THESIES OF DOCTORAL (PhD) DISSERTATION

IDENTIFYING THE FACTORS INFLUENCING THE ADOPTION OF ADVANCED INFORMATION TECHNOLGOES IN SMALL AGRICULTURAL ENTERPRISES

Mihály Tóth

Supervisor:

Dr. Róbert Szilágyi

associate professor

UNIVERSITY OF DEBRECEN

Károly Ihrig Doctoral School of Management and Business

Debrecen, 2023

Table of Content

1. BACKGROUND, OBJECTIVES AND HYPOTHESES OF THE RESEARCH

Due to the globalisation, increasing market competition and rapidly changing customer demands in the 21st century a number of changes are required regarding to the economic processes, both from a production and a business perspective. The amount of resources available is limited and therefore the practice of using them is difficult to sustain (BENOTSMANE et al., 2019). Sustainability is a key issue in agricultural production at environmental, economic and socio-economic levels (AGOVINO et al., 2019). The **concept of Industry 4.0** has become increasingly more recognised during the period of the research, which offers new opportunities in these areas by conceptually joining different aspects of data collection, data management, data analysis and process control, that can be applied to gain a complex understanding of each process, using large amounts of measurable (sensor) and administrative (economic and market) data, thus facilitating the identification and application of unknown relationships between factors. A gap can be observed between market leaders and laggards (JEFFERY, 2010), as the adoption of data-driven approaches that enable better decisions remains the prerogative of larger companies (REJIKUMAR et al., 2020), even though the implementation of solutions is not necessarily a matter of economic size or revenue. Based on current surveys, only 3.1% of the target group plans to use plant monitoring based on own equipment, while 2,2% using a service, even though it is considered the most well-known way of implementing measurement technologies. The lack of interest is not primarily due to high costs, but to a lack of a sense of need and necessary knowledge (KSH, 2020a). Having the right knowledge and infrastructure within the organisation is an essential element of the feasibility of adaptation. The use of basic digital tools is found to be almost linearly related to farm size class, with 24% of farms classified below 4 STÉ (standard production value) and 96% of farms classified above 500 STÉ (KSH, 2020c). According to the Farm Structure Survey, 53% of farms are below 4 STÉ and 78% below 15 STÉ, thus representing a significant share (KSH, 2020b). By comparing these two factors, it can be concluded that small and medium-sized farms are lagging in terms of general IT solutions, not to mention the adoption rate of the more complex systems, which are the topic of this research. Assessing the factors influencing the adaptation of data acquisition, data management and data analysis related solutions for small and medium sized farms is a common research topic (ZAMBON et al., 2019), with the aim of increasing efficiency and sustainability (SARKER et al., 2019).

The diffusion and rate of adoption of relevant tools and solutions could provide substantial benefits for both the organization and the community, regardless of their scale. However, the issue in practice is the low level of intention to use, which can also be observed among the representatives in the sector (arable crop production). The question arises as to what influences the intention of the adaptation in a greater extent. Related research focuses on the issue of technological acceptance, which seeks to find answers to the questions by adapting to the field. An emerging problem in the *literature* is the generality and inconsistency of related international surveys. The studies were typically assessed using structural equation modelling (PLS-SEM), consisting of the set of latent variables and the model structure. However, the issue of generality and inconsistency limits adaptation to the specificities of the sector. A complementary activity of mine is mainly focused on the development of related IT platforms (measurement tools and information systems based on the requirements of agricultural production) provides practical insight (TÓTH - DÉR - et al., 2019; TÓTH - FELFÖLDI - et al., 2019; TÓTH - SZILÁGYI, 2017) into the opportunities. The formulated objectives, regarding to the two main phase of the research are shown in Figure 1, supporting the identification of the factors that influence intention to use.

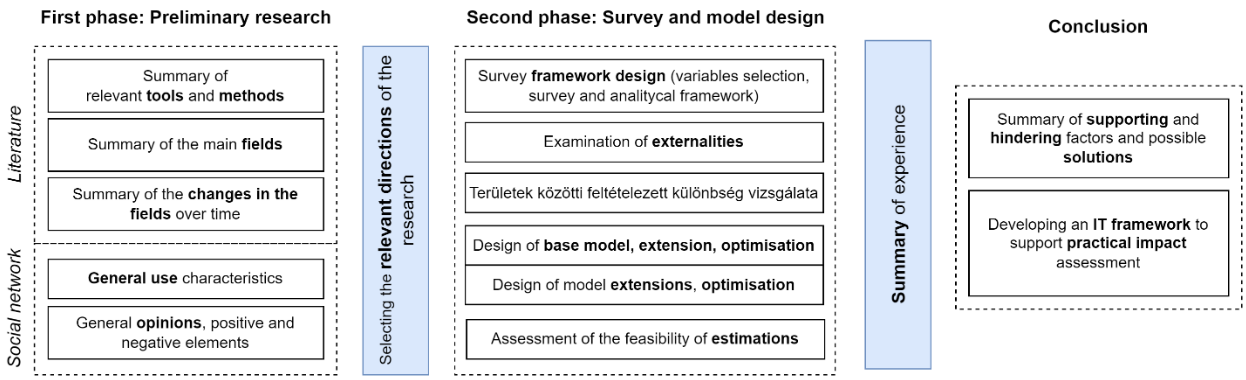

Figure 1: Summary and breakdown of objectives

Source: Own figure

1.1. Research design and process

The framework developed in the research, which includes the steps needed to achieve each *objective*, is presented below (Figure 2).

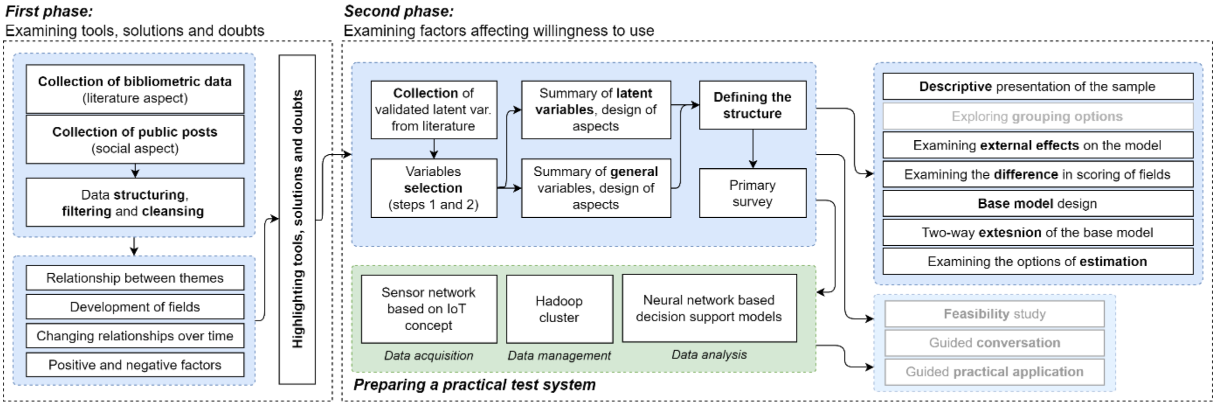

Figure 2: Schematic structure of the research

Source: Own figure

The first phase of the research begins with the implementation of *quantitative research field analysis*, which serve as a preliminary study, including **analysis of bibliometric** and **public entries** to provide understanding of topics, areas, and opinions. Based on the results, the second phase was developed, involving the first steps of **variable selection**. Subsequently, the **content**, **tools** and **methodological framework** for the primary data collection were developed, with the aim of **developing model variants** to support the assessment of the factors influencing the intention to use IT tools and solutions under the topic of smart agriculture. The latter process was conducted at several levels, preceded by supporting analyses, including the **identification of differences in scoring** by the users, also the **determination of external effects** on model variables. To support the procedures, a *survey* was developed based on a custom application. To lay the groundwork for a continuation, a practical **system concept** was developed, to investigate the relevance of the factors that influence the use of the tools and solutions, providing a basis for investigation in the future, regarding to the effect of the application on the user attitude and on the involved processes.

1.2. Formulated research questions

Research question 1: In what form do relevant tools and solutions appear within the field of agricultural production, regarding to their relationships, associations, and opinions, based on literature and public entries? Exploratory research supports the identification of factors relevant to the research. The objectives include a quantitative review of the literature to assess areas of focus, major themes, relationships and changes over time, while the processing of relevant posts in the social media provide insights into the relevant themes and their reception, identifying positive and negative perceptions of each aspect.

