UDC664.6+004.896

PERSPECTIVES OF USING ARTIFICIAL INTELLIGENCE ELEMENTS IN BREAD BAKING

V. Makhynko, L. Makhynko
National University of Food Technologies
O. Kozyr
Baking Enzymes, Lallemand

Key words:
Bread
Recipe design
Artificial intelligence
Machine learning
Baking industry

Article history:
Received 02.03.2023
Received in revised form 17.03.2023
Accepted 05.04.2023

Corresponding author:
V. Makhynko
E-mail:
mavam78@gmail.com

ABSTRACT

Enterprises of bread baking sector produce traditional everyday products and, at the same time, this sector remains one of the most conservative ones. There are always new types of raw materials and additives on the market, but new recipes are developed traditionally based on the results of test baking. However, a technologically justified solution is not always optimal regarding nutritional value.

The purpose of this work is consideration of the possibilities and prospects of using elements of artificial intelligence (AI) in bread baking based on examples from other sectors of the food industry and study the AI's ability to analyze and evaluate (categorise) existing recipes of bread products.

The initial stage of AI use is machine learning (ML), but today there are no unified electronic databases that can be used to fulfil the task. We took approved collections of recipes as the basis for filling such databases. In addition, technologically acceptable variations of the main recipe components and their mutual substitution were carried out. The base was analyzed using Microsoft Azure Machine Learning service and Google Cloud Machine Learning Engine.

Normalised database with 5,000 variants of recipes of bakery products that can be manufactured in conditions of both large industrial companies and small bakeries were created. The ML showed that the system effectively determines the main components of recipes and can independently, with high accuracy, classify recipes entered by the user into categories "bread" or "enriched bread".

The work showed the possibility and viability of using AI elements in the baking industry. The database can be used to design new bakery products with specified recipe composition and (after improvement) to model applications with specified chemical composition or biological value.

DOI: 10.24263/2225-2924-2023-29-2-17

ПЕРСПЕКТИВИ ВИКОРИСТАННЯ ЕЛЕМЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ В ХЛІБОПЕЧЕННІ

В. М. Махинько, **Л. В. Махинько**
Національний університет харчових технологій
О. М. Козир
Baking Enzymes, Lallemand

Хлібопекарська галузь, випускаючи продукти традиційного повсякденного вживання, водночас є однією з найбільш консервативних. На ринку постійно з'являються нові види сировини та добавок, але розроблення нових рецептур відбувається традиційним способом за результатами пробних випікань. При цьому технологічно виправдане рішення не завжди є оптимальним з позиції харчової цінності (і навпаки).

У статті розглянуто можливості та перспективи використання елементів штучного інтелекту (ШІ) в хлібопеченні на прикладі інших галузей харчової промисловості, вивчено спроможності ШІ до аналізу та оцінювання (категоризації) вже існуючих рецептур хлібних виробів.

Початковим етапом використання ШІ є проведення машинного навчання (МН), але на сьогодні відсутні уніфіковані електронні бази даних, що можуть бути використані для вирішення поставленого завдання. За основу для їх наповнення було взято затверджені збірники рецептур. Додатково провели технологічно допустиме варіювання основних рецептурних компонентів та їх взаємозаміну. Аналіз створеної бази здійснювали за допомогою Microsoft Azure Machine Learning та Google Cloud Machine Learning Engine.

Створено електронну нормалізовану базу з близько 5 тис. варіантів рецептур хлібобулочних виробів, що можуть виготовлятися в умовах як великих промислових підприємств, так і пекарень. Результати МН показали, що система ефективно визначає основні компоненти рецептури, а також здатна самостійно з високою точністю класифікувати введені користувачем рецептури за категоріями «хліб» чи «здобні вироби».

Показано можливість і перспективність використання елементів ШІ у хлібопекарській галузі. Створена база може бути використана для проєктування нових хлібобулочних виробів із заданим рецептурним складом, а також (після доопрацювання) — для моделювання виробів, що матимуть заданий хімічний склад чи біологічну цінність.

