Evaluation and Comparison of the Processes in the Frozen Vegetable Production Using Machine Learning Methods

Piotr Milczarski

Faculty of Physics and Applied Informatics, University of Lodz, Pomorska str. 149/153, Lodz, Poland

Abstract
In the paper, the study of the carbon footprint (CF) assessment in the frozen vegetable production processes is shown in order to receive low-carbon products. Three methods of clusterization have been chosen for the production assessment. The results of clusterization are evaluated by five classification methods: k-Nearest Neighbors, Multilayer Perceptron, C4.5, Random Forrest and Support Vector Machines with a radial basis kernel function. In the chosen model with five clusters, the best clusterization methods are k-means followed by Canopy.

Keywords
Carbon Footprint; clusterization; Canopy, k-means, Expectation-Maximization; k-Nearest Neighbors; Multilayer Perceptron; C4.5; Random Forrest; Support Vector Machines

1. Introduction

Greenhouse gas emissions from human activities have been a major contributor to global warming since the mid-twentieth century. Agriculture and land-use change contributed to 17% of global anthropogenic greenhouse gas emissions in 2010 [1]. By 2050 the population will be 9 billion people [2] to ensure supplying of food, agricultural production should be increased by 60%. Climate change can affect food availability; for example, an increase in temperature, a change in the structure of rainfall or extreme weather events may result in a reduction in agricultural productivity [3, 4]. Therefore, its main challenge has become to mitigate the threats that climate change poses to food security.

In response to the emerging threats of climate change, numerous programs, both global and regional, have been developed, the purpose of which is to slow down the growth rate of GHG concentration [5]. Achieving climate policy goals requires continuous monitoring of emissions and verification of the effectiveness of solutions for the development of a low-emission economy.

The adoption of an action plan for the reduction of gaseous emissions by EU countries in 2014 requires the reduction of GHG emissions by 30% by 2030, compared to the level in 2005 [6]. The methods of calculating the carbon footprint are most often based on well-known standards. Among them, the most used are:

- ISO14040: 2006 [7] – Environmental management-life cycle assessment: principles and framework,
- ISO14064-1: 2018 [8] – Greenhouse gases - Part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals,
- ISO/TS 14067:2018 [9] – Greenhouse gases - Carbon footprint of products - Requirements and guidelines for quantification,
- PAS2050 [10] – Specification for the assessment of the life cycle greenhouse gas emissions of goods and services.

Once the carbon footprint has been calculated, its detailed data helps to identify weaknesses, i.e. high-emission areas, that can be eliminated or improved. Thus, the carbon footprint is an indicator of sustainable development

ISIT 2021: II International Scientific and Practical Conference «Intellectual Systems and Information Technologies», September 13–19, 2021, Odesa, Ukraine
EMAIL: piotr.milczarski@uni.lodz.pl (A. 1);
ORCID: 0000-0002-0095-6796 (A. 1);

2. Carbon footprint assessment using Life Cycle Assessment (LCA) method

Carbon footprint calculation is used as a tool for assessing greenhouse gas emissions, helping to manage and reduce them. The carbon footprint is typically calculated using carbon emission factors and activity data that can be assessed through a Life Cycle Assessment (LCA). The carbon footprint analysis according to the LCA methodology is carried out by identifying potential environmental threats, usually throughout the entire life cycle of a product, i.e. from the extraction and processing of raw materials, their transport, through main production, distribution and use, to waste management [11]. However, in agricultural production, the emissions directly related to energy consumption are not dominant [12]. A large part of GHG emissions on farms is gas losses from farmland and livestock. While calculating the carbon footprint with the use of agricultural emission models according to the IPCC reports, all emission sources are taken into account, both those related to energy carriers and processes taking place in the agricultural environment.

LCA is a widely used approach to assess the actual environmental impact of a product from its production and use [11] [12] [13]. The standards for assessing the product carbon footprint in LCA are mainly PAS 2050 [10] and ISO / TS 14067 [9].