Research question 2: Do participants have different opinions on the fields related to relevant tools and solutions? In many cases, the difference between technological solutions is not considered from a user perspective, despite functional differences, that can influence the focus and the need for specific architecture of the model to be developed.

Research question 3: Is it possible to determine the impact of the general variables on the assessment of the latent factors that are to be included in the model? In order to investigate segmentation, it was considered to investigate external effects, sorted by general aspects. The scoring of the latent variables in the base model is often influenced by external factors, that can affect the evaluation of the model.

Research question 4: Is it possible to develop theoretical models that can explain the factors affecting the acceptance and use of relevant tools and solutions based on the selected latent variables? The objective includes the *development and two-way extension (optimisation)* of model variants supporting the assessment of usage patterns, attitude and intention to use relevant tools and solutions. This also involves the assessment of the factors influencing them, considering the specificities of the sector, highlighting the potential for application in decision support, thus supporting gaps in the literature.

Research question 5: Can factors affecting the acceptance and use of tools and solutions be estimated? Based on the models developed, it may be important to estimate the factors directly affecting intention to use (dependent parameters of the model) considering the selected variables for further simplification of survey after this research.

2. APPLIED DATABASE AND METHODS

The following chapter describes the main methods and data sources used in the research. Several customized solutions have been developed to achieve each objective, but the focus is now on commonly available approaches.

2.1. Data and methods used in the first phase of the research

The **secondary data** used for the baseline survey originate from two sources. In the case of the *bibliometric viewpoint*, the metadata is obtained from the Web of Science database, while in case of the *social viewpoint*, public entries from the Twitter social network were obtained, after determining the relevant logical sets, using a custom application, developed for the research (Figure 8).

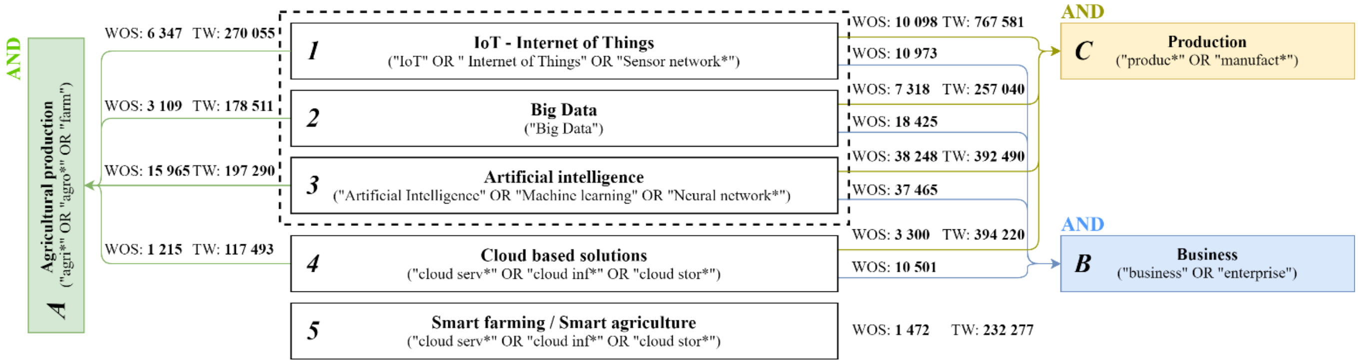

Figure 3: Applied query and data volume of secondary datasets after cleansing

Source: Own figure

The custom application was capable of appropriately structure, filter, cleanse and analyse the data, by the creation of a pipeline, determination for duplicate items (exact and partial matches), tokenisation, lemmatisation, as well as handling synonyms and misspellings using dictionaries developed during the research (Figure 3). In case of the literature, the first step was the keyword **co-occurrence analysis** (BATAGELJ - CERINŠEK, 2013), visualized using a force-directed graph (KAMADA - KAWAI, 1989). The subsequent analysis aims to map the **conceptual structure** of the framework based on the occurrence of keyword usage in the data set, performed using multiple correspondence analysis (MCA) after dimensionality reduction (ARIA - CUCCURULLO, 2017). **Thematic evolutionary** analysis is able to use co-word network analysis and cluster analysis to identify changes occurring at specific points in specific points of the time-series bibliometric data (COBO et al., 2011).

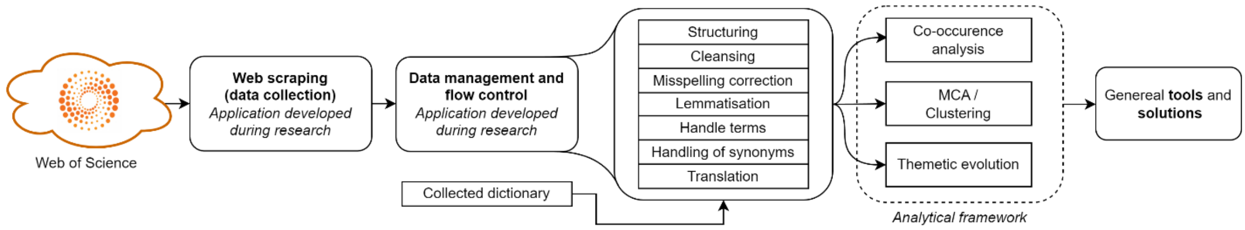

Figure 4: The developed workflow for the bibliometric analysis

Source: Own figure

In order to apply **sentiment analysis** to the public entries, a neural network of associated transformer layers was implemented, based on the BERT model with a characteristic (12-layer transformer) structure (DEVLIN et al., 2019). The entries classified in a given category were used to determine aspects by applying grammar rules (dependency), based on the Standford scheme (DE MARNEFFE et al., 2014). To achieve the objectives, the previously presented application was modified (Figure 5).

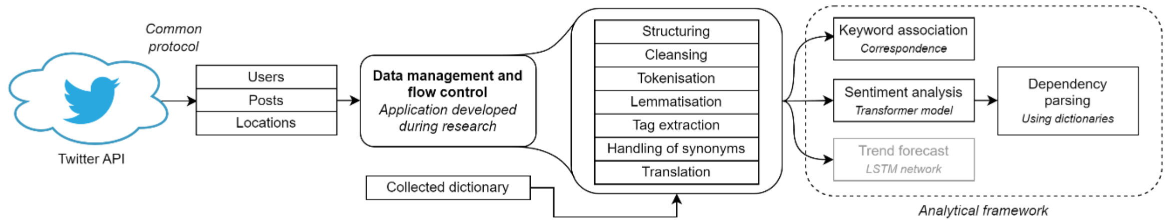

Figure 5: The developed workflow for the bibliometric analysis public entries

Source: Own figure

2.2. Data and methods used in the second phase of the research

As regards the primary data collection, the **survey** is one of the main elements of the research, targeting decision-makers and employees of small and medium sized farms engaged in agricultural activities. The variables are divided into several groups for efficient data management. Figure 6 shows the structure of the questionnaire, its sections, and the correspondences between them.

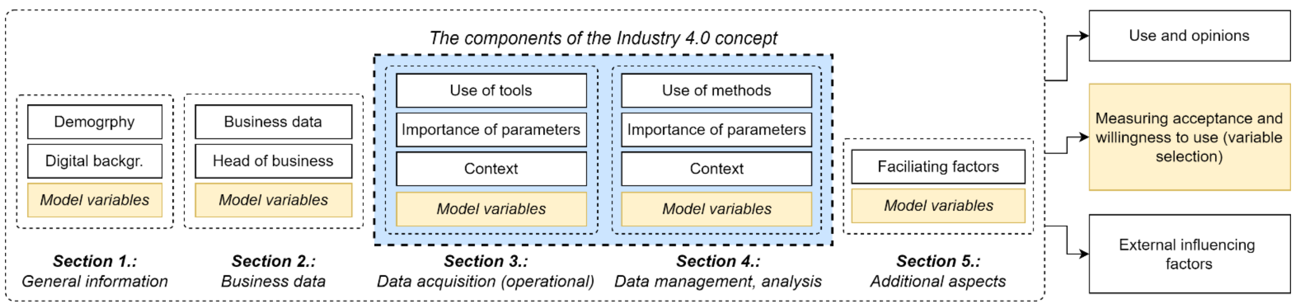

Figure 6: Main topics covered by the survey

Source: Own figure

The primary data collection was conducted through a questionnaire survey, published on the internet as part of a website presenting details of the research, thus increasing the information available to the respondent about the area. A **web application** (http://agrinfo.abadi-major.hu/) was developed to implement the custom solutions. Because of the volume, it provided an important feature to achieve a hierarchical design. The **data collection period** was due between 01.02.2022 and 01.04.2022, during which 172 surveys were completed, but the number of records was reduced to 135 by screening. Publication was done through personal contacts, professional groups, and advertisements. The approximate sample size for the PLS-SEM method was calculated apriori (FAUL et al., 2007), which resulted $n = 123$ records as an acceptable sample site, considering 90% confidence level, assuming an effect size of $f^2 = 0.15$. Due to the exploratory nature of the research, the 90% confidence level is considered acceptable (CONROY, 2015) The typical **sample size** that is used for calculations was $n = 135$ records after screening, however, due to the dynamic nature of the questionnaire, fewer records can be expected for the general variables. The methods outlined in Figure 8 below were used to achieve the objectives.

