10.18605/2175-7275/cereus.v16n3p197-208

ARTIGO ORIGINAL

<< Recebido em: 18/05/2024 Aceito em: 11/08/2024. >>

REVISTA CEREUS
ISSN: 2175-7275

Classification of Wheat Flour: a Case Study in a Food Company

Classificação de Farinhas de Trigo: um Estudo de Caso em uma Empresa Alimentícia

Jandrei Sartori Spancerski[1], José Airton Azevedo dos Santos[2]

ABSTRACT

Currently, the highly competitive environment requires food companies to optimize their processes. In this context, we suggest using artificial neural networks to optimize a food company's wheat flour classification process. The database, made available by the food company, presents 7666 observations. An algorithm based on the MLP (Multilayer Perception) architecture was implemented in the Python programming language. The Grid Search Cross-Validation technique was used to optimize the hyperparameters of the neural network. Experimental results showed that the MLP model presents an accuracy greater than 95% and a Kappa index of 0.949.

Keywords: Classification; Artificial intelligence; Python.

RESUMO

Atualmente, o ambiente altamente competitivo exige que as empresas alimentícias otimizem os seus processos. Neste contexto, sugerimos a utilização de redes neurais artificiais para otimizar o processo de classificação de farinha de trigo de uma empresa alimentícia. A base de dados, disponibilizada pela empresa alimentícia, apresenta 7666 observações. Um algoritmo baseado na arquitetura MLP (*Multilayer Perception*) foi implementado na linguagem de programação Python. A técnica *Grid Search Cross-Validation* foi utilizada para otimizar os hiperparâmetros da rede neural. Os resultados experimentais mostraram que o modelo MLP apresenta acurácia superior a 95% e índice Kappa de 0,949.

Palavras-chave: Classificação; Inteligência artificial; Python.

[1] Mestre em Tecnologias Computacionais para o Agronegócio (PPGTCA).. Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: jandreisst@gmail.com. ORCID: https://orcid.org/0000-0002-2568-5734

[2] Doutor em Engenharia Elétrica. Docente do Programa de Pós-Graduação em Tecnologias Computacionais para o Agronegócio (PPGTCA). Universidade Tecnológica Federal do Paraná (UTFPR). E-mail: airton@utfpr.edu.br. ORCID: https://orcid.org/0000-0002-4077-281X

1. INTRODUCTION

According to the Brazilian Agricultural Research Corporation, Wheat is the oldest cultivated plant and the most important cereal since prehistoric times (EMBRAPA, 2020). Its cultivation, due to its adaptability to different soils and climates, is widespread worldwide. In Brazil, wheat (*Triticum aestivum*) is primarily cultivated in the southern region, mainly in Paraná and Rio Grande do Sul. It is often used as a rotation crop with soybeans. Because soybeans are a summer crop and wheat is a winter crop (COSTA et al., 2008).

Wheat flour, a product obtained from grinding wheat grain, plays an essential role in human nutrition's economic and nutritional aspects. Wheat grain cultivation, harvesting, drying, and storage conditions directly influence its nutritional and technological characteristics. Consumer markets increasingly consider the requirement for quality wheat flour when purchasing wheat. The quality of wheat flour can directly influence the quality of the final product (COTRIM et al., 2016).

Numerous industrial activities require human interaction in decision-making processes; among them is the classification of wheat flour. Wheat flour, depending on its physical-analytical characteristics, can, in the food industry, be used for both biscuits and pasta. The characteristics of the raw material (flour) are different for each product (NITZKE; THYS, 2020).

Artificial Neural Networks (ANNs) are being applied in several areas of knowledge. They are applied to solve complex problems that are difficult to solve using conventional approaches. These applications include pattern recognition, simulation, and estimation, among others. ANNs are inspired by the way the brain processes information. They are composed of many highly interconnected elements (neurons) working to solve specific problems (PINHEIRO et al., 2020).

ANNs, like people, learn by example. They can memorize, analyze, and process large amounts of data (ESEN et al., 2019). According to Haykin (2001), artificial neural networks can be considered a processing scheme capable of storing knowledge and making it available for a given application.

The vast success of this technique can be attributed to its power, versatility, and ease of use. Neural networks are a very sophisticated modeling technique capable of making predictions of highly complex functions and relationships between data (BASTIANI et al., 2018).

In this context, we suggest using artificial neural networks to optimize a food company's wheat flour classification process.