Ключові слова: *хліб, проєктування рецептур, штучний інтелект, машинне навчання, хлібопекарська промисловість.*

Formulation of the problem. Food crisis is becoming a reality due to global changes in recent years: growth of Earth's population, recurrent epidemics and pandemics, military conflicts, and full-scale wars.

This situation requires new solutions at all stages of the food security — from planning and solving agricultural tasks minimizing the amount of food waste or designing mechanisms for its recycling. Extensive development that envisages only growth of planted areas and increased capacities of various branches of the food industry is impossible in current conditions and inexpedient considering existing technologies and means of production. So, methods to improve the intensity of food production, particularly by using various computer technologies, are being searched for. One of the most promising solutions in this area is the involvement of artificial intelligence (AI).

Analysis of recent research and publications. The basis of AI is the development of instrumental and software solutions to copy or simulate the behavior of human intellect. Development of this industry began in the middle of the last century. Initially, the goal was only to model the process of human learning and thinking. However, it was later found that AI can be learned in ways other than humans e.g. through classification of well-structured, specific and large datasets of tabular, text or visual information. This finally pushed the development of an essential component of AI — machine learning (ML), which analyzes and generalizes data sets and establishes specific dependencies, with the possibility of future predicting or generation of new elements of this set. The result was an autonomously functioning system that can fulfil tasks in a variety of industries and spheres of modern life (Chidinma-Mary-Agbai, 2020; Kumar, Rawat, Mohd, & Husain, 2021; Matthew, Omobayode, & Sarhan, 2020):

- Natural Language Processing — statistical and semantic analysis of natural languages to gain the ability to recognize language and analyze the received text, with its subsequent translation into other languages or forms of representation;
- Computer Vision — obtaining, recording, and analyzing visual information for its classification or use in automation systems (for example, object recognition and sorting, autonomous navigation, etc.);
- Robotics — the possibility to reproduce human actions with greater speed/accuracy, especially in spheres with a significant number of monotonous, repetitive operations;
- Data Mining — analysis of large data sets (including updatable) for both accumulation of statistical data, and establishment of hidden patterns that can be unnoticeable for human analysis.

Most of these tasks are solved at various stages of the food products creation, so as the availability of AI tools increases, so do the options for its application in agriculture and the food industry (Fisher, 2016; Jermsurawong, & Habash, 2015; Steluti, Junior & Marchioni, 2020). Without dwelling on the first stages, there were noted such compelling examples of AI use as satellite monitoring of planted areas' condition, the presence of the need for irrigation or use of means of plant protection, mapping of yield, and even cow care (Soltani-Fesaghandis, & Pooya, 2018; Kakani, Nguyen, Kumar, Kim, & Pasupuleti, 2020).The next stage where AI can be used is the analysis of harvested crops, with an assessment of the degree of their damage by pests, mechanical damage, the presence of dirt, pollution, or foreign impurities. A combination of Computer Vision and Robotics elements can significantly speed up the product sorting process, reducing the volume of low-skilled manual labor while simultaneously accumulating analytical information for further interaction with suppliers (Fisher, 2016; Jermsak, & Nizar, 2015; Steluti, Junior, & Marchioni, 2020). As for food industry companies, AI systems can be used there to analyze the sanitary condition of equipment (optical and ultrasonic detection of dirt and pollution) and to monitor workers' compliance with personal hygiene requirements at the workplace (Soltani-Fesaghandis, & Pooya, 2018; Vijay, Van Huan, Basivi, Hakil, &Visweswara, 2020).