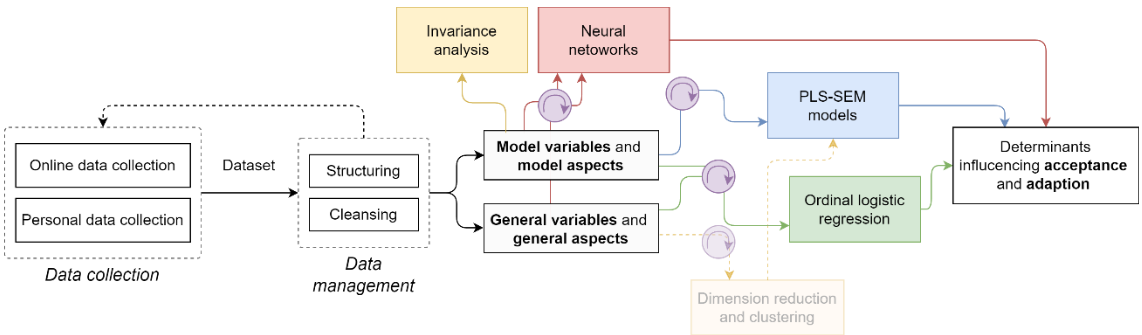

Figure 7: Methods and outputs used for survey

Source: Own figure

Difference between scores was assessed using **invariance analysis**, while the impact of general variables on the scoring of the model variables is determined using **ordinal logistic regression**. After assessing the general effects, **SEM** (Structural equation modelling) is applied. To estimate the dependent variables in the developed model, a multilayer neural **network based** was used based on the models. The shapes in purple represent the **iterative operations** to search for the optimal structure, which was performed using the application created in this research to support the selection of the most applicable variant in this case.

3. MAIN FINDINGS OF THE DISSERTATION

The first phase of the research involves a **quantitative research field analysis** based on literature and public entries. Considering the result, the second phase was developed, the first of which is the variable selection. This was followed by the development of the **content**, **tools** and **methodological framework** of primary data collection, with an aim of **creating**, **developing** and **extending** the model variants supporting the study of the factors influencing the intention to use of tools and methods based on smart agriculture, by using an appropriate **survey**.

3.1. Identifying key areas and topics based on the literature

The three main variants, including the relationship between the characteristics of the Internet of Things (data collection), Big Data (data management) and artificial intelligence (data analytics) and agricultural production (variant *"1A"*, *"1B"* and *"1C"*) have been interpreted in advance, but experience shows that the intermediate area of data management adequately expresses the main characteristics of the other two areas. Therefore, in order to avoid redundancy, the variant expressing the area of data management (*"2A"*) will be explained following the extension of the common areas. For general production, the emergence of related themes can be observed from the year 2014, while for agriculture from the year 2015, including the topics of models, impact, yield, estimation, climate change and framework. For both aspects, we can observe all three fields as relevant to the research, which further reinforces the overlap between them (Figure 13).

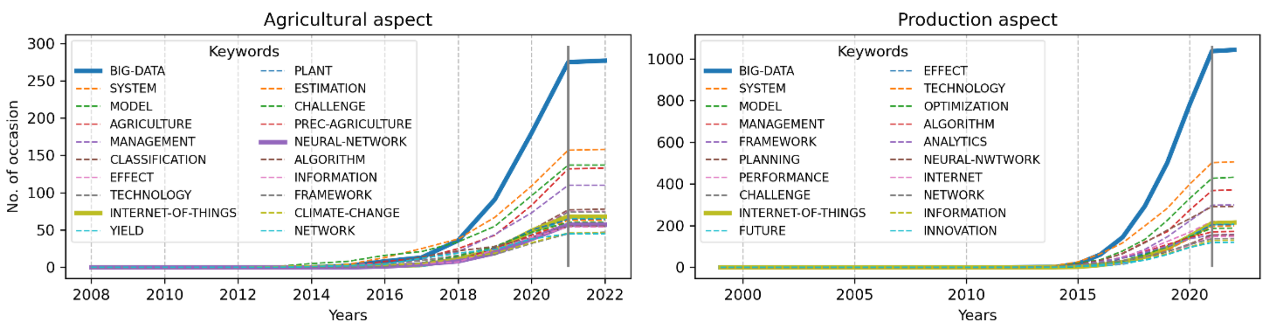

Figure 8: The emergence of data management

Source: Own figure

The analysis of co-occurrence presented below aimed to examine the themes defined by the keywords and the relationship or distance between them (Figure 9), based on current publications. The network structure shows the most frequented keywords, including the

Internet of Things concept, *sensor networks*, *agricultural production*, *Big Data* concept, *precision agriculture* and *artificial intelligence*. The full thesis also provides a detailed overview of the main interconnection points through network analysis, taking into account the PageRank values of the keywords. The items in the figure include specific **tools** and **solutions**, but also related topics that may express areas of **application**, **opportunities**, **benefits**, and **doubts** or **problems**.

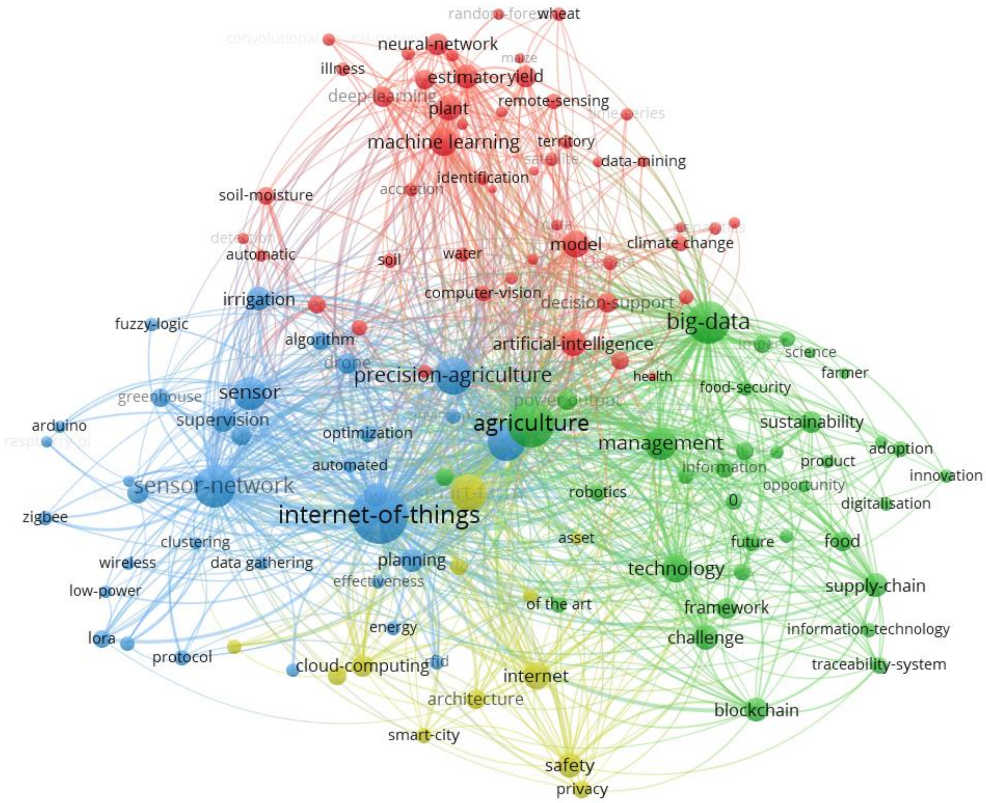

Figure 9: Network of the main topics in the aspect of agricultural production

Source: Own figure

The aim was then to define and make sense of the areas covered by the themes, resulting in relatively distinct groupings (Figure 10). The analysis resulted in 8 groups, reflecting the main areas. By highlighting the larger clusters, we can observe the keywords summarising the more general themes in the clusters highlighted in *red*. For the cluster next to this grouping, marked in *blue*, keywords related to sensor networks summarising technically relevant topics related to the Internet of Things concept have been classified. The grouping in *green* summarises some data management topics, but the options also include topics related to data analysis. I would like to highlight the emergence of the Hadoop ecosystem as a commonly used environment in business. The grouping above it, highlighted in *yellow*, includes some elements of data analytics underpinned by neural networks, with technical factors listed. Additional operations closely related and relevant to machine vision in the sector are included here. The smaller nearby located cluster, coloured in *pink*, covers more practical aspects of machine vision.

