Hungarian Association of Agricultural Informatics
European Federation for Information Technology in Agriculture, Food and the Environment

Journal of Agricultural Informatics. Vol. 13, No. 2
journal.magisz.org

Application of smart visualisation in the analysis of Hungarian Agriculture ICT

Róbert Szilágyi[1], Szilvia Botos[2], János Felföldi[3]

INFO
Received: 10.01.2023
Accepted: 18.01.2023
Available on-line 23.01.2023
Responsible Editor: László Várallyai

Keywords:
digitization, agriculture, ICT, smart visualisation, analysis.

ABSTRACT

Agriculture has a certain specialty in information technology, and the application of new ICT technologies is inevitable. First, we discuss the idea behind data, concepts such as data storytelling, types and proper use of visualisation. There are several well-known visualisation types, but we would like to point out some AI (Artificial Intelligence) based ones (decomposition tree, treemap, Q&A). Our institute has prepared a survey to investigate ICT usage and attitudes of companies in the agri-food sector. The importance of business processes that can be supported by ICT had to be evaluated by the companies. In this study, we evaluated the importance of some ICT indicators. To ensure the use of the mentioned intelligent visualisation, we present some examples from our institute's questionnaire on ICT use in agriculture. We found the smart visualisation usable and the idea of smart dashboards seems to be usable in practice.

1. Introduction

The development of information and communication technologies (ICT), the growth of database systems and digitalization are a hot topics nowadays. It is a great challenge to store, process, and interpret this data (Hermann et al., 2016). Digitalization is not only an ICT development but is also used in many business processes and has an impact on the organization (Horváth & Szabó, 2019). Most published articles deal with digitization only from a technological perspective (Dworschak & Zasier, 2014; Hermann et al., 2016) or from a theoretical perspective (Fettig et al., 2018; Dalenogare et al., 2018). The impact of digitalization is perhaps strongest in the engineering and electronics industries (Demeter et al., 2020; Nagy, 2018; Horváth and Szabó, 2019; Gauger et al., 2017). Digitization in agriculture, especially in the manufacturing and food industries, is frequently addressed in various studies.

In this paper, ICT is taken as a possible management support tool, and we are looking for the answer to what factors and drivers are important in agriculture. The following issues could be examined here:

- what are the main ideas about the visualisation,
- what about the self-ICT skill score (1-5 grade) of the specific sector,
- what about such visualisations like decomposition trees, treemaps and native (English) language based Q&A.

1.1. The methods of the research

Our institute at the University of Debrecen created a questionnaire to survey the ICT usage characteristics of companies operating in the agricultural food supply chain. The data from our survey

[1] Róbert Szilágyi
University of Debrecen
szilagyi.robert@econ.unideb.hu
[2] Szilvia Botos
University of Debrecen
botos.szilvia@econ.unideb.hu
[3] János Felföldi
University of Debrecen
felfoldi.janos@econ.unideb.hu

analysed in this paper were collected in the first quarter of 2022. The respondents were Hungarian companies operating in sectors related to agriculture (producers) and food industry (processors). The questionnaire was structured and the responses were collected by a market research company that contacted the heads or a member of the management of the companies. The sample was selected by stratified random sampling based on statistics from the Hungarian Central Statistical Office (HCSO). Based on the previous research activity of our institute (Fuezesi et al., 2016, Botos et al., 2020), we selected the fruit, vegetable, dairy and meat sectors. A total of 200 questionnaires were recruited for this study. In this article, we focus on showing an alternative visualisation for ICT analysis. To achieve our goal, we selected the food and vegetable sector.

1.2. Data Validation, Storytelling and visualisation

Data quality in the ETL process

The idea of the ETL (Extract Transform Load) process is well-known (Vassiliadis et al, 2002). The accuracy and relevance of data analysis rely on the ability to bring high data quality using the ETL process. The ETL process is complex and time-consuming, but important and unavoidable (Souibqui et al., 2019).