In the case of the CFOOD project, that is presented in the paper, the focus is on the optimization of the frozen food production process, so we consider a segment of the product life cycle from the moment of raw material delivery to the shipment of the finished frozen food to the recipient

According to the adopted LCA methodology, the carbon footprint of a product consists of carbon footprints generated at the following stages of its production. Hence the total CF for a given product or its unit value can be expressed by the following formula [14][15][16]:

$$CF = \sum_{i=a}^{r} CF_i \tag{1}$$

where: i is each of the stages of the product life cycle, $i = a, m, t, u,$ and r, relate to the extraction of raw materials, production, transport, use as well as the recycling and disposal stage, respectively.

3. Carbon footprint assessment in CFOOD project

In the case of the CFOOD project, we focus on the optimization of the frozen food production process, so we consider a segment of the product life cycle from the moment of raw material delivery to the shipment of the finished frozen food to the recipient. The production process can be divided into several smaller stages:

- S1 – initial cooling of the raw materials before the processing;
- S2 – the raw material preparation for the production;
- S3 – raw material pre-processing on the production line;
- S4 – product freezing in the cold tunnel;
- S5 – product preparation to a coldstore.

Each of the process stages is connected to electric meter units. Each production stage has also a preparation phase that is measured separately, e.g. S1 has a preparation phase that is denoted pS1, etc.

In the research section, we have tested several clusterization methods and choose three: Canopy, k-Means (KM) and Expectation-Maximization (EM) [17][18]. We have tested several options with the cluster numbers and chosen five clusters for each method that should represent according to our experience some real-time situations that occur during the production and their accounting systems:

- Optimal production – the product has the temperature from -25°C till -18°C at the end of the line;
- Close to optimal – during the high season through-output should be higher, hence the energy consumption should be lower, the product temperature is allowed to be from the range -6°C and -18°C.
- Wrong accounting of some parameters e.g. operators mistakes resulting in too high or too low results e.g. the through-output.
- Malfunction of the energy meters. It is a different situation from the above one and might result in random results.

The clusterization model with five clusters should have at least 60 processes. After a year of the process measurement, till June 2021, we have collected 152 results only for the frozen onion production and 75 for the spinach. The other vegetables have less than 50 cases. Nonetheless, the other production e.g. broccoli and cauliflower should also be optimized. That is why in the current work, the results of clusterization of 35 broccoli processes and 42 cauliflower ones are presented in the current paper.

In the previous work [15][16] to assess the onion and spinach production processes we have prepared the set of verified data and to assess the trustworthiness of the production data we have compared the results of processes classification using 5 classifiers: k-Nearest Neighbors, Multilayer Perceptron [17], C4.5, Random Forrest and Support Vector Machines with a radial basis kernel function [17]. In the current paper, the focus is on unsupervised methods i.e. clusterization [17] into the broccoli and cauliflower processes.

Table 1
K-means clusterization of broccoli production, the units for stages i-th stage pS1, S1 etc. are in kWh/ton, for pt in ton/h, for et in kWh/h

	Broccoli Clusters K-Means				
Attribute	0	1	2	3	4
pS1	0.08	0.32	0.04	4.19	0.09
S1	1.34	1.35	1.51	4.25	2.08
S2	0.16	0.03	0.23	0.09	0.08
pS3	0.06	0.05	0.03	0.11	0.06
S3	0.91	1.14	0.70	0.21	1.38
pS4	7.68	2.29	0.12	6.54	0.25
S4	49.10	55.69	3.07	13.19	6.40
pS5	0.01	0.18	0.00	0.18	0.01
S5	0.18	1.51	0.03	0.24	0.17
pt	1.56	1.46	1.80	2.11	2.12
et	98.67	91.01	9.91	57.77	20.32
instances	4	4	3	22	2