The text of the article was organized into five sections. The Introduction presents an overview of the topic and the work's objective. Section 2 addresses the theoretical foundation, providing a conceptual basis for understanding the relevant terms used in the work. Section 3 presents the methodologies used to classify flour from a food company. In Section 4, the results obtained from the application of the MLP Neural Network model are presented. Final comments and conclusions conclude the work.

2. MATERIALS AND METHODS

2.1 Wheat Flour

Wheat flour is obtained from *Triticum species* (except *Triticum durum*) by grinding the processed wheat grain (BRASIL, 2005).

The quality of wheat flour is decisive for the characteristics of the final products. As the range of products that contain flour in their composition is vast (bread, biscuits, cakes, and pasta), the characteristics of the raw material are entirely different for each one, which is called typified flour. Typified wheat flour is a flour that has characteristics that are within the specification range of a given product, providing a final product with excellent visual, aromatic, and palatable characteristics, with an attractive nutritional value and at a competitive cost (NITZKE; THYS , 2020).

The characteristics mentioned above are different due to several factors, the main ones being the type of wheat used to make the flour (genetic factors), the place where this wheat was planted (soil nutrition), and the climatic conditions in which this wheat was cultivated. Several parameters are analyzed in the laboratory to obtain information regarding wheat flour's technological and industrial quality or even flour mixtures (BROCA, 2021; NITZKE; THYS, 2020).

2.2 MLP Neural Networks

Multilayer Perceptron (MLP) networks are composed of layers and nodes. The layers are related to their location in the network; each has a function, generally having an input layer, an output layer, and the so-called hidden layers between them.

The layers contain nodes (also called neurons), which carry information and are also the connection points between one layer and another. The MLP Network is formed by a tangle of node connections between layers (HAYKIN, 2001; SAS, 2020).

MLP networks are widely used in regression and classification problems. The backpropagation algorithm is the most used for training neural networks (MOREIRA, 2020; SAS, 2020).

2.3 Cross Validation

This work used cross-validation to divide the data set into training and validation. In cross-validation, the data is separated into approximately equal sets. Each is used only once for validation, while the others are used for training. The idea is that all sets are once treated as the validation set; that is, several analyses will be carried out, and the final accuracy result is obtained by averaging the accuracies of each iteration (PINHEIRO et al., 2020).

2.4 Grid Search Techniques

The parameters of a neural network must be adjusted to obtain good results (HAYKIN, 2001). These parameters are known as hyperparameters. This process involves testing different network parameters and seeing which offers better returns. In this work, the Grid Search Cross-Validation technique was used to adjust (optimize) the hyperparameters of the neural network. This technique will test all possible combinations of hyperparameters, compare results, and choose the combination with the best accuracy.

2.5 The Company Under Study

The company, the subject of this study, was created in the 1970s and is located in Paraná. It currently has approximately 600 employees. It can produce more than 200 tons of food daily, divided into biscuit and pasta production units.

2.6 Methodological Classification

The types of research are classified according to their nature, approach, research objectives, and technical procedures (GERHARDT; SILVEIRA, 2009).

The classification in relation to (GIL, 2010):

- ✓ **Nature**: it is of the applied type. Type of research that aims to generate knowledge for practical application.

- ✓ **Approach**: classified as qualitative-quantitative. A modality that combines the characteristics of both quantitative and qualitative research.
- ✓ **Objectives**: the research is descriptive since a specific population is described. It aims to identify relationships and associations between variables.
- ✓ **Technical procedures**: this is operational research, where mathematical models, statistics, and algorithms assist decision-making.

2.7 Research Procedures

The procedures adopted in the development of the research follow the steps indicated in Figure 1:

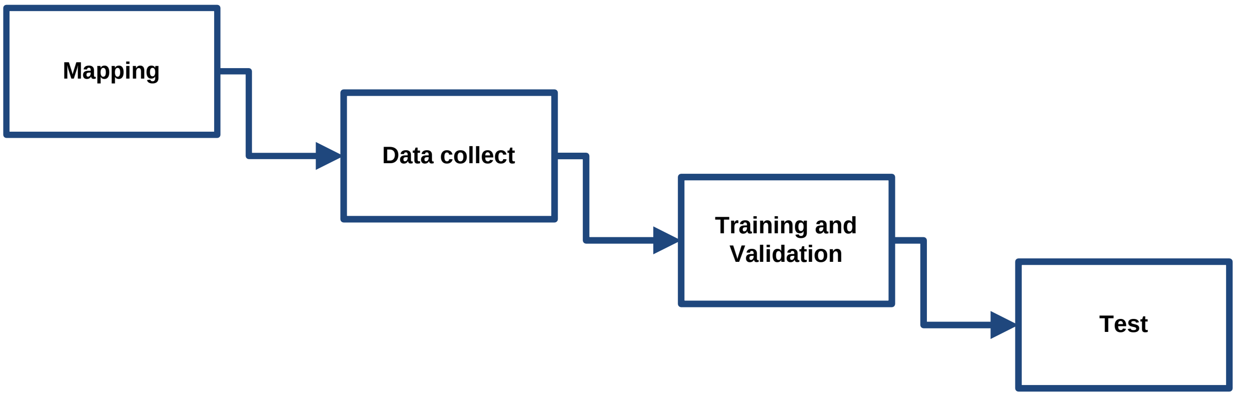

Figure 1. Research steps
Source: The authors (2023).

The Mapping stage was used to learn about the procedures carried out by the company laboratory under study. Next, Data Collection took place directly in the company's database. Information about all analyses performed by the laboratory is contained in the database. Subsequently, data pre-processing was carried out, consisting of variable selection and data cleaning. At the end of this stage, a unified database free of inconsistencies was obtained. Subsequently, the best MLP model was selected using the Grid Search Cross-Validation technique in the Training and Validation stage. Finally, in the Test Stage, the MLP model was tested on data that did not participate in the Training and Validation stages.

2.8 Data base

Initially, a data set with 6644 analyses was obtained to classify the flour type. The data set is made up of the following variables: Company: flour supplier; DU: Days between the manufacture of wheat flour and its use in the production process (days); UM: Wheat flour moisture percentage (%); P: Tenacity (mm); L: Extensibility (mm); P/L: Tenacity/extensibility ratio; W: Gluten strength (10^{-4}J); Ie: Elasticity index (%); FN: Falling Number or falling number (s); GU: Moist gluten (g); GS: Dry gluten (g); Index: Gluten content (%); Bu: Ash on a wet basis (g); Bs: Ash on a dry basis (g); L*: Brightness; a* and b*: Chromaticity coordinates.

In the classification, the class will be the Product (Dough or Biscuit), while the predictors will be the variables presented previously.

Some of the variables from the company's database are explained below (EMBRAPA, 2020).

- ✓ **Company**: the company variable refers to wheat flour suppliers. These companies transform wheat grain into flour, each with their grain suppliers, making them a preliminarily important variable for this study since wheat produced under different conditions generates different flours.
- ✓ **DU**: the Manufacturing Date and Entry Date variables were transformed into one DU variable (Days until use). It is observed that wheat flour arrives from the supplier and is sometimes not used on the same day or week. This difference in dates influences the quality of the flour.
- ✓ **Product**: This variable refers to which product type is intended for flour (Biscuit or Pasta), given its characteristics.
- ✓ **Ash**: represents the mineral or ash content of the wheat grain or flour and is expressed as a percentage.
- ✓ **Falling Number**: measures the intensity of activity of the α-amylase enzyme in the grain, with the result expressed in seconds.
- ✓ **Gluten strength**: represents the dough's deformation work and indicates the flour's baking quality (flour strength).
- ✓ **Tenacity**: the maximum pressure required to expand the dough.
- ✓ **Extensibility of the dough**: dough's ability to stretch without breaking.
- ✓ **Tenacity/extensibility relationship**: expresses mass balance.
- ✓ **Elasticity index**: is correlated to the recovery phenomena of the initial shape after deformation, allowing a better prediction of the rheological behavior of the dough used in industrial baking and biscuit production.

2.9 Metrics

Various metrics can be used to measure a model's performance. The following metrics were used in this work (CANKURT; SUBASI, 2015):

- ✓ **Accuracy**: In classification problems, accuracy is one of the most used metrics. This metric indicates the percentage of correctly classified examples (Equation 1).

$$Accuracy = \frac{ICC}{ICC + ICI} \quad (1)$$

Where: *ICC*: correctly classified instances and *ICI*: incorrectly classified instances.

- ✓ **Confusion Matrix**: This matrix is generated by classifying the data set into categories. The Confusion Matrix Indicates the efficiency of a classifier for samples from different classes.
- ✓ **Kappa index** (*k*) (Equation 2): evaluates the agreement level between two data sets. It provides an idea of how far the observations differ from those expected, indicating how legitimate the interpretations given by the Confusion Matrix are.
- ✓

$$k = \frac{Po - Pe}{1 - Pe} \quad (2)$$

Where: P_o: relative acceptance rate and P_e: hypothetical acceptance rate.

