Good Practice
CFExpert
Application of Machine Learning Methods in Carbon Footprint Optimization

Piotr Milczarski
Faculty of Physics and Applied Informatics,
University of Lodz, Lodz, Poland
piotr.milczarski@uni.lodz.pl

BIOSTRATEG

FACULTY OF PHYSICS AND APPLIED INFORMATICS
University of Lodz

LCA4Regions
Interreg Europe

województwo łódzkie

Unia Europejska
Europejski Fundusz Rozwoju Regionalnego

Agenda

- Foundations of the research
- CFOOD project
- Related work
- Production and its Carbon Footprint assessment
- Clusterization of the processes
- Classification and evaluation of the processes
- *CFExpert – a tool to assess CF in the production*
- Conclusions and future work

Summary

- In the research, that our system *CFExpert implements,* 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.

2019 IARW Global Top 25 List

	Company Name	Locations	Cubic Feet	Cubic Meters
1	Americold Logistics	Argentina, Australia, Canada, China, New Zealand, United States	980,172,516	27,755,394
2	Lineage Logistics [i]	Belgium, Netherlands, United Kingdom, United States	964,931,093	27,323,805
3.	United States Cold Storage	United States	311,986,000	8,834,460
4.	AGRO Merchants Group [ii]	Australia, Austria, Brazil, Chile, Ireland, Netherlands, Poland, Portugal, Spain, United Kingdom, United States	255,741,569	7,241,795
5	Nichirei Logistics Group, Inc. [iii]	France, Japan, Netherlands, Poland	180,886,031	5,122,122
6	Kloosterboer	Canada, Germany, France, Netherlands, Norway, South Africa, Sweden, United States	171,182,686	4,847,354
7	NewCold Advanced Cold Logistics	Australia, France, Germany, Netherlands, Poland, United Kingdom, United States	169,504,655	4,799,837
8	VersaCold Logistics Services	Canada	132,510,534	3,752,280
9	Cloverleaf Cold Storage Co. [iv]	United States	128,920,223	3,650,614
10	Emergent Cold Storage	Australia, New Zealand, Sri Lanka, Vietnam	121,470,455	3,439,660

Life cycle assessment

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.

Related work

- Scherer M, Milczarski P. Machine-Learning-Based Carbon Footprint Management in the Frozen Vegetable Processing Industry. *Energies*. 2021; **14**(22):7778. https://doi.org/10.3390/en14227778
- 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
- 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.
- Z. Stawska, P. Milczarski, et al., ”The carbon footprint methodology in CFOOD project.” *International Journal of Electronics and Telecommunications,* 2020, 66(4), 781–786

CFExpert Foundations

- An expert program for counting the carbon footprint (*CFExpert*) was established, with the help of which technological processes are verified in order to minimize the carbon footprint in technologies and products and to assess economic effects. This tool can be adapted to various industries. The processes were gathered by the Leader of the consortium IBPRS in the partner company Unifreeze.
- The main **beneficiaries** of this good practice are **the companies** in the region and in the country where **the production** of goods **has a significant amount of CO_2 equiv per unit**.

Production processes in CFOOD project

In the CFOOD project 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.

Foundation of the research

In the research section, we have tested several of clusterization methods and choose three: Canopy, k-Means (KM) and Expectation-Maximization (EM). 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 its accounting system:

- **Optimal production** – the product has the temperature -25^{o}C ± 2 ^{o}C at the end of the line and the production goes without obstacles
- **Close to optimal** – during the high season energy consumption should be lower so as to make through-output higher. That is why the product temperature is allowed to be between -6^{o}C and -18^{o}C.
- **Wrong accounting** of some parameters. Sometimes, operators of the system can make mistakes. That will result 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

Foundation of the research

In Tables there are the results of the production processes using the following 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 .

Foundation of the research

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.