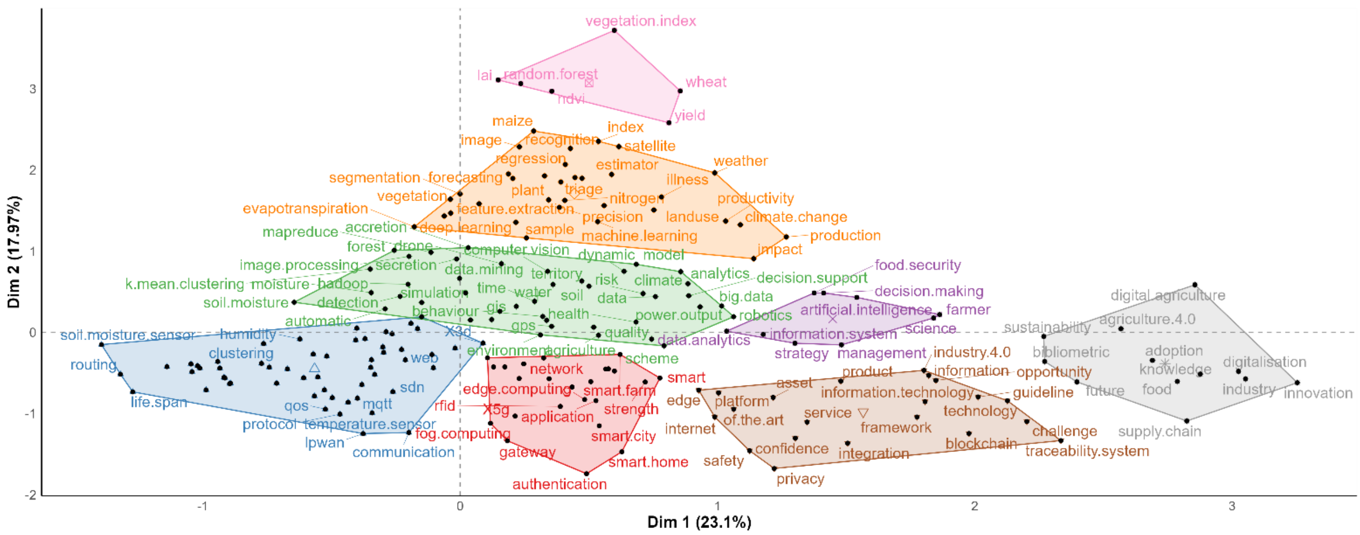

Figure 10: The main fields based on the relevant topics

Source: Own figure

The thematic evolution aims to examine how the relationship between the topics has changed over time, or to identify which topics have contributed to the emergence of the fields (Figure 11). At the first breakpoint, *sensor networks* were already in use, but the concept of the *Internet of Things* emerged through the increased ubiquitous influence of *wireless technology* and *information technology*. A number of new topics have emerged, including applications in *decision support*, the emergence of *data connectivity standards*, and energy-efficient operation. We are able to determine the emergence of the *Hadoop framework*, which is almost essential for the efficient implementation of distributed data storage, given the increasing volume of measurement data alongside business data. The next period saw the emergence of *machine learning* and hence new possibilities for data analytics, until finally, today, we have a more streamlined approach with the components that these have brought together. Today's biggest challenge raises the question of a *framework* that enables the integrated application of these components in decision support.

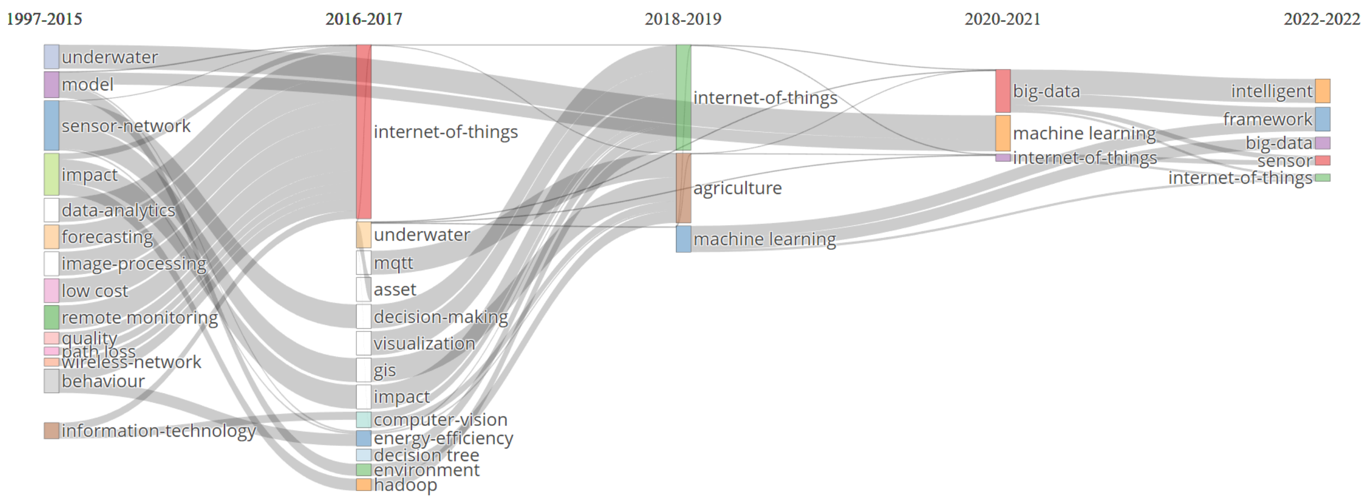

Figure 11: Thematic evolution of the main topics

Source: Own figure

3.2. Determining the community's views based on public entries

The aim of the analysis of public entries is to provide an overview of the community's opinion by extracting the positive and negative factors and aspects contained in the posts, classified according to sentiment analysis, with attention to the doubts. To follow the previous structure, only posts related to data management are presented, which have been present on social network since the beginning of the period (2013) under study (Figure 12). Overall, 41.5% of the posts expressed a positive, while 3.1% expressed a negative opinion. The prevalence of positive, neutral, and negative opinions, their reliability (weight) and the number of positive, neutral and negative opinions per date are shown in Figure 12 below.

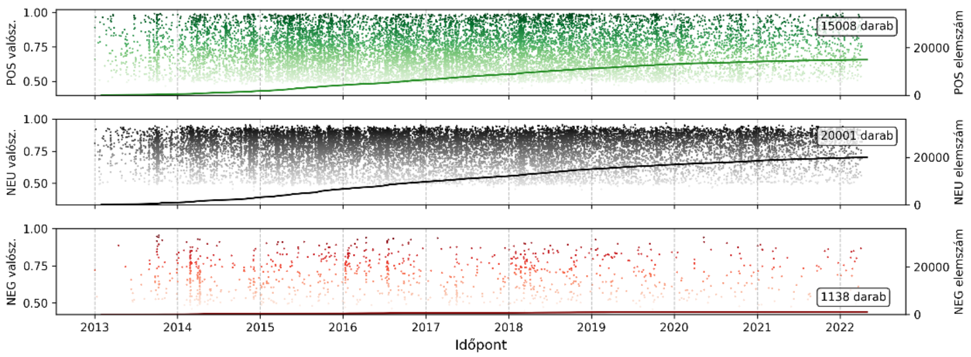

Figure 12: Number and classification of posts published regarding to the field

Source: Own figure

By defining aspects, it was possible to unravel 14,891 coherent grammatical structures, regardless of the polarity of the opinion. A two-level network was then constructed, expressing the relationship between *signifiers* (green shapes) and *objects* (blue shapes) (Figure 13) by constructing a co-occurrence matrix. The entries expressing positive opinions during the review included *low cost*, *high coverage*, *vertical integration*, *innovative solution*, *environmental impact*, *new platform*, *high yield*, *high quality*, *good practice*, *high performance,* and *operational efficiency*. Conversely, negative factor combinations include new *cyber-attacks*, *profitable metrics*, *long time horizon*, *small economies*, *data protection*, *complexity of use*, *high cost*, *scalable solutions* (their presumed absence), *low data security*, *backward compatibility*, *conditional accuracy,* and *unprecedented pressure* (presumably on the need for use).