In conditions of a highly competitive medium, one of the most promising spheres of AI use in the food industry is the study of market consumer preferences. After all, it is known that only 20% of all novelties can gain a foothold on the market for sufficient time to compensate for their development's organizational and technological costs incurred. The 80% losses can be minimized by conducting various online and offline surveys, with subsequent processing of results obtained by AI methods. The outcome will ensure modelling consumer taste preferences and predicting their reaction to a new product (Carlos, 2018; Jabeen, Nargis, & Lehmann, 2019; O'Brien, & Rivas, 2019; Guiné, 2019).

The final link of this technological chain where AI can be used is designing new recipes. The review of scientific sources shows the existence of theoretical justification and practical developments in creating new culinary dishes and even composing a whole menu. At the same time, in the confectionery industry, AI is proposed to be used only to predict and model cookie recipes (Kicherer, Dittrich, Grebe, Scheible, & Klinger, 2018; Müller, & Bergmann, 2017; Ohene, 2017). There was found no information on the design of recipes of bakery products using AI that can be manufactured both in conditions of large industrial companies and small bakeries.

The study of the possibility of creating new bread product recipes using AI elements has scientific and practical value.

There are no such works yet, which has led to the need for the creation of an initial (recipe) database to carry out the ML process. Today, such databases can be created by exporting (parsing) recipes from open sources (in particular, from various culinary or relevant websites) or by manually entering data. The first method has clear advantages and such main disadvantages as the inability of checking the data obtained and the lack of unification in their presentation (for example, various dimensions of recipe components dosage values). So, while being practical for the initial filling of recipe database, such sources require considerable technological analysis and improvement later.

The purpose of the research is to create a database for ML based on officially approved collections of recipes used by large industrial companies, small bakeries, and public catering facilities.

Materials and methods All bakery product recipe groups were divided into two classes (the minimum number to ensure the possibility of classifying systems by ML methods). The first class included bread itself, made of wheat or rye flour or their mixture, as well as rolls and buns. The second class included enriched bread with recipes containing mostly highest- or first-quality wheat flour and a considerable amount of sugar and fat. It was also decided not to include special-purpose products (for patients with diabetes mellitus, coeliac disease, phenylketonuria, renal failure, etc.) and products with non-traditional recipe components. These applications have insignificant share in the total mass of recipes that will not considerably affect the ML process.

Analysis of open sources (collections of recipes issued in different years) showed that they have a maximum of 400 recipe compositions of traditional bakery products, which is not enough to carry out the ML process. At the same time, it was noted that recipes of some products differ only in a slight change in amounts of certain recipe components (for example, yeast or salt), which are within technologically acceptable limits. This way allows the creation of new recipes based on existing ones.

So, it was decided to expand the database for ML based on traditional recipes by variation of the main recipe components (as a percentage of flour mass in basic recipes):

- compressed yeast ± 0.5%;
- table salt ± 0.3%;
- the wheat and rye flour ratio for bread-type products based on their mix ranges from 40:60 to 60:40 with a 10% step (for recipes that allow a wider variation — from 20:80 to 80:20).

The possibility of mutual substitution of raw materials was also considered. Recipes containing either native or powdered milk that was allowed by regulatory documentation were created. Additionally, for all recipes containing compressed yeast, new formulations were created in parallel with their replacement by dry (instant) yeast in a ratio of 3:1. For products with rye and rye-wheat flours with the use of sourdough, it was decided to introduce 2...5% of dry sourdough powder (depending on the amount of rye flour in the base recipes), with its variation step of ± 0.5%, for use in conditions of discrete dough preparation. In addition, traditional recipes' collections do not give the amount of water needed for dough preparation. Therefore, it should be calculated depending on the moisture content in the final product. It was then decided to use this factor to increase the number of recipe options: the calculated moisture content value varied within ± 1%.

Results and discussion. Abovementioned introduction of the chosen solutions allowed to increase the total number of recipes in the database for ML from 400 to almost 6 thousand (over 4 thousand recipes of bakery products that could be classified as "Bread" and almost 2 thousand recipes of "Enriched bread" type of application). Recipes were normalized before the use: the sum of recipe components was reduced to 1, with a recalculation of their ratio, respectively (fig. 1).