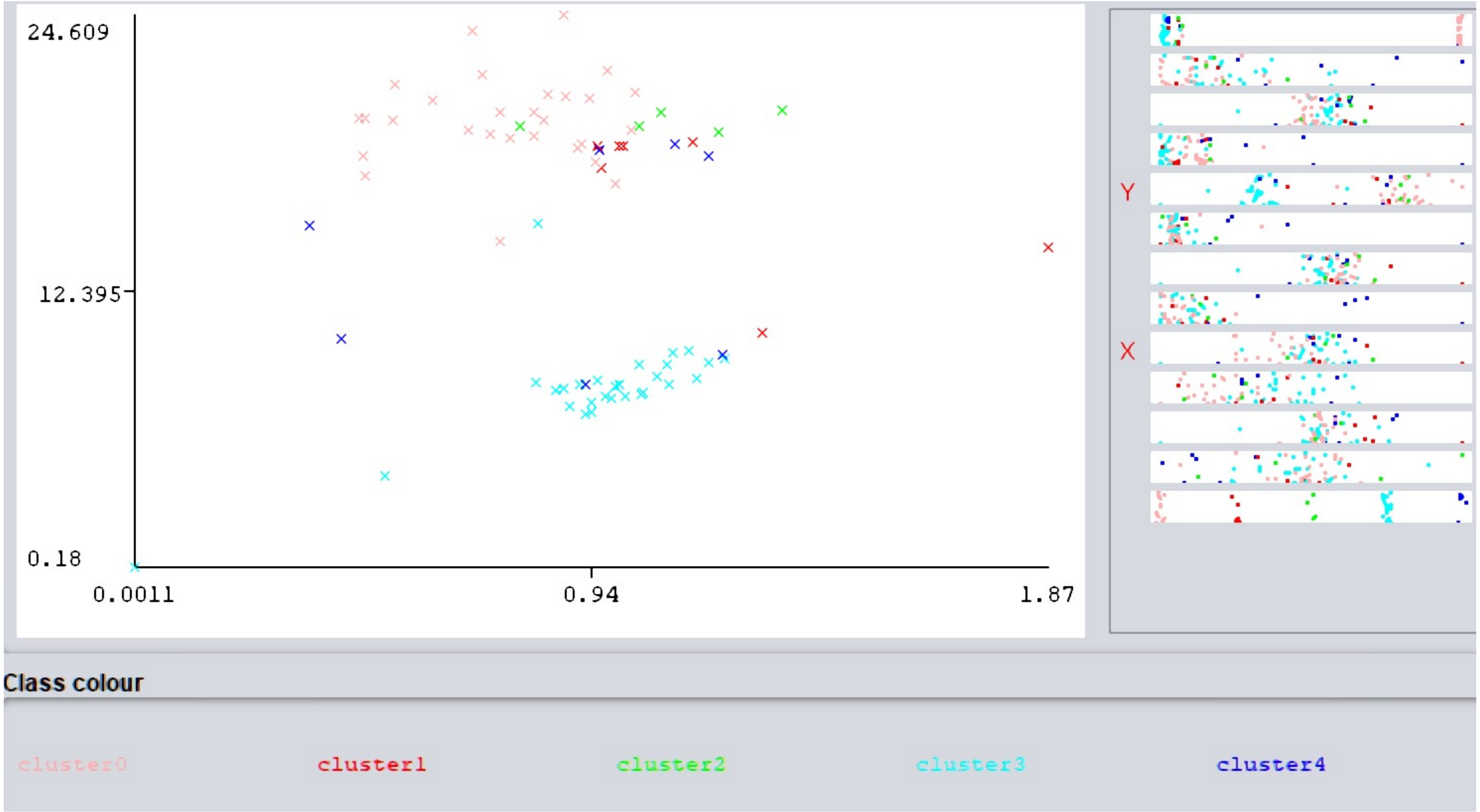
The other vegetables have around 50 cases.

Clusterization results of the onion processes

TABLE III. THE K-MEANS CLUSTERIZATION RESULTS OF THE CENTROIDS PARAMETERS OF THE CHOSEN STAGES.

Cluster	Division	pS1	S1	S3	pS4	S4	pt
0	1	27.82	53.25	14.20	30.38	32.83	1.86
1	3	6.23	3.43	0.47	25.60	15.17	3.45
2	20	0.14	7.16	1.05	0.89	30.56	2.18
3	**97**	**0.15**	**4.79**	**1.05**	**1.01**	**28.95**	**2.34**
4	31	0.06	2.72	0.67	1.16	34.11	2.02

The k-means cluster division for spinach production in the space (S4, S2): where the stage S4(x-axis), S2 (y-axis).



Clusterization results of the spinach processes

TABLE VII. THE K-MEANS CLUSTERIZATION RESULTS OF THE CENTROIDS PARAMETERS OF THE CHOSEN STAGES AND PARAMETERS.

Cluster	Cases	S1	S2	S3	S4	pt	Et
0	28	28.88	19.95	16.91	0.77	1.58	141.8
1	7	39.75	16.82	19.20	1.17	1.69	192.7
2	5	35.23	19.86	18.40	1.09	1.56	143.7
3	**28**	**31.3**	**8.05**	**15.06**	**0.95**	**1.88**	**137.7**
4	7	30.3	14.14	15.14	0.88	1.92	161.7

Clusterization assessment

To assess and to choose the clusterization method we have used five machine learning methods as in our previous work. All the clusterization results were assessed by the classification methods with the same parameters:

- 3NN (kNN) 3-Nearest Neighbors;
- Multilayer Perceptron (MLP) with a hidden layer with 16 nodes for the both productions with a learning rate equal to 0.79 and momentum equal to 0.39;
- 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)

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

Clusterization assessment

TABLE IX. THE ASSESSMENT OF THE CLUSTERIZATION METHODS BY FIVE MACHINE LEARNING METHODS.

Classifier	Classification results [%]		
	Canopy	KM	EM
3NN	90,7	**94.7**	90,7
C4.5	93.3	**97.3**	98.7
MLP	96.0	**94.7**	97.3
RF	100	**100**	100
SVM	100	**98.7**	100

TABLE V. THE ASSESSMENT OF THE CLUSTERIZATION METHODS BY FIVE MACHINE LEARNING METHODS.

Classifier	Classification results [%]		
	Canopy	KM	EM
3NN	98.7	**98.0**	87.5
C4.5	98.0	**99.3**	98.7
MLP	92.1	**97.4**	89.5
RF	100	**100**	100
SVM	96.1	**96.7**	88.8

CFExpert

The research is financed by the Polish National Center for Research and Development, grant CFOOD number:

BIOSTRATEG3/343817/17/NCBR/2018.