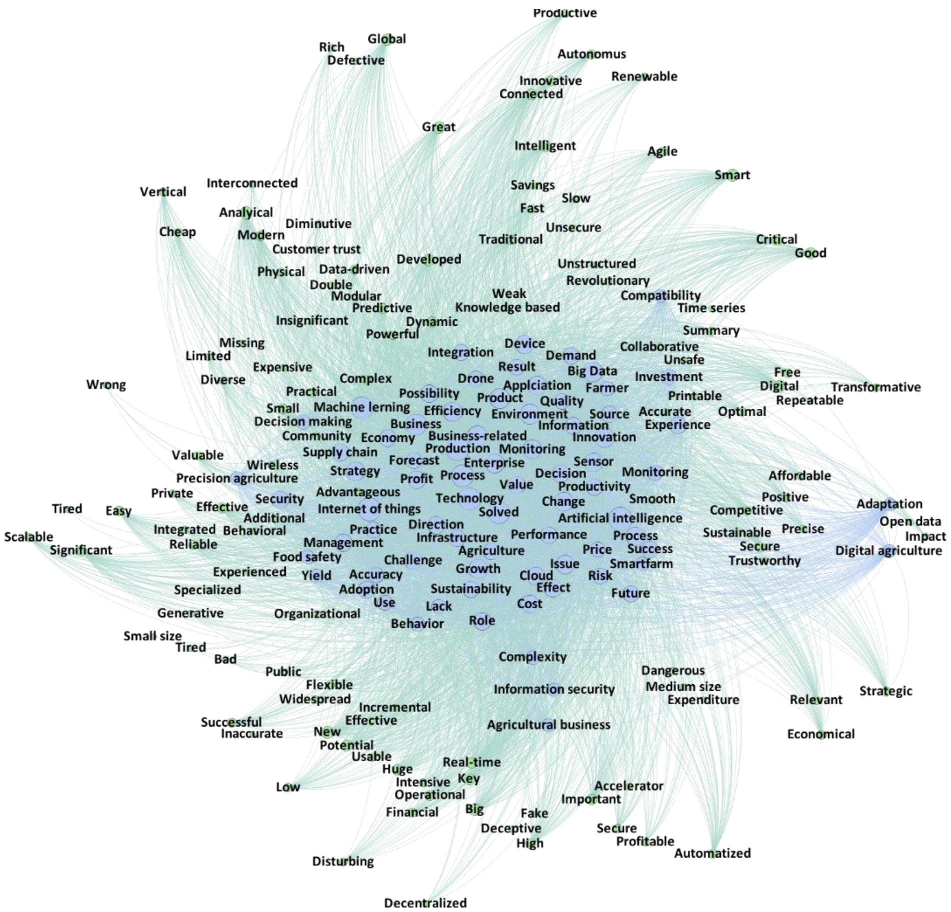

Figure 13: Main topics and their indicators based on the joined data sets

Source: Own figure

3.3. Supporting the analysis of intention to use and influencing factors

Based on the results presented in the previous chapters, a set of criteria was summarized to select the most relevant points from the literature and public entries (Figure 19). This allowed the focus points to be highlighted in the survey to be developed, taking into account the main **tools** and **solutions** emerging from the literature and the **doubts** expressed in the community contributions, thus supporting the assessment of the intention to adapt.

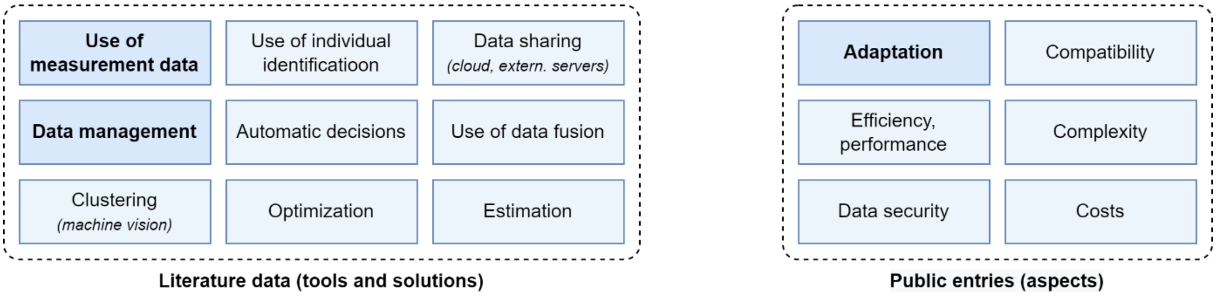

Figure 14: Comparison of literature and community posts, selection of priority factors

Source: Own figure

A questionnaire-based **survey** was designed and conducted to obtain the data needed to develop a model to support the identification of the *application characteristics*, *acceptance*, and *drivers* of the relevant technology.

3.3.1. Selection of relevant variables

The selection of variables relevant to the sector and the territory is a key task of the research. This was carried out in **three stages**, the main stages of which are summarised in Figure 15 below. The *first step* was a review of the literature on relevant research (survey of the intention to use IT tools in agriculture). As a result, 5,520 latent variables and 198 PLS-SEM model variables were summarised, along with their specific results (weights and model fit), in the form of a non-relational database. Subsequently, as a *second step*, the results were summarised by content in order to select the latent variables to be included in the survey. In a *third step*, the models were constructed based on the survey and the set of latent variables used was reduced further based on the results, thus achieving the objective of model specialisation (Figure 15).

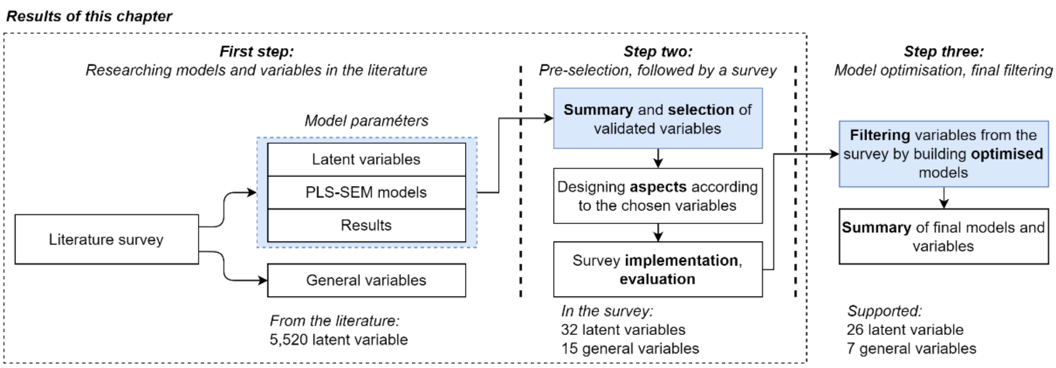

Figure 15: Selection process of the applied variables

Source: Own figure

In terms of variables, a logical distinction is made between **general variables** and **model variables**. For the general variables, the aim was to assess factors that were hypothesized to directly or indirectly influence the scoring the model variables, explaining the segmentation in the sample. Since the survey was conducted in increased detail to assess each factor in the form of the general variables, I tried to aggregate them by constructing aspects, thus a personnel aspect *(A1)*, a general technology aspect *(A2)*, an organisational characteristics aspect *(A3)*, and a practical aspect of data collection and management *(A4 and A5)*. The primary purpose of using **model variables**, i.e., latent variables, is related to the model that will be developed as a result of the research. The variables of the UTAUT2, TAM3 and TOE models have been used in the design of the model structure, but the set have been extended by the implementation of other latent variables to meet the specificities of the field. In developing the model, no relevant literature on the subject was available at the time of the summary, thus alternative fields were reviewed. As many literatures treat the Big Data concept (as data management) and artificial intelligence (as data analysis) as related areas, in contrast to the practice, a similar model structure has been developed. For the model variables, a 5-point Likert scale was used, as it is typical for models measuring technology acceptance (HOLDEN - KARSH, 2010). The principles for grouping the latent variables were defined empirically, thus forming 8 model aspects. Aspects can be further sorted into primary and secondary groups based on the importance of the role they are assumed to play in the model design. Figure 16 shows the implemented **primary latent variables** and their classification.