In Tables 1-3 and 4-6 there are clusterization results of the broccoli and cauliflower production processes. The units for stages i-th stage pS1, S1 etc. are in kWh/ton, for pt in ton/h, for et in kWh/h. The results are achieved using the chosen clusterization methods with five clusters:

- Canopy: max-candidates = 100; periodic-pruning = 10000 ; min-density = 2.0; T2 radius = 0.804 and T1 radius = 1.005
- k-Means (KM) with Euclidean distance, max-candidates = 100, periodic-pruning = 10000, min-density = 2.0, T1 = -1.25 and T2 = -1.0.
- Expectation–Maximization (EM) with max-candidates = 100, “minimum improvement in log likelihood” = 1E-5, “minimum improvement in cross-validated log likelihood” = 1E-6, and “minimum allowable standard deviation” = 1E-6.

Table 2
Canopy clusterization of broccoli production

	Broccoli Cluster Canopy				
Attribute	0	1	2	3	4
pS1	0.09	0.39	0.08	0.13	0.13
S1	2.85	1.53	0.13	6.92	0.71
S2	0.11	0.03	0.10	0.11	0.05
pS3	0.02	0.06	0.05	0.00	0.07
S3	0.44	1.25	0.63	0.14	0.63
pS4	1.59	1.75	5.22	0.14	5.36
S4	16.85	58.77	45.3	10.65	43.53
pS5	0.01	0.24	0.00	0.00	0.22
S5	0.21	1.74	0.00	0.21	0.42
pt	2.00	1.35	1.55	1.90	1.92
et	42.19	85.69	82.9	33.65	100.1
instances	16	3	3	8	5

Table 3
EM clusterization of broccoli production

	Broccoli Cluster EM				
Attribute	0	1	2	3	4
pS1	0.09	0.33	0.02	89.74	0.25
S1	3.17	13.28	1.16	6.92	1.46
S2	0.08	0.11	0.23	0.14	0.06
pS3	0.01	0.02	0.04	2.16	0.06
S3	0.27	0.55	0.77	0.14	1.01
pS4	0.30	1.86	4.55	129.4	3.27
S4	8.60	38.08	20.92	11.29	52.48
pS5	0.01	0.05	0.00	3.61	0.14
S5	0.18	0.68	0.02	0.27	1.02
pt	2.13	2.07	1.71	1.96	1.55
et	26.84	104.9	44.61	465.0	95.07
instances	19	2	5	1	8

Figures 1 and 2 show the energy consumption during the production on the energy meters of the chosen stages S1, S2, S3 and S4 for the chosen broccoli process with ID 373 and the cauliflower process with ID 365.