The created recipe database was analyzed on the Microsoft Azure Machine Learning and Google Cloud Machine Learning Engine platforms. For training, the typical settings were used to divide the created data set into three groups: a training set (80% of the total data), a validation set (10%) and a testing set (10%). All the data array was pre-randomized using a random number generator.

Both ML instruments showed approximately the same efficiency in terms of determining the feature importance of the recipe's components (fig. 2), and in terms of the ability to classify new recipes (referring them to the categories of "bread" or "enriched bread"). It is certain (fig. 2) that the level of particular ingredients in the formulation impacts product classification and the model has been correctly identifying those ingredients. These results are well aligned with industry standards and baker's classification of the final product.

The score threshold (S.T.) for the proposed model was set at the level of 95%. This means that the model category classification will be considered as positive if the accuracy of the prediction achieves 95% or higher. As it can be seen (fig. 3), the share of correct classification predictions created by the model ("Accuracy") for both product groups does not decline below 96%, and the share of correct positive predictions created by the model ("Precision") is 100%.

	A	C	D	K	L	M	N	BL
1	Application	Wheat flour, breadmaking (extraction rate 75%)	Rye flour, bolted (extraction rate 85 %)	Yeast, baker`s, compressed	Yeast, dried	Salt, table	Sourdough, Rye, Bocker 350	Water, tap, drinking, average values
5222	Bread	0.3986	0.1708	0.0140	0.0000	0.0090	0.0100	0.3976
5223	Bread	0.4006	0.1718	0.0110	0.0000	0.0070	0.0100	0.3996
5224	Bread	0.4000	0.1710	0.0110	0.0000	0.0090	0.0100	0.3990
5225	Bread	0.3994	0.1712	0.0110	0.0000	0.0100	0.0100	0.3984
5226	Bread	0.3980	0.1710	0.0170	0.0000	0.0070	0.0100	0.3970
5227	Bread	0.3976	0.1698	0.0170	0.0000	0.0090	0.0100	0.3966
5228	Bread	0.3970	0.1700	0.0170	0.0000	0.0100	0.0100	0.3960
5229	Bread	0.3990	0.1710	0.0140	0.0000	0.0070	0.0100	0.3990
5230	Bread	0.3980	0.1710	0.0140	0.0000	0.0100	0.0100	0.3970
5231	Bread	0.3986	0.1708	0.0000	0.0040	0.0070	0.0100	0.4096
5232	Bread	0.3972	0.1707	0.0000	0.0050	0.0090	0.0100	0.4082
5233	Bread	0.3970	0.1700	0.0000	0.0060	0.0100	0.0100	0.4070
5234	Bread	0.3964	0.1702	0.0140	0.0000	0.0080	0.0130	0.3984
5235	Bread	0.3980	0.1710	0.0110	0.0000	0.0070	0.0130	0.4000
5236	Bread	0.3974	0.1702	0.0110	0.0000	0.0090	0.0130	0.3994
5237	Bread	0.3970	0.1700	0.0110	0.0000	0.0100	0.0130	0.3990
5238	Bread	0.3956	0.1698	0.0170	0.0000	0.0070	0.0130	0.3976
5239	Bread	0.3954	0.1692	0.0170	0.0000	0.0080	0.0130	0.3974
5240	Bread	0.3944	0.1692	0.0170	0.0000	0.0100	0.0130	0.3964
5241	Bread	0.3970	0.1700	0.0140	0.0000	0.0070	0.0130	0.3990
5242	Bread	0.3956	0.1698	0.0140	0.0000	0.0100	0.0130	0.3976