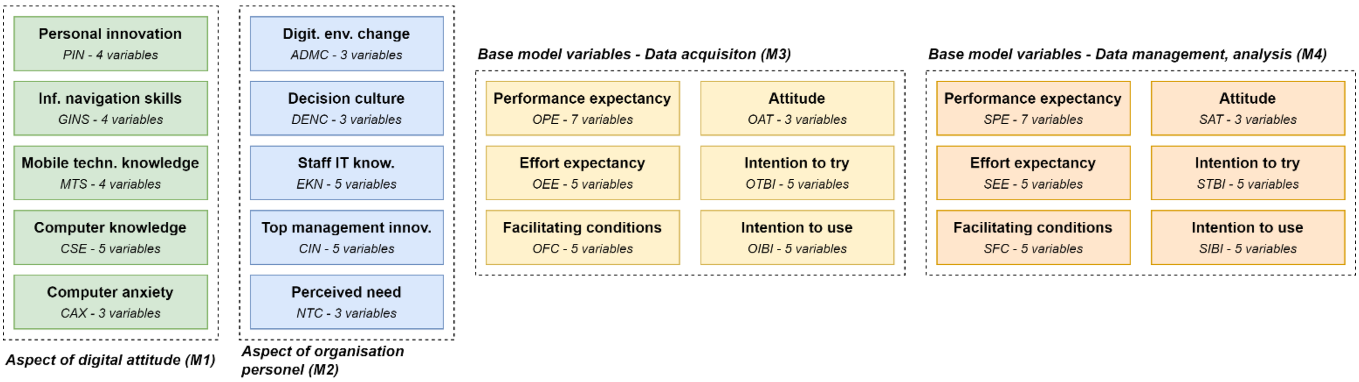

Figure 16: Summary of primary latent variables by aspect

Source: Own figure

To complement this, Figure 17 below shows the **additional latent variables** implemented to investigate the effects of some elements of the primary variables in the base model by iteratively extending it according to specified rules.

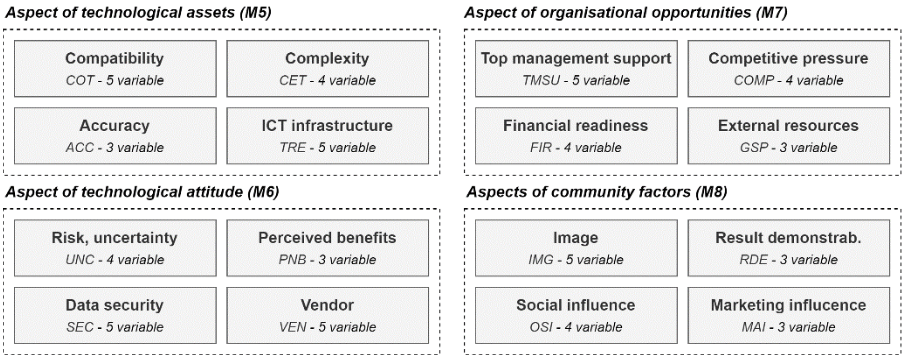

Figure 17: Summary of secondary latent variables by aspect

Source: Own figure

3.3.2. Demographic, IT and organisational characteristics of the research participants

The descriptive presentation of the sample provides a context specific information of the sample by presenting factors such as *demographic characteristics*, *IT-related experiences*, and the *characteristics of the business unit* associated with the participant. Without being exhaustive, the data on usage and interest should be highlighted, which show that high usage rate for the relevant items of the data collection tools can be determined (individual recording of assets and goods, use of fixed sensors), although there is also a high rate of isolation. However, the characteristics of data utilization based on the measured items of data management (decision support, business management or case management applications, customised data mining and analysis) has the lowest adoption rate, despite the high interest. This reflects the importance of examining the factors that express the intention to use, since the *two areas cannot create real value* in decision support for production and business processes *without the support of each other*.

3.3.3. Examining the divergence between the relevant fields based on the base model

In the presentation of the composition and structure of the latent variables, it was mentioned that the latent variables included in the basic model aspect (*M3* and *M4*) were measured twice after the subject of the measured variables had been rephrased, thus separately examining the opinion towards the subject of data collection (section 3 of the questionnaire) and data management (section 4 of the questionnaire) to examine the differences between the scoring of the intention to use in these areas. The results of the analysis of invariance to determine are presented in Table 1 below in an abbreviated form, considering the most stringent levels, where the first section expresses the results of the model and the second section their comparison.

Table 1: Results of the analysis of invariance at the most restrictive level of the characteristic

Model	*Khi-sq.*	*df*	*p*	*RMSEA*	*SRMR*	*AIC*	*Khi-sq. diff.*	*df diff.*	*p*
Strict inv. (OPE)	516,424	95	**<0,001**	0,274	0,113	1426,90	3,611	7	0,823
Strict inv. (OEE)	165,946	47	**<0,001**	0,207	0,068	1140,03	5.131	4	0,400
Strict inv. (OFC)	266,191	47	**<0,001**	0,281	0,121	1510,35	3,540	5	0,617
Strict inv. (OAT)	30,785	15	**0,009**	0,134	0,139	660,96	3,807	3	0,283
Strong inv. (OTBI)	84,796	42	**<0,001**	0,131	0,083	921,00	3,484	4	0,480
Strong inv. (OIBI)	168,519	42	**<0,001**	0,226	0,078	1081,56	1,026	4	0,906

Source: Own calculations

The results show that the invariance conditions are fulfilled in all cases, in most cases up to strict invariance. The condition of strict invariance was not met for the two dependent latent variables (intention to try and intention to use), so partial invariance can be determined, but the presence of strong invariance can be interpreted as an appropriate result (DIMITROV, 2010), since at this level the mean and weight of the indicators are already compared.

3.3.4. Impact of general variables on model variables

The next step was to examine the effects influencing the scoring of the latent variables in the model being developed. In this way, the question is that which combination of general variable influences most the evaluations expressing opinions on the factors included in the base model, thus directly or indirectly affecting attitude, intention to try and intention to use. To answer these questions, ordinal logistic regression was applied, supported by an application that allows the creation of model structured based on different variable combinations, the calculation of results and the management of metadata. A list of combinations based on *general variables* has been generated for the calculation of alternatives. Thereafter, the list of combinations is corrected by pre-determined constraints were used to examine their possible effect on the independent latent variables of the base model. The process of the optimisation is illustrated in Figure 18. Variables that were measured but not used are shown in grey.

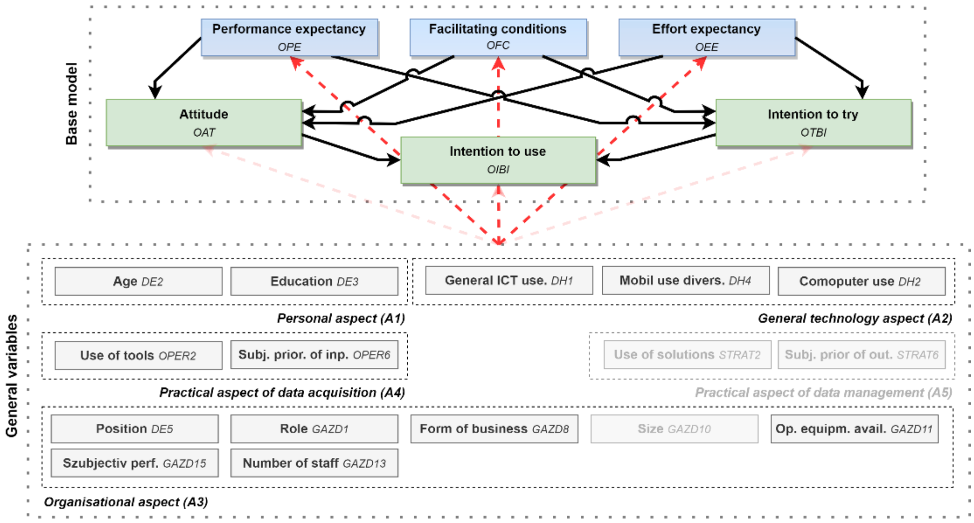

Figure 18: Examination of the effects on the latent variables of the base model (M3)

Source: Own figure and structure

The aspect of the **data acquisition focused base model (M3)** consists of a set of key latent variables in this case from the UTAUT2 model, including independent items expressing performance expectancy *(OPE)*, effort expectancy (OEE) and facilitating conditions *(OFC)*, as well as dependent items for attitude *(OAT)*, intention to try *(OTBI)* and intention to use *(OIBI)*, but only the independent items were used in these analyses. Based on the results, several variable groupings can be observed, presented in Figure 19 with the number of cases.