IBPRS

INSTYTUT BIOTECHNOLOGII
PRZEMYSŁU ROLNO-SPOŻYWCZEGO
im. prof. Wacława Dąbrowskiego
PAŃSTWOWY INSTYTUT BADAWCZY

UNIFREEZE

UNIFREEZE

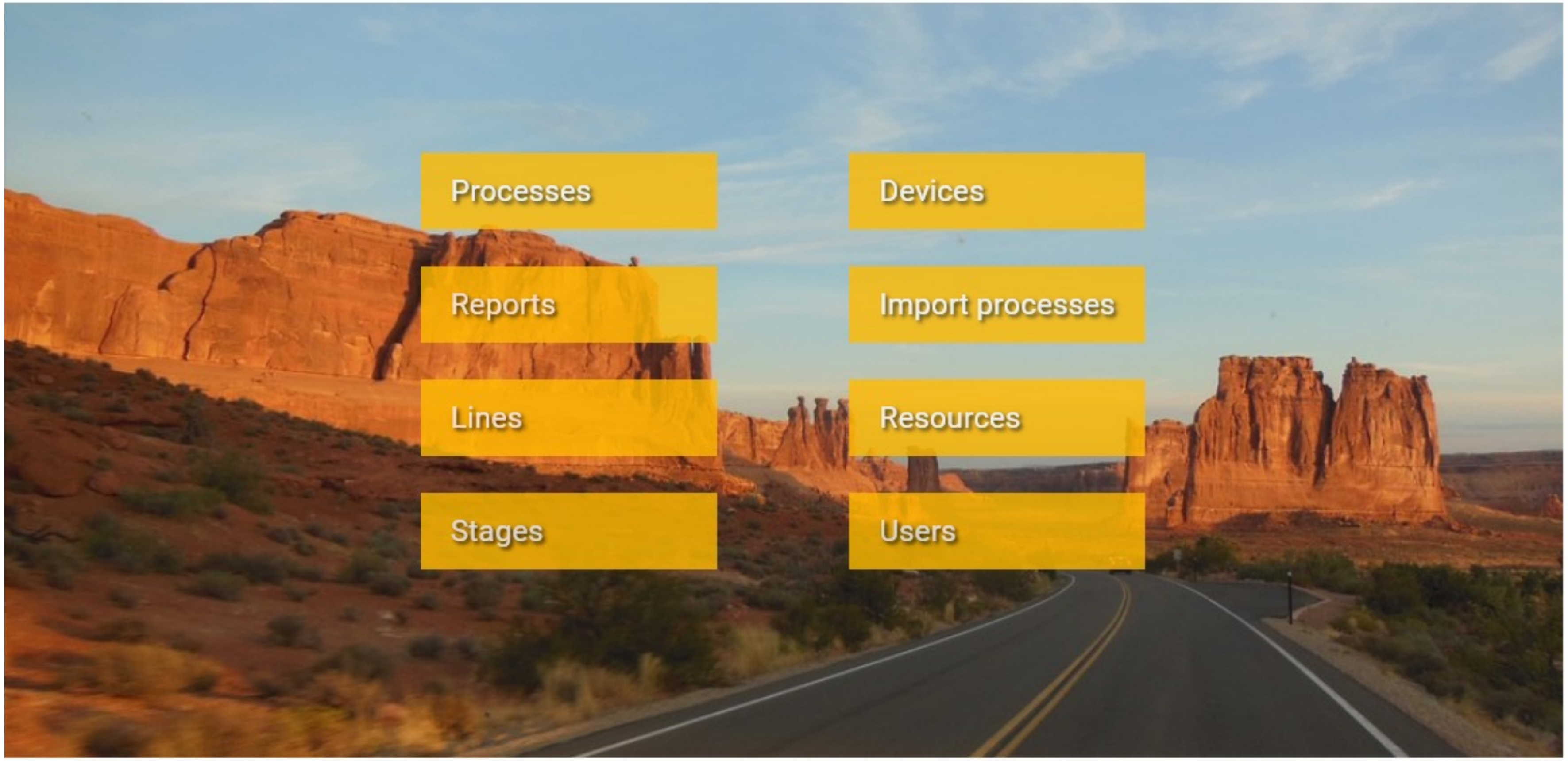
Production processes in CFOOD project

In the CFOOD project 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.

CFExpert



Projekt „Opracowanie innowacyjnej metody obliczania śladu węglowego dla podstawowego koszyka produktów żywnościowych" realizowany w ramach Strategicznego Programu Badań Naukowych i Prac Rozwojowych „Środowisko naturalne, rolnictwo i leśnictwo" – BIOSTRATEG.
Nr umowy: BIOSTRATEG3/343817/17/NCBR/2018.

Procesy
Zapisane raporty
Linie
Etapy
Urządzenia
Importuj procesy
Zasoby
Użytkownicy

Dodaj nowy proces | Procesy dla linii | Linia: cebula uni | Generuj raport | Zwykły