Levels of agreement are classified according to the criteria presented in Table 1 (LANDIS and KOCH, 1977).

Table 1. Agreement Levels – Kappa Index

Kappa Coefficient Value	Level of Agreement
<0	There is no agreement
0 - 0,20	Minimum agreement
0,21 - 0,4	Reasonable agreement
0,41 - 0,60	Moderate agreement
0,61 - 0,80	Substantial agreement

Source. Landis e Koch (1977).

3. RESULTS AND DISCUSSION

3.1 Descriptive Analysis

In this work, a descriptive analysis of the data was carried out. Table 2 presents the Tenacity (*P*) and Extensibility (*L*) analysis.

Table 2. Descriptive data analysis

Description	*P*	*L*
Observations	6644	6644
Average (*mm*)	112,73	69,4
Standard Deviation (*mm*)	25,53	21,15
Coefficient of variation (%)	26,35	30,47

Source: The authors (2023).

It can be seen from the data presented in Table 2 that Tenacity and Extensibility averaged 112.73 mm and 69.4 mm, respectively. It can also be seen that the coefficients of variation are 26.35% and 30.47%, which is considered high and indicates data variability.

3.2 Training and Validation

Initially, all combinations of hyperparameters presented in Figure 2 were tested.

```
parametros = {'batch_size': [10, 30],
              'epochs': [100, 300, 500],
              'optimizer': ['adam', 'adagrad', 'adamax'],
              'kernel_initializer': ['random_uniform', 'normal'],
              'activation': ['relu', 'sigmoid', 'softmax', 'softplus',
                             'softsign', 'tanh', 'selu', 'elu', 'exponential'],
              'neurons': [7, 14, 28, 56],
              'loss': ['binary_crossentropy', 'poisson', 'kl_divergence'],
              'dropout': [0.2, 0.3]}
```

Figure 2. Network hyperparameters
Source: The authors (2023).

Where:

- ✓ **dropout**: is a technique to reduce overfitting, which controls the percentage of neurons randomly deactivated during training.

- ✓ **epochs**: the number of times the learning algorithm accesses the training set to update the neuron weights.
- ✓ **neurons**: to bring more or less complexity to the model, it is allowed to add or remove neurons.
- ✓ **activation**: indicates which activation function will be used by the artificial neurons.
- ✓ **batch_size**: corresponds to the number of training cases used in each epoch.
- ✓ **optimizers**: are algorithms that change neural network attributes, such as weights and learning rate, to reduce losses.
- ✓ **loss**: estimates the model loss so that the weights can be updated.
- ✓ **kernel_initializer**: initializes the weights.

The MLP model optimized, through this technique, presented the parameters shown in Figure 3.

melhores_parametros - Dicionário (8 elementos)

Chave	Tipo	Tamanho	Valor
activation	str	1	relu
batch_size	int	1	10
dropout	float	1	0.25
epochs	int	1	500
kernel_initializer	str	1	normal
loss	str	1	binary_crossentropy
neurons	int	1	56
optimizer	str	1	adam

Figure 3. Optimized hyperparameters
Source: The authors (2023).

It is observed that the processing period for the hyperparameters lasted five days, three hours, and five minutes, using a computer with an Intel® processor, model i5, 7th generation. A total of 11664 different models were tested during this period.

3.3 Test

The company's database was reaccessed to collect data in production and test the optimized model. One thousand twenty-two new data were collected from the company's database.

Table 3 presents the correctly and incorrectly classified instances, the kappa index, and the confusion matrix. In the confusion matrix, the Biscuit class had 14 flour samples wrongly classified as Dough, and for the Dough class, 21 flour samples were classified as Biscuit. Presenting 96.57% of instances classified correctly (987).

Table 3. Results of applying the MLP model – Test set

Summary	**Test**			
Correctly classified instances	987 – 96,57%			
Incorrectly classified instances	35 – 3,43%			
Kappa Statistics	0,949			
Total Instances	1022			
		a	**b**	
Confusion matrix	**a**	484	14	**a= Cookie**
	b	21	503	**b= Mass**

Source: The authors (2023).

Analyzing the results in Table 3, with emphasis on the result of the Kappa index, it can be classified as very good, according to the methodology of Landis and Koch (1977) (Table 1), thus showing that the samples collected were consistent with the classified information.

4. FINAL CONSIDERATIONS

In this study, wheat flour from the food industry was classified. In this classification, a database with 7666 observations was used.