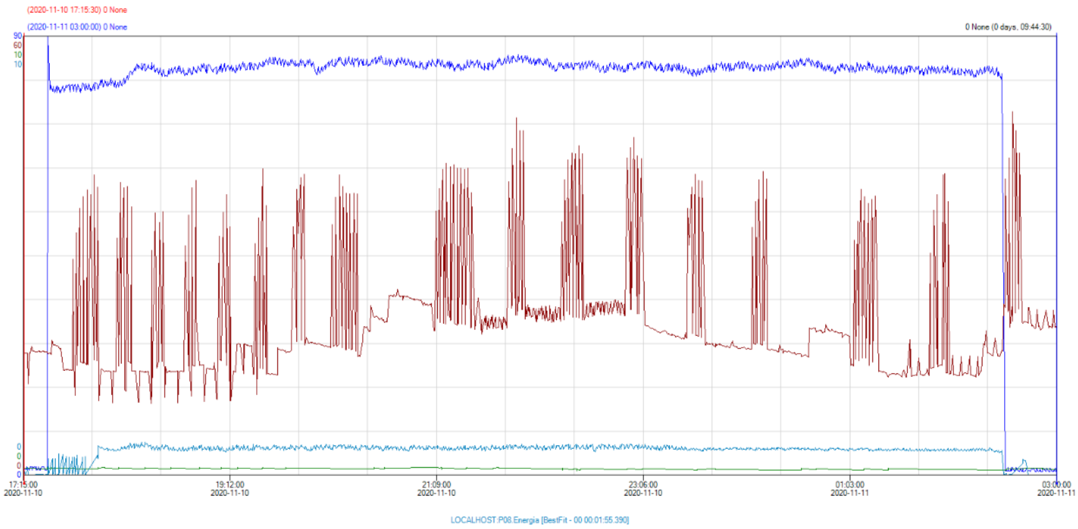

Figure 1: Example of energy consumption for the broccoli production, process ID 373; the colors of the stages: S1 – brown, S2 – green, S3- light blue, S4 - dark blue.

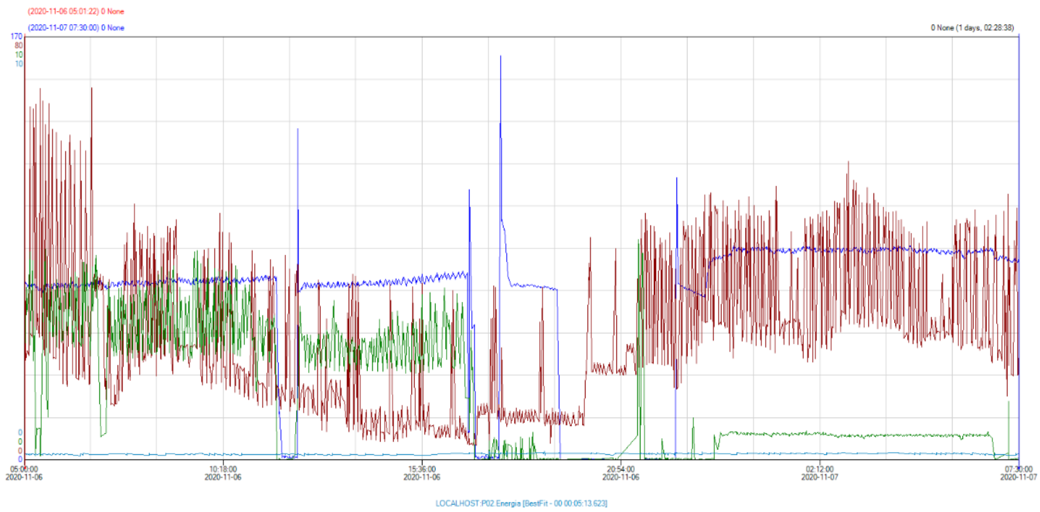

Figure 2: Example of energy consumption for the cauliflower production, process ID 365; the colors of the stages: S1 – brown, S2 – green, S3- light blue, S4 - dark blue.

Table 4

K-means clusterization of cauliflower production

	Cauliflower Clusters K-Means				
Attribute	0	1	2	3	4
pS1	0.52	0.18	5.46	6.97	519.2
S1	24.27	2.48	7.08	1.00	2.28
S2	1.13	0.10	0.14	0.06	0.05
pS3	0.17	0.06	0.16	3.20	157.7
S3	8.41	0.97	1.71	0.55	1.21
pS4	0.43	5.22	3.67	22.58	678.1
S4	28.30	57.14	17.50	3.14	5.55
pS5	0.02	0.22	0.14	0.84	48.59
S5	0.69	1.31	0.33	0.06	0.24
pt	1.86	1.37	2.07	1.64	2.22
et	127.0	92.66	79.17	81.15	3332
instances	3	5	17	15	2