Procesy do raportu	Id	Linia	Waga	Data rozpoczęcia	Data zakończenia	energia [kW]	CO_2 [kg]
☑	Proces 6	cebula uni	79730	2020-02-03	2020-02-04	9084.94	0.09
☑	Proces 7	cebula uni	52010	2020-02-04	2020-02-05	3782	0.06
☑	Proces 8	cebula uni	53240	2020-02-05	2020-02-06	4628.19	0.07
☑	Proces 9	cebula uni	104240	2020-02-06	2020-02-08	11214.43	0.08
☑	Proces 10	cebula uni	32904	2020-02-07	2020-02-08	2836.45	0.07
☑	Proces 11	cebula uni	79730	2020-02-03	2020-02-04	9084.94	0.09
☑	Proces 12	cebula uni	52010	2020-02-04	2020-02-05	3781.98	0.06
☑	Proces 13	cebula uni	53240	2020-02-05	2020-02-06	4628.19	0.07
☑	Proces 14	cebula uni	104240	2020-02-06	2020-02-08	11214.43	0.08
☑	Proces 15	cebula uni	32904	2020-02-07	2020-02-08	2836.45	0.07
☑	Proces 16	cebula uni	12590	2020-02-11	2020-02-11	833.97	0.05
☑	Proces 17	cebula uni	50080	2020-02-11	2020-02-12	4104.9	0.06
☑	Proces 18	cebula uni	47660	2020-02-12	2020-02-13	4266.35	0.07
☑	Proces 19	cebula uni	54870	2020-02-14	2020-02-15	3640.23	0.05
☑	Proces 20	cebula uni	19200	2020-02-17	2020-02-18	1304.31	0.05
☑	Proces 21	cebula uni	83420	2020-02-18	2020-02-20	6423.27	0.06
☑	Proces 22	cebula uni	117940	2020-02-20	2020-02-22	7228.39	0.05
☑	Proces 23	cebula uni	37040	2020-02-26	2020-02-26	2460.08	0.05
☑	Proces 24	cebula uni	34020	2020-03-02	2020-03-03	4000.68	0.09
☑	Proces 25	cebula uni	89850	2020-03-03	2020-03-05	2392.92	0.02
☑	Proces 26	cebula uni	15300	2020-03-05	2020-03-05	1468.63	0.07
☑	Proces 27	cebula uni	16630	2020-03-06	2020-03-06	707.65	0.03

Etapy

Urządzenia

Importuj procesy

Zasoby

Użytkownicy

$0 \le r \le 0.4$ | $0.4 < r < 0.6$ | $0.6 \le r \le 1$

					Ślad węglowy na tonę asortymentu [kg CO_2/t]						
		Średnia	2233.27	148.94	0.07	0.01	0	0	0.06	0	0.07
		Odchylenie standardowe	380.01	43.7	0.03	0.01	0	0	0.03	0	0.03
		Mediana	2255.6	144.46	0.07	0.01	0	0	0.06	0	0.07
ID procesu	Czas rozpoczęcia	Długość [h]	Przepustowość [t/h]	Przepustowość [kg CO_2/h]	Całkowity	komora d04	rampa 02	przed tunelem 08	tunel duży 09	za tunelem 10	energia
11	2020-02-03 08:40	32.33	2465.88	214.95	.09	.00	.00	.00	.08	.00	.09
12	2020-02-04 23:20	20.92	2486.53	138.32	.06	.00	.00	.00	.05	.00	.06
13	2020-02-05 23:45	19.42	2741.97	182.35	.07	.01	.00	.00	.06	.00	.07
14	2020-02-06 22:35	39.00	2672.82	219.98	.08	.00	.00	.00	.08	.00	.08
15	2020-02-07 20:30	14.58	2256.27	148.79	.07	.00	.00	.00	.06	.00	.07
16	2020-02-11 05:20	5.58	2254.93	114.27	.05	.01	.00	.00	.04	.00	.05
17	2020-02-11 22:50	19.25	2601.56	163.13	.06	.00	.00	.00	.06	.00	.06
18	2020-02-12 22:35	19.17	2486.61	170.28	.07	.01	.00	.00	.06	.00	.07
19	2020-02-14 17:20	20.00	2743.50	139.24	.05	.00	.00	.00	.05	.00	.05
20	2020-02-17 22:35	8.75	2194.29	114.03	.05	.01	.00	.00	.04	.00	.05
21	2020-02-18 23:20	31.75	2627.40	154.77	.06	.01	.00	.00	.05	.00	.06
22	2020-02-20 20:35	47.67	2474.27	116.01	.05	.01	.00	.00	.04	.00	.05
23	2020-02-26 00:00	15.92	2327.12	118.24	.05	.01	.00	.00	.04	.00	.05
[illegible]	2020-03-02	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Pobierz jako arkusz Excela