Fig. 1. Example of a normalized recipe

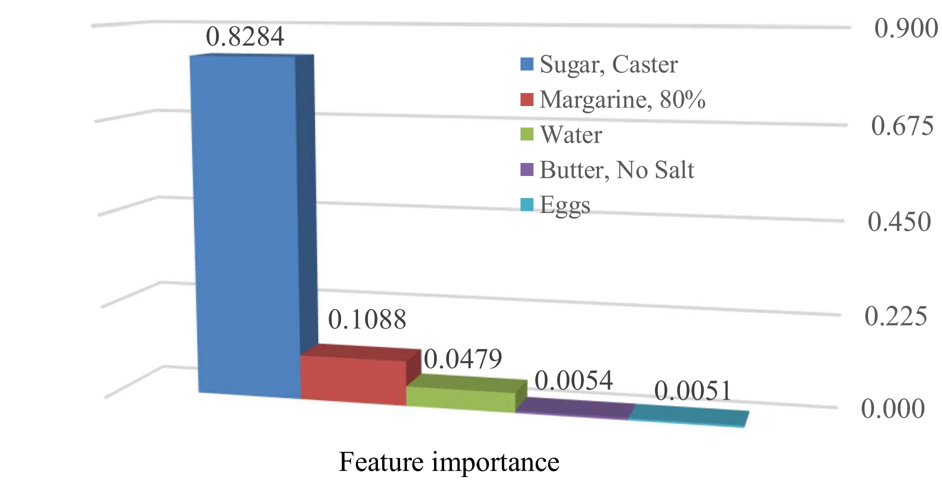

Fig. 2. Feature importance of the recipe ingredients

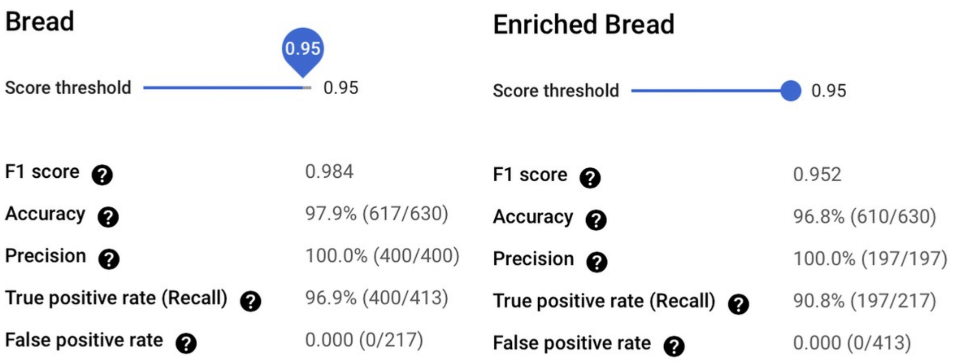

Fig. 3. Model evaluation summary of both application categories

At the same time, the share of real positive results for bread reaches almost 97%, and for enriched bread — almost 91%.

Conclusions

Analysis the model evaluation outcome allows to conclude that the process of preparing the database and the ML is considered to be carried out correctly based on the following:

Feature importance of the recipe ingredients has been identified correctly with fat and sugar level being the key factors that impacts the classification of the final application. Within Ukrainian regulation, if the total level of sugar and fat is higher than 14%, the application should be classified as "Enriched bread".

There are no false positive cases meaning that the model managed to correctly characterize each recipe and assign a category to it.

True positive rate (recall) is higher than 90% overall between 2 application categories proving that only 10% or less of the positive class were predicted with lower than 95% accuracy.

100% precision across both types of the application underline that all classified recipes with the S.T. >0.95 were done correctly.

The overall F1 score is >0.95 (as a harmonic mean of Precision and Recall) confirms that the obtained model, performs correct product classification with high achieved accuracy, despite learning on the unevenly distributed dataset.

Thus, it can be stated that the model is ready for further use in the development of bakery related AI systems. The simplest option can be generation of new recipes based on raw material specified by the user and technologically acceptable limits of their introduction. It will also be possible to select recipe components to create products with a specified chemical composition or a specific biological value. However, this will require connecting existing databases of the chemical composition of raw materials to the created system and taking into account changes in the product's main components during the technological process. Such work is planned for the future and will be covered in further publications.