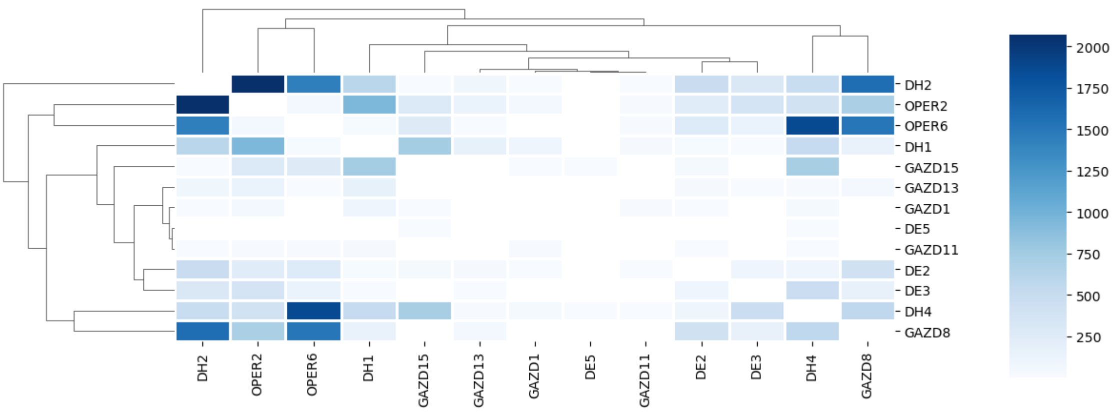

Figure 19: Frequency of variables affecting the scoring of the base model (M3)

Source: Own data and figure

The results showed that the personal aspect *(A1)* was relevant for performance expectancy *(OPE)* and effort expectancy *(OEE)* in terms of age and education. The general technological aspect *(A2)* was similarly relevant for computer use and was relevant for the facilitating conditions *(OFC)* in terms of the general diversity of device use. The organisational aspect *(A3)* was found to be relevant for all three variables considering subjective performance, role in business, form of business and asset availability. The practical aspect *(A4 and A5)* was relevant for all three variables based on subjective priority of data collection tools and input parameters.

3.3.5. Designing the base model to assess intention to use

In the following, the base model was developed and interpreted, which serves as a groundwork during the extension with latent variables expressing field and sector specific characteristics. The set of independent variables used and modified to develop the model are known from the UTAUT2 theoretical model, including performance expectancy *(OPE)*, effort expectancy *(OEE)* and facilitating conditions *(OFC)*. Before the survey, latent variables considered irrelevant were omitted from the original model. In the UTAUT model, the dependent variable is the behavioural intention *(BI)* as originally stated in the source, but this has been extended as discussed earlier. Accordingly, I have tried to implement latent variables representing increments, whereby attitude *(AT)*, intention to try *(TBI)* and intention to use *(BI)* were examined separately. The result of this extension is shown in Figure 20.

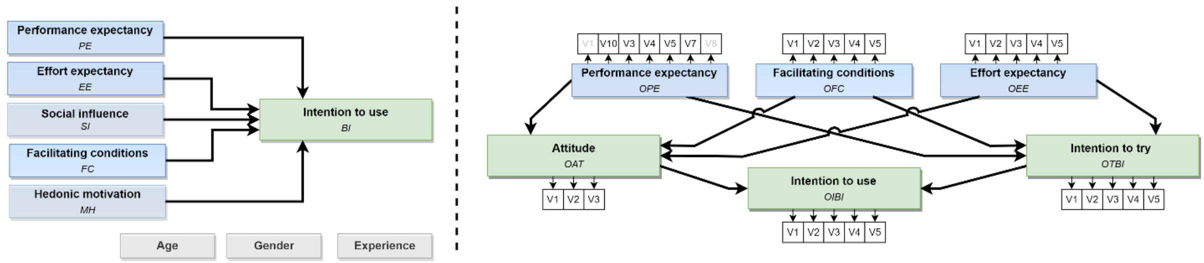

Figure 20: Initial changes to the UTAUT2 base model

Source: Own figure and structure

As a first step, the overlap between the latent variables used was examined, and then the indicators supporting the demonstration of different levels of validity specific to the methodology were calculated, including internal consistent reliability, convergent validity, discriminant validity, multicollinearity, and external model weights. The path coefficients of the developed base model are summarised in Table 2 below.

Table 2: Path coefficients of the base model (ADGY-A-2)

Path	*Path coeff.*	*Path. mean*	*Path. std.*	*t*	*p*
			Negative effect		
OEE → OTBI	-0,327	-0,329	0,113	2,888	**0,002**
OFC → OTBI	-0,017	-0,010	0,139	0,124	0,428
			Positive effect		
OPE → OTBI	0,721	0,710	0,127	5,675	**<0,001**
OPE → OAT	0,718	0,715	0,079	9,079	**<0,001**
OTBI → OIBI	0,516	0,511	0,128	4,022	**<0,001**
OAT → OIBI	0,322	0,327	0,110	2,940	**0,002**

OFC → OAT	0,176	0,179	0,077	2,287	0,017
OEE → OAT	0,003	0,003	0,070	0,047	0,461

Source: Own calculations

Overall, the value of the coefficient of determination for attitude (OAT) was $R^2 = 0.679$, while for intention to try the value was $R^2 = 0.305$. The value of the associated coefficient of determination for the intention to use (OIBI) was $R^2 = 0.483$. The structure of the model, which was developed without optimization based on the literature and the additional ideas, is shown in Figure 21. The width of the lines representing the paths is based on the standardised path coefficients.

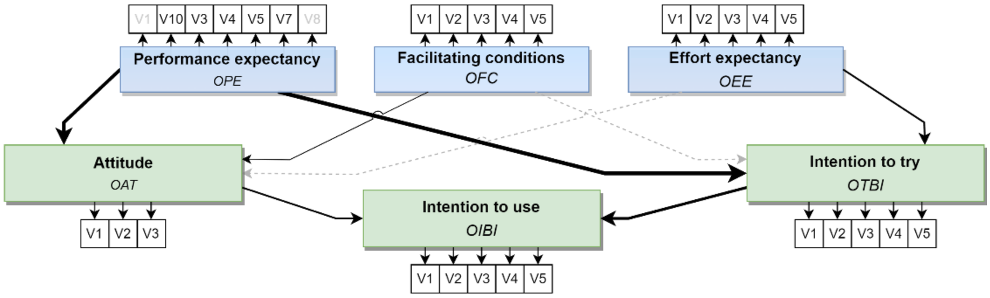

Figure 21: Structure of the base model (ADGY-A-2)

Source: Own figure and structure

The development of the base model has established the applicability of the UTAUT2 model for measuring technology acceptance, but despite its frequent use in the literature, the model can be considered rather general. The selection of variables presented earlier allows the model to be extended to adapt to the specificities of the field and the sector.

3.3.6. Extension plan of the base model

The base model developed, despite initial modifications, was found to be too general and not necessarily adapted to the specificities of the field and sector. Consequently, two steps of variable selection have been presented earlier, resulting in the availability of primary and secondary variables categorised by assumed importance in order to extend the model.

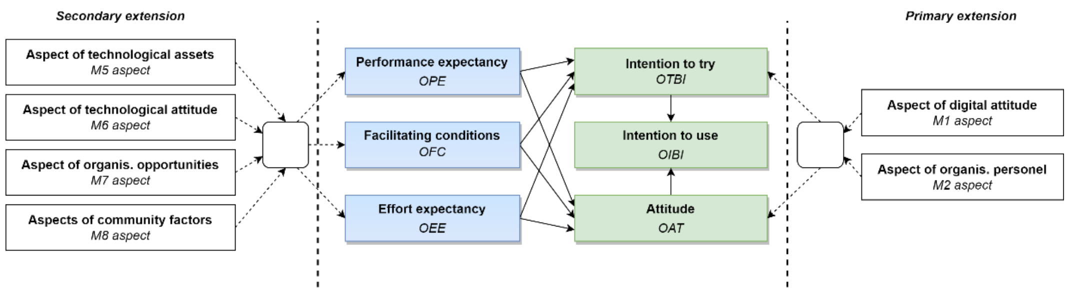

Figure 22: Primary and secondary model extension plan

Source: Own figure and structure

In case of the primary extension, the independent variables of the base model are extended using the latent variables included in the primary model aspects, while maintaining the original structure of the base model. In case of the secondary model extension, attention is turned to explaining the aggregated independent variables in the UTAUT2 model using the latent variables included in the secondary model aspects (Figure 22).