Etapy

Urządzenia

Importuj procesy

Zasoby

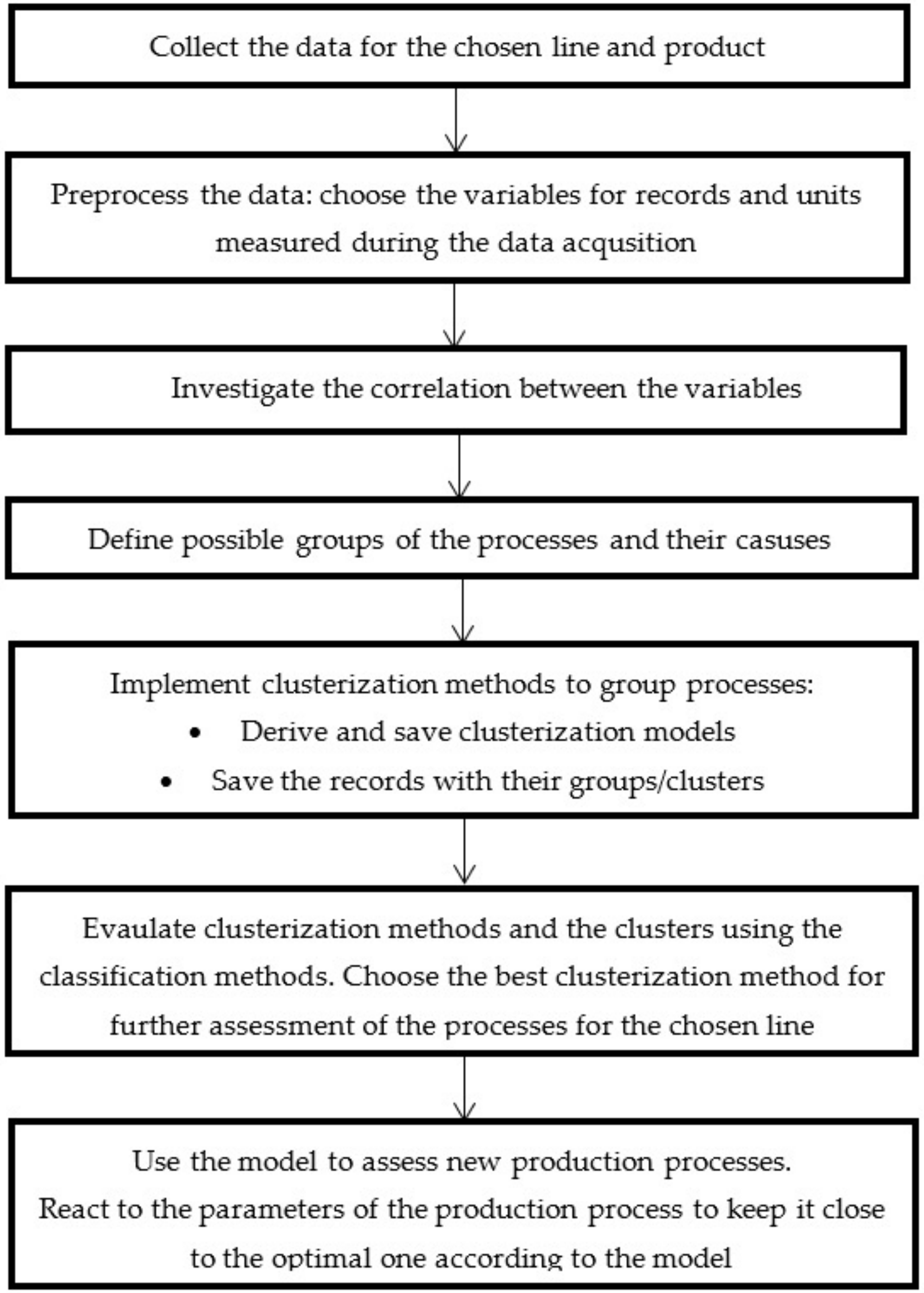
Użytkownicy

Modele			1	2230.05	197.20	.10	.01	.00	.00	.09	.00	.10
			2	2082.90	123.73	.06	.01	.00	.00	.05	.00	.06
			3	2661.40	215.18	.08	.01	.00	.00	.08	.00	.08
			4	2585.30	122.49	.05	.00	.00	.00	.04	.00	.05
			5	1966.08	153.26	.08	.01	.00	.00	.07	.00	.08
						Ślad węglowy na tonę asortymentu [kg CO_2/t]						
ID rocesu	Czas rozpoczęcia	Długość [h]	Id klastra	Przepustowość [t/h]	Przepustowość [kg CO_2/h]	Całkowity	komora d04	rampa 02	przed tunelem 08	tunel duży 09	za tunelem 10	energia
	2020-02-03 07:40	32.33	1	2465.88	214.95	.09	.00	.00	.00	.08	.00	.09
	2020-02-04 22:20	20.92	4	2486.53	138.32	.06	.00	.00	.00	.05	.00	.06
	2020-02-05 22:45	19.42	3	2741.97	182.35	.07	.01	.00	.00	.06	.00	.07
	2020-02-06 21:35	39.00	3	2672.82	219.98	.08	.00	.00	.00	.08	.00	.08
0	2020-02-07 19:30	14.58	5	2256.27	148.79	.07	.00	.00	.00	.06	.00	.07
1	2020-02-03 08:40	32.33	1	2465.88	214.95	.09	.00	.00	.00	.08	.00	.09
2	2020-02-04 23:20	20.92	4	2486.53	138.32	.06	.00	.00	.00	.05	.00	.06
3	2020-02-05 23:45	19.42	3	2741.97	182.35	.07	.01	.00	.00	.06	.00	.07
4	2020-02-06 22:35	39.00	3	2672.82	219.98	.08	.00	.00	.00	.08	.00	.08
5	2020-02-07 20:30	14.58	5	2256.27	148.79	.07	.00	.00	.00	.06	.00	.07
6	2020-02-11 05:20	5.58	2	2254.93	114.27	.05	.01	.00	.00	.04	.00	.05
7	2020-02-11 22:50	19.25	4	2601.56	163.13	.06	.00	.00	.00	.06	.00	.06
8	2020-02-12	19.17	3	2486.61	170.28	.07	.01	.00	.00	.06	.00	.07