Data Validation

In the validation process, we refer to good-quality data characteristics (Gualo et al., 2021) to assess the quality of the dataset we use. Using characteristic several steps of data validation are performed. The most important steps are the following:

Data Accuracy: In assessing the accuracy of data and initial data exploration was conducted.

Data Completeness: In assessing the data completeness we conducted data sorting to search for missing data.

Data Consistency: To assess the data consistency, data matching and cleaning were conducted using. Only after the data validation process can we start the discussion of the story of our data.

Data storytelling

Most researchers and scholars did not agree on the definition of data storytelling is, many scholars have centered on how the use of visual charts and narratives can support communication and insights (Knaflic, 2015, Lee et al., 2015,). Jeffrey Heer used the terms narrative and visualisation to refer to Data storytelling with a chain of events coupled with visual charts. (Heer et al., 2010).

Data Visualisation Techniques

Various techniques and methods have been used to present insights using data visualisation from standard to more sophisticated and complex techniques that deal with complex data sets (Heer et al., 2010). Simple visualisation techniques include but are not limited to charts, tables, graphs, maps, infographics, and dashboards.

Situating diagrams in graphic communication

A major problem with visual methods in the natural and social sciences is the different terminology used to describe diagrams (Umoquit et al., 2013). Table 1 summarizes the differences and similarities between three categories of graphic communication used as qualitative data collection methods: Charts, Tables, and Lists. Although there may be overlap between the categories, they are discussed as distinct and mutually exclusive categories (Umoquit et al., 2013).

1.3. Machine Learning and Business Intelligence

Data-driven research has been energized by the growth of Big Data analytics and developments in machine learning. Therefore, the machine learning method is one of the possible approaches for creating a Big Data decision support tool. Tsolaki et. al (2022) grouped data mining and machine learning methods by application area.

Table 1. Differences in diagrams and tables and lists (adapted from Umoquit, 2013)

	Diagrams	**Tables and lists**
Purpose	Ability to simplify complex ideas	Highly structured ideas and variables
Abstraction	Some level of abstraction	Concrete
Structure (internal)	More direct in terms of representation; structure inherent where pre-agreed notation is used	Most direct in terms of representation
Notation	May use pre-set or agreed conventions or notations	Uses pre-set or agreed conventions or notations
Spatial Meaning	Spatial arrangements of signs usually carry meaning	Spatial arrangements of signs carry meaning
Verbal/visual spectrum	Composite of both visual and verbal	Verbal signs dominate over visual

AI visualisation is one of the distinctive features of Power BI. Key-influencer visualisation, decomposition tree, and anomaly detection are the three different types of AI visualisation in Power BI. In this research, we will use decomposition trees (Ehrenmueller, 2020, Patriarca et al., 2022).

Implementing machine learning techniques often requires a thorough understanding of mathematics and computer science; however, due to the high level of expertise required, ML is difficult to use. A self-service framework ML can therefore be used by analysts with less technical knowledge (Patriarca et al., 2022). Numerous research papers have addressed the self-service deployment of machine learning for business intelligence using Microsoft Power BI (Patriarca et al., 2022; Aspin, 2021, 2022).

Decomposition Tree Visual: Self-Service Decision Tree

The concept of decision tree is used by decomposition tree visualisation to represent data along multiple dimensions. In the visualisation, the data is summarised in a graph that also allows the user to dive into the data dimensions in any order. (Friedman, 2001).

The decomposition tree in Power BI allows the user to visualise data across multiple dimensions. It automatically aggregates the data and allows the user to drill down into the dimensions in any order. In addition, it is a visualisation with artificial intelligence, so the user can prompt it to find the next dimension and break it down according to the user's criteria. Thank you to this tool, it is possible to perform root cause analysis and ad hoc investigations. (Hart, 2020)

The gradient boosting algorithm uses statistical regression and classification as primary methods for predictive data mining. The idea behind the gradient boosting procedure is to train a system using a mathematical model to maximize a differentiable loss function from a set of random "inputs" and "outputs" (Friedman, 2001). The procedure automatically aggregates data and allows drilling down into selected dimensions in any order. It is an artificial intelligence (AI) visualisation, which makes it a valuable tool for ad hoc investigations. There is an "AI Splits" or "AutoAI" to figure out where to look next in the data (Aspin, 2021).

