней свободы: $f_1 = 1$ и $f_2 = 4$,
при уровне значимости $\alpha = 0.05$ находим табличное значение критерия Фишера: $F_{таб} = 7.71$

Fig. 6. Obtaining the mathematical model of the object and checking its adequacy.

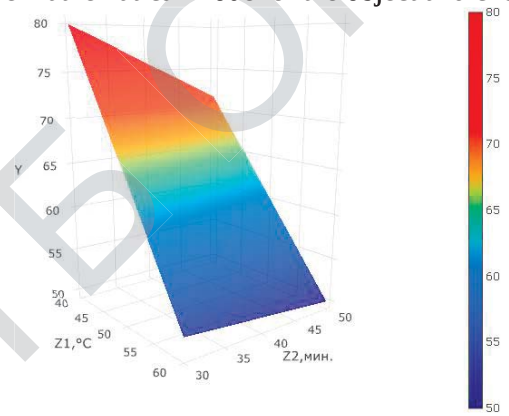

Fig. 7. Dependence of the indicator of biological activity.

Conclusions. The first example creates the preconditions for the development of a management algorithm, which, after entering the data of flour quality indicators, can automatically implement a certain technology of kneading dough. The control actions will be performed due to changes in the direction of the spiral, the direction of movement, the speed of the spiral, the speed of the bowl, the time of kneading, the temperature of the leaven.

In the second example, the analysis of the results of the data shows that the optimal parameters of extraction of biologically active substances from the fruits of the aronia are: the extraction temperature is 40 °C, the time is 30 minutes. The indicator of biological activity is 80 units. act.

LIST OF REFERENCES

1. О.М. Жигайло, В.В. Борис. Кластерний аналіз даних в автоматизованих системах простежуваності//Автоматизація технологічних та бізнес-процесів. –О.,2018. – Том 10, №1 – С. 39-46.

2. Myers R., Montgomery D., Anderson-Cook C. Response surface methodology: process and product optimization using designed experiments. – 4th ed. Hoboken, New Jersey : John Wiley & Sons. 2016. 825p.

XI МІЖНАРОДНА НАУКОВО-ПРАКТИЧНА КОНФЕРЕНЦІЯ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ І АВТОМАТИЗАЦІЯ – 2018

ОДЕСА
4 – 5 ЖОВТНЯ, 2018

Збірник включає доповіді учасників XI Міжнародної науково-практичної конференції «Інформаційні технології і автоматизація – 2018»

Редакційна колегія: Котлик С.В., Хобін В.А.

Комп'ютерний набір і верстка: Шамрай О.А.

Відповідальний за випуск: Котлик С.В.

ДЛЯ НОТАТОК