Pobierz jako arkusz Excela

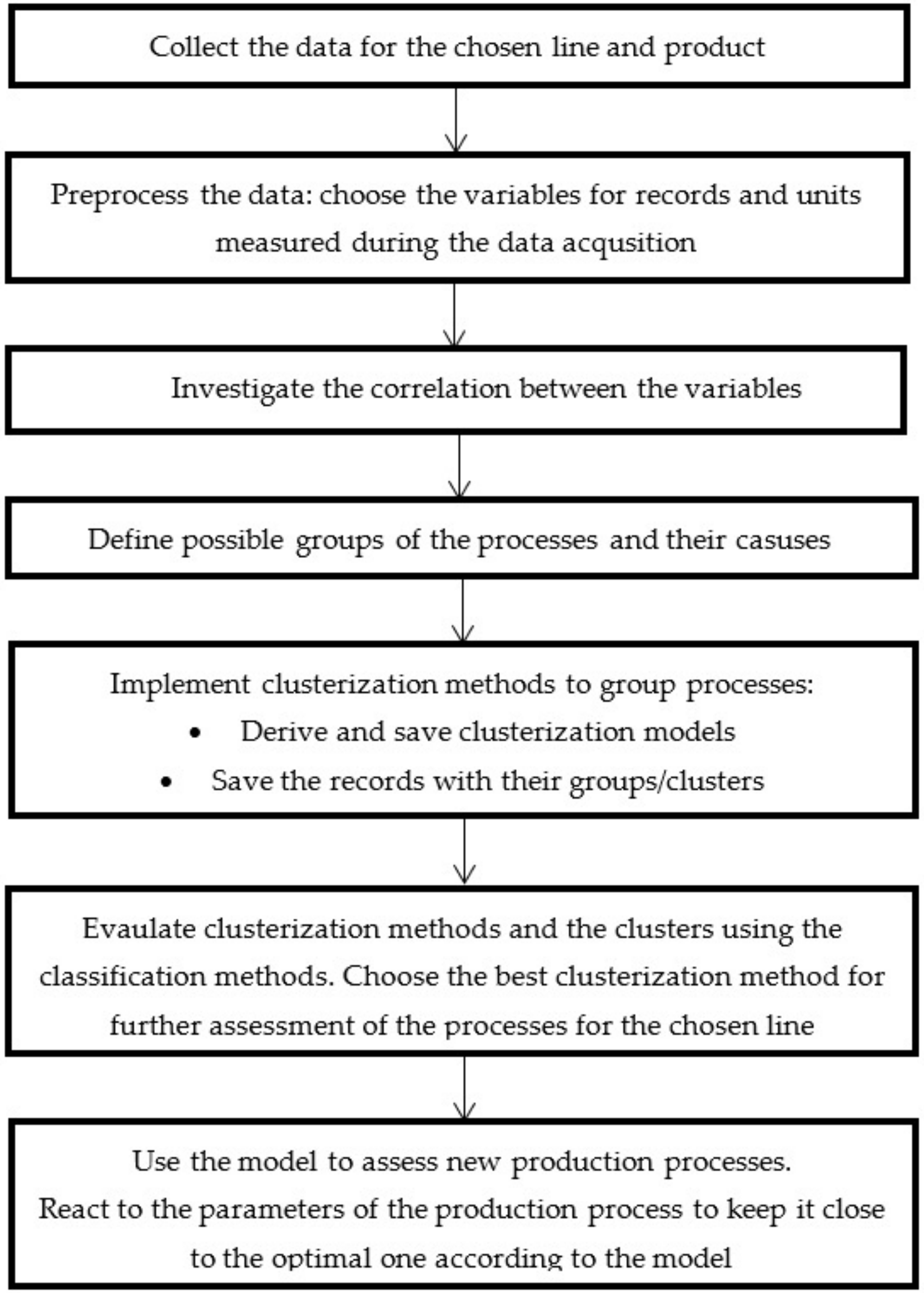
How to use it?

- Collect the data. The data should correspond to the production processes that are stored in the databases, e.g., the energy consumption values from the electric meters, etc. In our case, for the onion production we had five electric power consumption meters. Each of the meters corresponded to one of the production stages S1, S2, ..., S5.
- Preprocess the data to obtain the corresponding parameters, factors, units, and values in order to prepare the dataset for further research. In our case, we recalculated the raw data to obtain the average energy consumption of the production stages, the current average production output, and the average energy utilization in one hour (average power).
- Build the correlation matrices to investigate relationships between the energy consumptions of individual technological processes.



How to use it?