2. The application of smart visualisation

The first decomposition tree (Figure 1) is about paying for services. The analysis value is the average of the company's ICT skill value (1-5). This value is from the company's IT relevant persons' opinion. They explain by values were the following: Communication, Applying IT experts, Data storage and Marketing. The responders were from the fruit and vegetable sectors as we previously mentioned. We used the so-called auto-AI function (automatic Artificial Intelligence) when we select always the high value (in this case the „yes answer" related way). Except for the communication service all the high value were 5.00, surprisingly the communication was lower 4.61. It seems the companies prefer not to pay for the communication services, or the free services are good enough.

Figure 1. Decomposition tree of Average ICT skill in the aspect of pay for services, own source

Figure 2 is about the different aspects of Average ICT skill (service use). The analysis value is the average of the company's ICT skill value (1-5). This value is from the company's IT relevant persons' opinion. The Explain were the following Production control, Accounting system, Online Datastore, Communication, Marketing application, Resource management, and Logistic application. The responders were from the fruit and vegetable sectors as we previously mentioned. We used also the so-called auto-AI function when we select always the high value. The results are the same values, except for the production control.

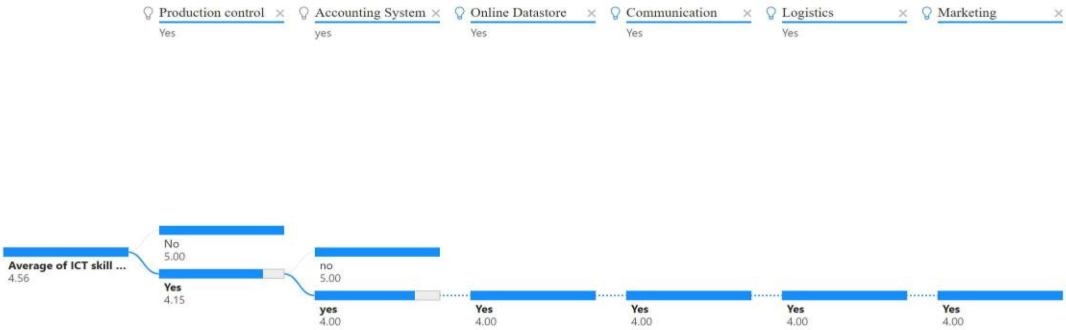

Figure 2. Decomposition tree of Average ICT skill in the aspect of service use, own source

Figure 3 is an example to use a treemap visualisation. Each level of the hierarchy is represented by a coloured rectangle (branch) containing smaller rectangles (leaves). Power BI bases the size of the space inside each rectangle on the measured value. The rectangles are arranged in size from top left (largest) to bottom right (smallest). In this case, we see the whole database. The sector's colour codes and the size of the rectangles make easily understand the relative size of the answers.

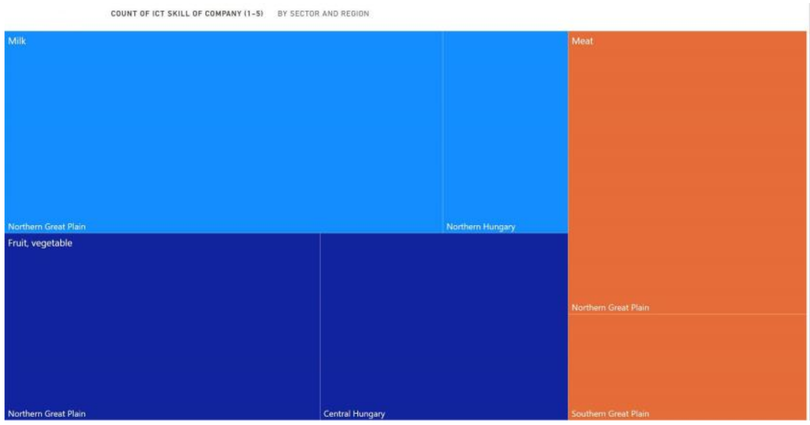

Figure 3. The treemap of the ICT skill in the aspect of sector and region, own source

doi: 10.17700/jai.2022.13.2.681

The Q&A visual gives the user to ask questions about the recent data using the natural (English) language. While we type our questions, the application uses the autocomplete features. The Q&A underlines the word where there is no direct match. In Figure 4 we can see the results for the „*average ICT skill by regions barchart”* written question. The order of visualisation is in descending.