Table 5

Canopy clusterization of cauliflower production

	Cauliflower Cluster Canopy				
Attribute	0	1	2	3	4
pS1	5.23	0.50	519.2	0.70	0.10
S1	4.52	24.42	2.28	14.62	7.16
S2	0.11	1.60	0.05	0.35	0.08
pS3	1.35	0.09	157.7	0.01	0.01
S3	1.34	8.24	1.21	0.77	2.72
pS4	11.26	0.36	678.1	0.11	0.18
S4	17.43	26.35	5.55	4.30	11.93
pS5	0.42	0.01	48.59	0.00	0.01
S5	0.37	0.55	0.24	0.13	0.58
pt	1.80	1.87	2.22	1.67	1.81
et	83.16	123.6	3332	36.75	44.63
instances	27	2	2	3	8

4. Evaluation of the clusterization

In the discussion presented in Tables 1-6 and, the optimal clusters have been highlighted. All values for the stages and their preprocessing phase are in kWh/ton, the production through output (pt) in [ton/h]. K-means and EM seem to provide the best assessment of the processes because it's the best cluster that has the lowest energy consumption from the three optimal clusters for each clusterization.

Table 6
EM clusterization of cauliflower production

	Cauloflower Cluster EM				
Attribute	0	1	2	3	4
pS1	3.44	0.50	0.17	34.90	519.2
S1	4.13	23.95	2.13	0.06	2.28
S2	0.10	0.94	0.10	0.00	0.05
pS3	0.11	0.13	0.08	16.03	157.7
S3	1.31	6.59	0.96	0.00	1.21
pS4	2.13	0.34	5.53	113.2	678.1
S4	11.01	22.59	54.4	0.28	5.55
pS5	0.09	0.01	0.19	4.24	48.59
S5	0.23	0.58	1.11	0.01	0.24
pt	1.89	1.94	1.47	1.55	2.22
et	48.6	112.4	94.3	363.0	3332
instances	27	4	6	3	2

To assess and to choose the clusterization method we have used five machine learning methods as in our previous work [11][12]. All the clusterization results were assessed by the classification methods with the same parameters. In Tab. 5 there are classification results of the production processes using the following classifiers:

- 3NN (kNN) 3-Nearest Neighbors;
- Multilayer Perceptron (MLP) with a hidden layer with 16 nodes for both productions with a learning rate equal to 0.79 and momentum equal to 0.39 [13];
- binary tree C4.5 with a confidence factor equal to 0.25, with a minimum number of instances per leaf equal 2;
- Random Forrest (RF) with the bag size percent equal to 100, with maximum depth unlimited, number of execution slots equal to 1 and 100 iterations;
- Support Vector Machine (SVM) with a radial basis function (RBF) given by the Eq. (2):

$$K(x,y) = exp(-0.05*(x-y)^2) \qquad (2)$$

Table 7
Evaluation of the broccoli clusterization by the chosen classifiers

Classifier	Broccoli evaluation results [%]		
	Canopy	KM	EM
3NN	85.7	97.1	97.1
C4.5	94.3	100	97.1
MLP	97.1	94.3	97.1
RF	100	100	100
SVM	100	100	100

Table 8
Evaluation of the cauliflower clusterization by the chosen classifiers

Classifier	Cauliflower evaluation results [%]		
	Canopy	KM	EM
3NN	90.5	90.5	85.7
C4.5	95.2	97.6	97.6
MLP	92.9	81.0	92.9
RF	100	100	100
SVM	100	100	100

5. Conclusions

In the paper, three clusterization methods have been shown that allow us to assess the processes and their impact on energy consumption and hence, the carbon footprint. We have shown that all the clustering methods point out the processes that are proper from the manufacturing point of view. In the paper, the results for the broccoli and cauliflower production taking into account 35 and 42 corresponding processes respectively have been shown. Currently, we collect new processes for the other vegetable products. The will be analyzed using the clustering methods shown above

The k-means classifier is fast and simple, it has significant disadvantages because it is sensitive to emissions that distort the average value. Although it gives EM the best results in the assessment of the whole production it is planned to use k-SVD and fuzzy k- means methods in future work.