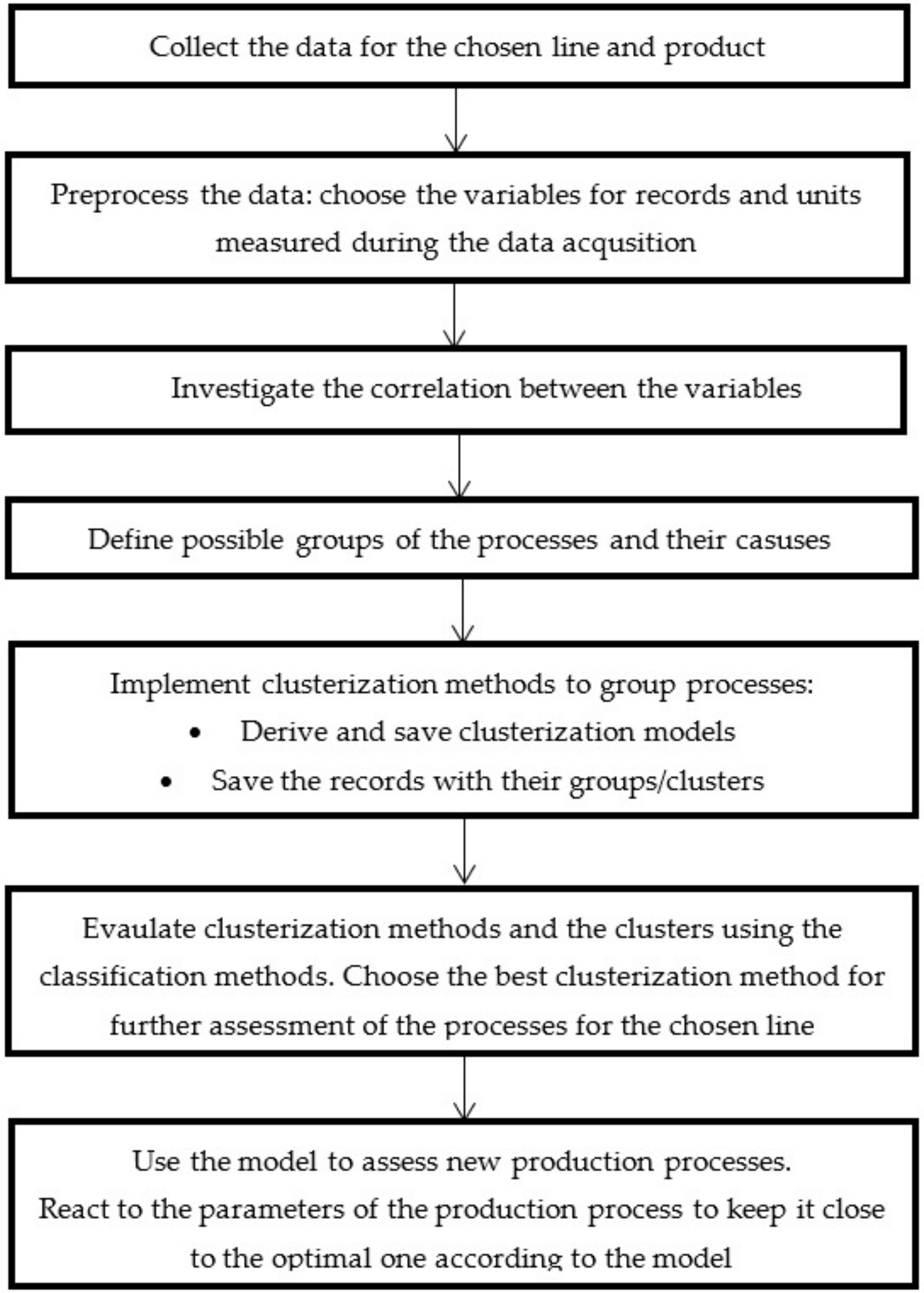
- Use unsupervised learning methods. In our case, we chose clusterization methods. As a result, we obtained different models with the data of the processes divided into a chosen number of categories. Each category (cluster) obtained after clusterization was then evaluated to indicate desired processes with a low CF that did not affect the product quality. Some processes with much lower or higher values of the CF at a given stage than those defined by the technology range were examined by the managers.
- Evaluate the processes with the groups/clusters as classes. Some machine learning supervised methods were applied to the data with clusters as classes. The target of this stage of the research was to choose one or two of the best clusterization methods that could be used by the trained clusterization models to assess the subsequent processes.



How to use it?