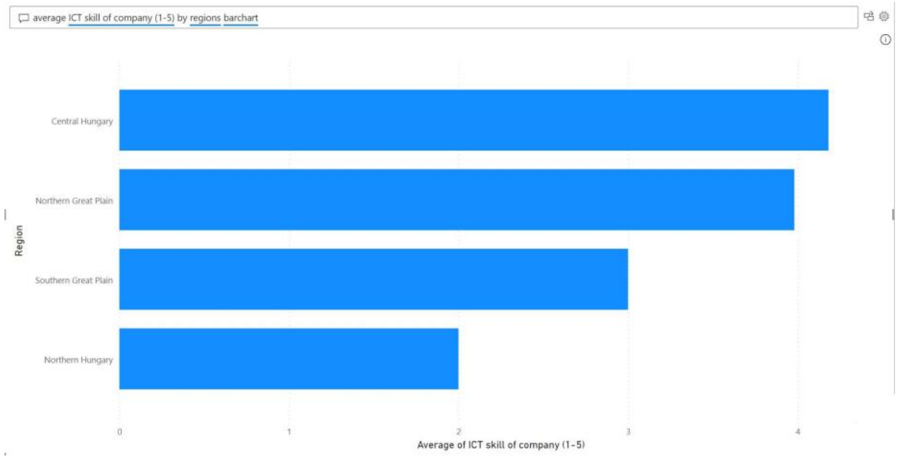

Figure 4. Q&A function example “average ICT skill by regions barchart”, own source

Conclusion

This paper is related to our study analysing the readiness of companies in the food sector for appropriate use of ICT. In the analysis process we have had experience with the Power BI Desktop. There is a group of visual types that are not text-based, charts or maps. Many business intelligence tools have such functionality, and Power BI is no exception. What makes Power BI special is its high use of AI-based visualisations such as the so-called decomposition tree and Q&A. We believe that providing advanced information services to food consumers would be a benefit to businesses, including SMEs. Our proposal is aimed primarily at SMEs, as ICT adoption is critical for them. ICT solutions are widely used for various business activities, and companies can gain a competitive advantage by applying them.

References

Aspin, A. (2021) Miscellaneous Visual Styles. In: Pro Power BI Theme Creation. Apress, Berkeley, CA. DOI 10.1007/978-1-4842-7068-4_11

Aspin, A. (2022) Third-Party Visuals. In: Pro Power BI Dashboard Creation. Apress, Berkeley, CA. DOI 10.1007/978-1-4842-8227-4_10

Botos, S., Szilágyi, R., Felföldi, J. and Tóth, M. (2020) Readiness for ICT Based B2C Information Flow – Case Study of the Hungarian Food Sector, AGRIS on-line Papers in Economics and Informatics, Vol. 12, No. 2, pp. 41-51. ISSN 1804-1930. DOI 10.7160/aol.2020.120204.

Dalenogare, L. S., Benitez, G. B., Ayala, N. F. and Frank, A. G. (2018) The expected contribution of Industry 4.0 technologies for industrial performance, International Journal of Production Economics, Vol. 204, pp. 383-394. ISSN 0925-5273. DOI 10.1016/j.ijpe.2018.08.019.