Initially, null and redundant values were eliminated from the database. Subsequently, the Grid Search Cross-Validation technique was used in the training and validation stage to test 11644 network configurations—the network with optimized parameters obtained, in the test set, 96.57% of correctly classified instances. Also, an excellent Kappa index (0.949) was obtained. Therefore, the results showed good performance of the MLP neural network in classifying the flours tested.

It was also observed that the Grid Search Cross-Validation technique is essential for selecting the best model. However, it significantly increases computing time.

For the conditions established in this work, it was concluded that MLP Neural Networks could significantly benefit the food industry in classifying flour.

REFERENCES

BASTIANI, M.; SANTOS, J. A. A.; SCHMIDT, C. A P.; SEPULVEDA, G. P. L. (2018). Application of data mining algorithms in the management of the broiler production. **Revista Geintec**, v. 8, n. 1, 2018.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa n.º 8, de 2 de junho de 2005. Aprova o Regulamento Técnico de Identidade e Qualidade da Farinha de Trigo, conforme o anexo desta Instrução Normativa. **Diário Oficial da União**, Brasília, DF, 2 de junho de 2005.

BROCA, A. **Estudo da qualidade do trigo e da farinha de trigo destinada a panificação em um moinho no sul do Brasil**. Dissertação de Mestrado.Laranjeiras do Sul. Universidade Federal da fronteira Sul. 2021.

CANKURT, S.; SUBASI, A. Comparasion of linear regression and neural network models forecasting tourist arrivals to turkey. **Eurasian Journal of Science & Engineering**, 2015.

CARVALHO, N. M.; NAKAGAWA, J. **Sementes: ciência, tecnologia e produção**. 5 ed. São Paulo: Funep, 2012.

COSTA, M. G.; SOUZA, E. L.; STAMFORS, T. L. M.; ANDRADE, S. A. C. Qualidade tecnológica de grãos e farinhas de trigo nacionais e importados. **Food Science and Technology**, Campinas, 2008.

COTRIM, M. F.; ALVAREZ, R. C. F.; SERON, A. C. C. **Qualidade fisiológica de sementes de trigo em resposta a aplicação de Azospirillum Brasilense e ácido húmico**. Revista Brasileira de Engenharia de Biossistemas, v. 10, n. 4, 2016. DOI: https://doi.org/10.18011/bioeng2016v10n4p349-357

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. **Trigo: germinação e posterior extrusão para obtenção de farinha integral extrusada de trigo germinado**. (2020). Disponível em: https://www.infoteca.cnptia.embrapa.br/bitstream/doc/852529/1/pdo74.pdf. Acesso em: 17 abr. 2020.

ESEN, H.; OZGEN, F.; ESEN, M.; SENGUR, A. Artificial neural network approaches for modeling of a solar air heater. **Expert System with Applications**, v. 1, 2019. DOI: https://doi.org/10.1016/j. eswa.2009.02.073

GERHARDT, T. E.;SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5 ed. São Paulo: Atlas, 2010.

HAYKIN, S. **Neural networks: a comprehensive founda**tion. New Delhi: Pearson Prentice Hall, 2001.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1 , 1977.

MOREIRA, S. **Rede neural perceptron multicamadas**. Disponível em: https://medium.com/ensina-ai/rede-neural-perceptron-multicamadas-f9de8471f1a9. Acesso em: 08 jun. 2023.

NITZKE, J. A.; THYS, R. C. S. **Avaliação da qualidade tecnológica/industrial da farinha de trigo**. 2020. Disponível em: https://www.ufrgs.br/napead/projetos/avaliacao-farinha-trigo/index.php. Acesso em: 18 fev. 2023.

PINHEIRO, T. C.; SANTOS, J. A. A.; PASA, L. A. Gestão da produção de frangos de corte por meio de redes neurais artificiais. **Revista Holos**, v. 2, 2020. DOI: https://doi.org/10.15628/holos. 2020.9043

SAS. **Redes Neurais: O que são e qual sua importância**. Disponível em: https://www.sas.com/pt_br/insights/analytics/neural-networks.html. Acesso em: 27 jan. 2020.

TAKEITI, C. Y. **Trigo**. (2015). Disponível em: https://www.embrapa.br/agencia-de-informacao-tecnologica/tematicas/tecnologia-de-alimentos/processos/ grupos-de-alimentos/cereais-e-graos/trigo. Acesso em: 23 abr.. 2023.

WOLKE, R. L. **What einstein told his cook 2: further adventures in kitchen science**. 2 ed. New York: W. W. Norton, 2005.