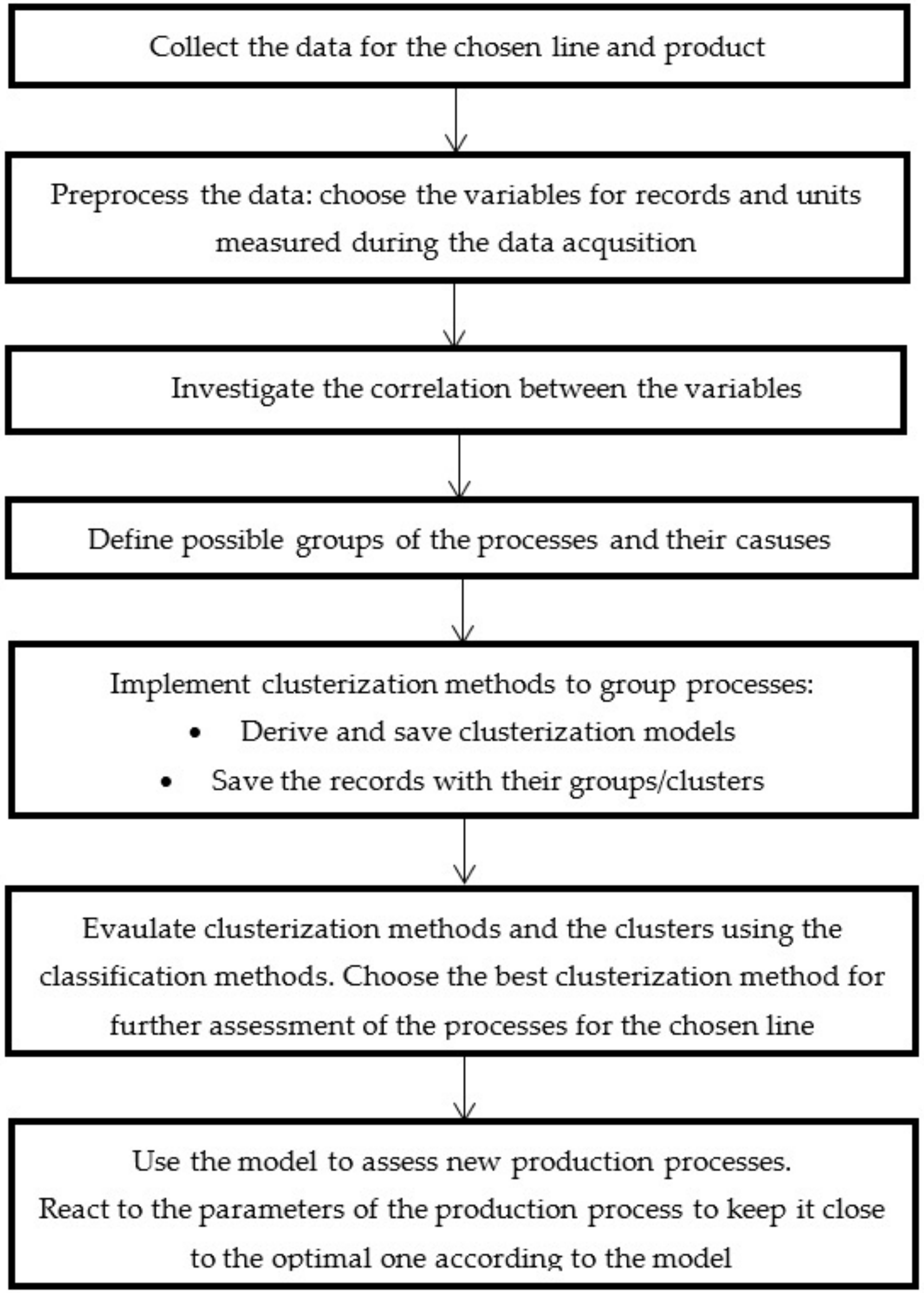
- Obtain the resulting model for the management and optimization of the chosen production line. In our case, for the onion production, the best model was the k-means model. This was stored to be used for the validation of new processes.
- Validate the current production using the obtained model and provide the production manager with a tool to assess the production process for a given category.

The managers may react by setting the optimal parameters, e.g., to lower the production output in order to lower its temperature while maintaining the standard, to raise the power in the freezing tunnel to lower the temperature of the processed materials, or to raise the output when the temperature is lower than required, in order to lower the CF related to production, etc.



Advantages

- This practice can be applied everywhere, in many supply chains and production processes, including in Industry 4.0. It allows to effectively monitor the effectiveness of processes related to the reduction of CO2 emissions.
- Currently the distribution of *CFExper*t software is prepared to be commercialized by the leader of the project. *CFExpert* is flexible to model different processes, products and production.
- The product does not need to be adapted to be used in other type of industry. Its use can be an extremely promising and objective option for R&D, science and technology and certification centers, or regulators such as the Agricultural Advisory Centre, the Office of Technical Inspection, the Energy Regulatory Office, etc.
- It is a tool that can be successfully used to monitor and evaluate processes in terms of LCA/CF. A specific study case can be applied to Farm to Fork processes. Machine Learning & Internet of Things elements are universal and easily transferable to other IT tools.



CFExpert

In *CFExpert,* we present a method of automatic evaluation and optimization of production processes towards low-carbon-emissions products.

The tool supports the management of production lines and is based on unsupervised machine learning methods, i.e., canopy, k-means, and expectation-maximization clusterization algorithms. For different production processes, a different clustering method may be optimal. Hence, they are validated by classification methods (k-nearest neighbors (kNN), multilayer perceptron (MLP), binary tree C4.5, random forest (RF), and support vector machine (SVM)) that identify the optimal clusterization method.

CFExpert

Using the *CFExpert* with real-time production parameters for a given process, we can classify the process as optimal or non-optimal on an ongoing basis.

The production manager can react appropriately to sub-optimal production processes.

If the process is not optimal, then during the process the manager or production technologist may change the production parameters, e.g., speed up or slow down certain batches, so that the process returns to the optimal path.

This path is determined by a model trained via the proposed method based on the selected clustering method.

The method is verified on an onion production line with more than a hundred processes and then applied to production lines with a smaller number of cases. We use data from real-world measurements from a frozen food production plant. Our research demonstrates that proper process management using machine learning can result in a lower carbon footprint per ton of the final product

Conclusions and future work

- The tangible result of the project is a unique expert system for production process optimisation containing three levels of reporting, production scalability, intuitive GUI. The *CFExpert* system is able to quantify & show comparable processes in production, i.e. correct, incorrect or erroneous measurements due to faulty equipment. It has been pilot tested in a manufacturing company that is a consortium member. The system has proven itself in real industrial conditions. Applies to SDGs 9,12,13.
- Although, k-means gives 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.

Acknowledgments

The paper is written as a part of the project CFOOD that is supported by The National Centre for Research and Development, Poland, grant number BIOSTRATEG3/343817/17/NCBR/2018.

Thank you for the attention!

CFExpert
Application of Machine Learning Methods in Carbon Footprint Optimization

Piotr Milczarski

Faculty of Physics and Applied Informatics, University of Lodz, Lodz, Poland

piotr.milczarski@uni.lodz.pl

FACULTY OF PHYSICS AND APPLIED INFORMATICS
University of Lodz