Demeter, K., Losonci, D. and Nagy, J. (2020) Road to digital manufacturing – A longitudinal case-based analysis, Journal of Manufacturing Technology Management, ISSN 1741-038X. DOI 10.1108/JMTM-06-2019-0226

Dworschak, B. and Zaiser, H. (2014) Competences for cyber-physical systems in manufacturing– first findings and scenarios, Procedia CIRP, Vol. 25, pp. 345-350. ISSN 2212-8271. DOI 10.1016/j.procir.2014.10.048.

Ehrenmueller-Jensen, M. (2020) Adding Smart Visualizations. In: Self-Service AI with Power BI Desktop. Apress, Berkeley, CA. DOI 10.1007/978-1-4842-6231-3_5

Fettig, K., Gačić, T., Köskal, A., Kühn, A. and Stuber, F. (2018) Impact of Industry 4.0 on Organisational Structures, In 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1-8. DOI 10.1109/ICE.2018.8436284.

Friedman, J. H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(5), 1189-1232. http://www.jstor.org/stable/2699986

Füzesi, I., Lengyel, P., Csobán, K. and Szilágyi, R. (2016) Analysis of ICT Support in Hungarian Meat Sector, Agrárinformatika / Journal of Agricultural Informatics, Vol. 7, No 2, pp. 60-69. ISSN 2061-862X. DOI 10.17700/jai.2016.7.2.306.

Gauger, C., Gehres, B., Quinn, M., Schmieg, F. and Xu, G. (2017) Building the digital car company of the future, The Boston Consulting Group, USA, April.

Gualo, F., Rodriguez, M., Verdugo, J., Caballero, I., Piattini, M. (2021) Data quality certification using ISO/IEC 25012: Industrial experiences. Journal of Systems and Software, 176, 110938. DOI 10.1016/j.jss.2021.110938

Heer, J., Bostock, M., Ogievetsky, V. (2010) A tour through the visualization zoo. Communications of the ACM, 53(6), 59. DOI 10.1145/1743546.1743567

Horváth, D. and Szabó, R. Z. (2019) Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities?, Technological Forecasting and Social Change, Vol. 146, pp. 119-132. ISSN 0040-1625. DOI 10.1016/j.techfore.2019.05.021.

Knaflic C.N. (2015) Storytelling with Data: A Data Visualization Guide for Business Professionals, Wiley, pp288, ISBN: 978-1-119-00225-3

Lee, B., Riche, N. H., Isenberg, P., Carpendale, S. (2015) More Than Telling a Story: Transforming Data into Visually Shared Stories. IEEE Computer Graphics and Applications, 35(5), 84–90. DOI 10.1109/mcg.2015.99

Nagy, J., Jámbor, Z. and Freund, A. (2020) Digitalisation in the Food Industry – Case Studies on the Effects of IT and Technological Development on Companies, AGRIS on-line Papers in Economics and Informatics, Vol. 12, No. 4, pp. 77-92. ISSN 1804-1930. DOI 10.7160/aol.2020.120406.

Patriarca, R., Di Gravio, G., Cioponea, R., Licu, A. (2022) Democratizing business intelligence and machine learning for air traffic management safety. Safety Science, 146, 105530. DOI 10.1016/j.ssci.2021.105530

Souibgui M., Atigui F., Zammali S., Cherfi S., Yahia S. B.(2019) Data quality in ETL process: A preliminary study, Procedia Computer Science, 159, 676-687, ISSN 1877-0509, DOI 10.1016/j.procs.2019.09.223.

Tsolaki, K., Vafeiadis, T., Nizamis, A., Ioannidis, D., Tzovaras, D. (2022) Utilizing machine learning on freight transportation and logistics applications: A review. ICT Express. DOI 10.1016/j.icte.2022.02.001

Umoquit, M., Tso, P., Varga-Atkins, T., O'Brien, M., Wheeldon, J. (2013) Diagrammatic Elicitation: Defining the Use of Diagrams in Data Collection. *The Qualitative Report*, *18*(30), 1-12. DOI 10.46743/2160-3715/2013.1487

Vassiliadis P., Simitsis A., and Skiadopoulos S. (2002) Conceptual modeling for ETL processes. In Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP (DOLAP '02), DOI 10.1145/583890.583893