References

Carlos, H. (2018). Creating Recipes using Machine Learning and Computational Creativity. *Thesis to obtain the Master of Science Degree in Information Systems and Computer Engineering.*

Chidinma-Mary-Agbai (2020). Application of artificial intelligence (AI) in food industry. GSC Biological and Pharmaceutical Sciences, 13(01), 171—178. doi:10.30574/gscbps.2020.13.1.0320.

Fisher, D. H. (2016). Meal Planning from an Abstraction Hierarchy of Menus and Recipes. *ICCBR Workshops*. 235—242.

Guiné, R. (2019). The use of artificial neural networks (ANN) in food process engineering. *International Journal of Food Engineering*, 5(1), 15—21. doi:10.18178/ijfe.5.1.15-21.

Jabeen, H., Nargis T., & Lehmann, J. (2019). EvoChef: Show Me What to Cook! Artificial Evolution of Culinary Arts. *Computational Intelligence in Music, Sound, Art and Design. EvoMUSART 2019.* Lecture Notes in Computer Science, vol. 11453. Springer, Cham., 156—172. doi:10.1007/978-3-030-16667-0_11.

Jermsurawong, J. & Habash, N. (2015). *Predicting the Structure of Cooking Recipes*. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 781—786, Lisbon, Portugal. Association for Computational Linguistics. doi:10.18653/v1/D15-1090.

Kicherer, H., Dittrich, M., Grebe, L., Scheible, C. & Klinger, R. (2018). What you use, not what you do: Automatic classification and similarity detection of recipes. *Data & Knowledge Engineering.* 117, 252—263. doi:10.1016/j.datak.2018.04.004.

Kumar, I., Rawat, J., Mohd, N., & Husain, S. (2021). Opportunities of Artificial Intelligence and Machine Learning in the Food Industry. *Journal of Food Quality,* 2021, 1—10. doi:10.1155/2021/4535567.

Matthew N. O. Sadiku, Omobayode I. Fagbohungbe, & Sarhan M. Musa. (2020). Artificial Intelligence in Food Industry. *International Journal of Engineering Research and Advanced Technology (ijerat)* (E-ISSN 2454-6135), 6(10), 12—19. doi:10.31695/IJERAT.2020.3649.

Müller, G. & Bergmann, R. (2017, June). Cooking Made Easy: On a Novel Approach to Complexity-Aware Recipe Generation. *In ICCBR (Workshops)*, 229—236.

O'Brien, M. & Rivas, P. (2019). *Optimizing Cookie Recipes for Ratings Using Machine Learning and Deep Vector-to-Sequence Recurrent Neural Models*. 21st International Conference on Artificial Intelligence (ICAI 2019). 469—472.

Ohene, M. (2017). A Proposed General Formula to Create and Analyze Baking Recipes. *ICCBR(Workshops)*. 245—252.

Soltani-Fesaghandis, G. & Pooya, A. (2018). Design of an artificial intelligence system for predicting success of new product development and selecting proper market-product strategy in the food industry. *International Food and Agribusiness Management Review*. 21. 1—18. doi:10.22434/IFAMR2017.0033.

Steluti, J., Junior, J., & Marchioni, D. (2020). Development of quality index to classify meal healthiness through photos: First step for app of meal assessment using Machine Learning. Proceedings of the Nutrition Society, 79(OCE2), E274. doi:10.1017/S0029665120002220.

Kakani, V., Nguyen, V. H., Kumar, B. P., Kim,H. & Pasupuleti, V. R. (2020). A critical review on computer vision and artificial intelligence in food industry. Journal of Agriculture and Food Research, 2, 100033. doi:10.1016/j.jafr.2020.100033.