6. Acknowledgements

The paper is co-financed by the Polish National Center for Research and Development, grant CFOOD number BIOSTRATEG3/343817/17/NCBR/2018.

7. References

[1] O. Edenhofer, R. Pichs-Madruga, Y. Sokona, E. Farahani, S. Kadner, K. Kadner, A. Seyboth, I. Adler, S. Baum, G. Myhre, et al. “Climate Change 2014: Mitigation of Climate Change” *Working Group III Contribution to the IPCC Fifth Assessment Report*, Cambridge University Press: Cambridge, UK, 2015.

[2] Food and Agriculture Organization of the United Nations (FAO). Regional Strategy for Sustainable Hybrid Rice Development in Asia, Food and Agriculture Organization of the United Nations Regional Office for Asia and the Pacific: Bangkok, Thailand, 2014.

[3] D.B. Lobell, W. Schlenker, J. Costa-Roberts, “Climate trends and global crop production since 1980”, *Science 2011*, 333, 616–620.

[4] R.Y.M. Kangalawe, C.G. Mungongo, A.G. Mwakaje, E. Kalumanga, P.Z. Yanda, “Climate change and variability impacts on agricultural production and livelihood systems in Western Tanzania”. Clim. Dev. 2017, 9, 202–216.

[5] ECE Strategies and policies for air pollution abatement. United Nations, New York and Geneva, 2007.

[6] *European Council Conclusions 2014. 2030 Climate and energy policy framework. Conclusions* – 23/24 October 2014, EUCO 169/14, http://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ec/145397.pdf

[7] ISO14040 - Environmental management-life cycle assessment: principles and framework. International Organization for Standardization, Geneva, 2006.

[8] ISO14064-1 - Greenhouse gases - Part 1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals. International Organization for Standardization, Geneva, 2018.

[9] ISO/TS 14067 - Greenhouse gases - Carbon footprint of products - Requirements and guidelines for quantification. International Organization for Standardization, Geneva, 2018.

[10] PAS 2050 (2011) “The Guide to PAS2050-2011, Specification for the assessment of the life cycle greenhouse gas emissions of goods and services. British Standards Institution, 2011.

[11] M.A. Renouf, C. Renaud-Gentie, A. Perrin, C. Kanyarushoki, F. Jourjon, “Effectiveness criteria for customised agricultural life cycle assessment tools”, *J. Clean. Prod. 179*, 2018, 246–254

[12] D. Perez-Neira, A. Grollmus-Venegas, “Life-cycle energy assessment and carbon footprint of peri-urban horticulture. A comparative case study of local food systems in Spain”, *Landscape and Urban Planning* 172, 2018, 60-68

[13] A. Nabavi-Pelesaraei, S. Rafiee, S.S. Mohtasebi, H. Hosseinzadeh-Bandbafha, K. Chau, “Energy consumption enhancement and environmental life cycle assessment in paddy production using optimization techniques”, *J. Clean. Prod. 162*, 2017, 571-586

[14] P. Milczarski, A. Hłobaż, P. Maślanka, B. Zieliński, Z. Stawska, P.Kosiński, "Carbon footprint calculation and optimization approach for CFOOD project", CEUR Workshop Proceedings 2683 (2019) 30-34

[15] P. Milczarski, B. Zieliński, Z. Stawska, A. Hłobaż, P. Maślanka, P. Kosiński, "Machine Learning Application in Energy Consumption Calculation and Assessment in Food Processing Industry", ICAISC (2) (2020), Springer LNAI 12416, 369-379.

[16] Z. Stawska, P. Milczarski, et al., ”The carbon footprint methodology in CFOOD project.” *International Journal of Electronics and Telecommunications*, 2020, 66(4), 781–786

[17] P. Harrington, “Machine Learning in Action.” Manning Publ. 2012.

[18] A.P Dempster, N.M. Laird, D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm". *Journal of the Royal Statistical Society, Series B*. 39 (1), 1977, 1–38