3.3.7. Primary extension of the basic model

To achieve this goal, the effects on the two previously known dependent variables, namely attitude *(OAT)* and intention to try *(OIBI)*, were assessed using the latent variables in the model aspects *M1*, *M2* and *M3*, from which the latter already used for the baseline model. Skipping the previously mentioned steps (checking the variables and prerequisites), the substantive part of the model calculation is summarised in Table 3 below, which provides an overview of the path coefficients and their significance through bootstrap sampling.

Table 3: Path coefficients for the base model (ADGY-D)

Path	*Path coeff.*	*Path. mean*	*Path. std.*	*t*	*p*
			Negative effect		
OEE → OTBI	-0,434	-0,426	0,115	3,778	**<0.001**
CAX → OTBI	-0,147	-0,14	0,094	1,561	**0,059**
OEE → OAT	-0,110	-0,107	0,082	1,339	**0,090**
ADMC → OTBI	-0,048	-0,039	0,099	0,485	0,314
			Positive effect		
OPE → OTBI	0,674	0,657	0,113	5,988	**<0.001**
OPE → OAT	0,668	0,657	0,098	6,848	**<0.001**
OTBI → OIBI	0,495	0,491	0,132	3,737	**<0.001**
OAT → OIBI	0,320	0,325	0,112	2,848	**0,002**
DENC → OAT	0,185	0,187	0,079	2,350	**0,009**
CSE → OAT	0,142	0,140	0,063	2,251	**0,012**
CSE → OTBI	0,135	0,128	0,093	1,453	**0,073**
OFC → OAT	0,132	0,133	0,074	1,800	**0,036**
NTC → OAT	0,092	0,096	0,055	1,672	**0,047**
DENC → OTBI	0,068	0,065	0,099	0,682	0,248

Source: Own calculation

The weight of the variables included in the base model has changed slightly due to differences in the variable composition, while significant change can be observed in the impact of the variables on the effort expectancy *(OEE)* and attitude *(OAT)*. In addition to the items already identified, the new variables included in the model are reviewed below. In terms of negative effects, only the effect of computer anxiety *(CAX)* on intention to try *(OTBI)* was found to be significant at path coefficient $\beta = -0.147$, while the effect of digital environment change (ADMC) was not significant despite its inclusion in the model. In terms of positive effects, the effect of the decision-making culture *(DENC)* on attitude *(OAT)* can be highlighted with a coefficient of $\beta = 0.185$, followed by the effect of the computer skills *(CSE)* variable on attitude *(OAT)* and intention to try *(OTBI)*. To the smallest extent, perceived need *(NTC)* had a positive effect on attitude *(OAT)*. For both aspects, 2 of the 5 latent variables included were included in the model. The resulting structure can ensure that different approaches can be taken into account when applying the model, thus providing a broader perspective for further assessments. The model structure is illustrated in Figure 23 below.

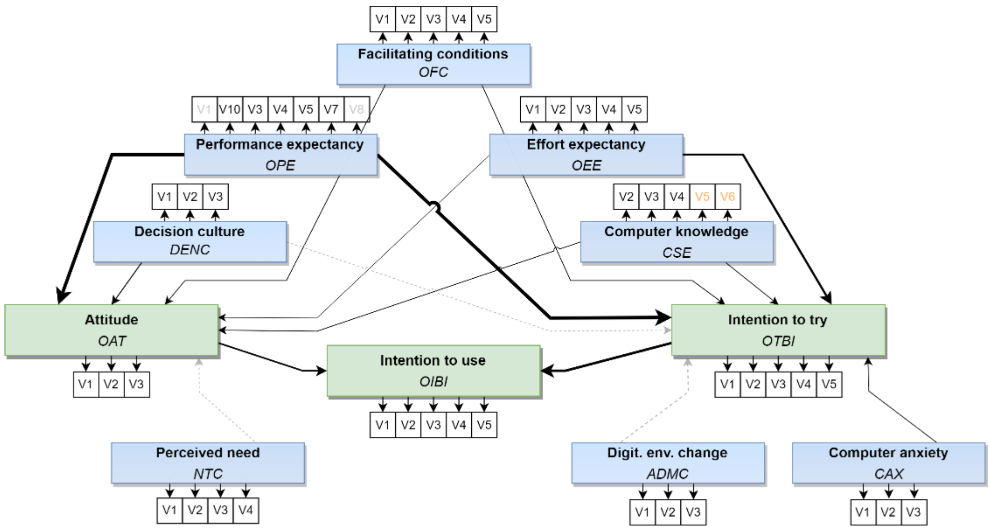

Figure 23: The selected base model after optimisation (ADGY-D)

Source: Own figure and structure

Despite the limitations, the performance of the base model has been improved. Although the latent variables implemented do not constitute as much weight as the aggregated variables included in the base model, their interpretation may lead to additional implications when formulating strategies to increase the intention of adaptation.

3.3.8. Secondary extension of the base model

The independent variables, including the performance expectancy *(OFC)*, effort expectancy *(OEE)* and facilitating conditions *(OFC)*, included in the UTAUT2 model used to develop the base model are aggregated, i.e., the characteristics of several latent variables have been included in their design (VENKATESH et al., 2003), so that as fewer variables express a phenomenon. Despite the frequent use in the literature, it is questionable to what extent the variables are able to represent the meaning in the actual context. During the first two steps of variable selection, a number of additional latent variables were measured, including the items included in the aspects *M5*, *M6*, *M7* and *M8*. The generation of the combinations and models, the calculations as well as the metadata management was performed using a dedicated application developed during the research. All combinations of the secondary variables were assigned to the dependent variables, resulting in the calculation of 65,518 model variants. The design pattern of the model generation is outlined in Figure 24 below.

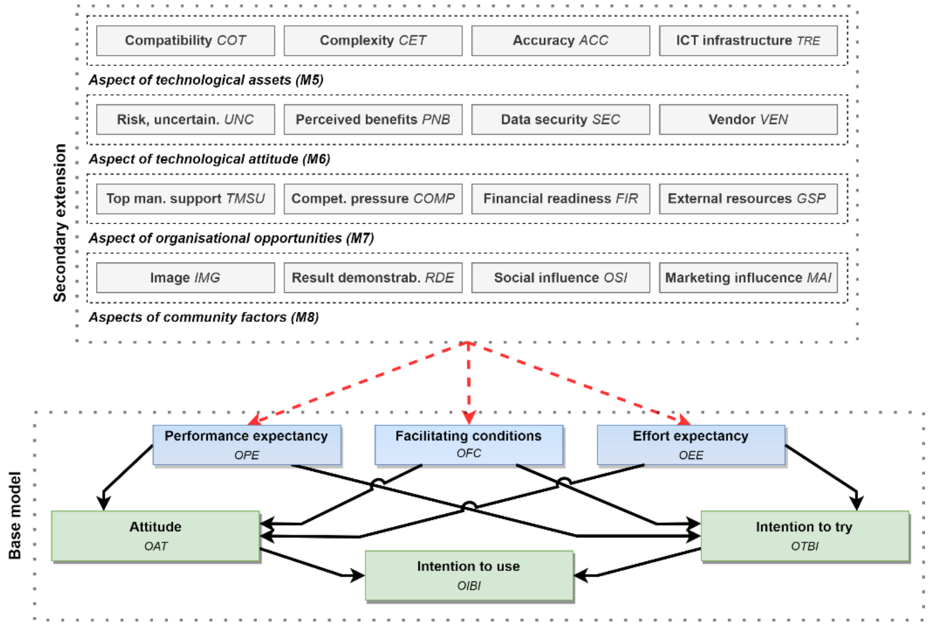

Figure 24: Strategy for secondary model extension

Source: Own figure and structure

The resulting metadata was used to query the set of technically plausible model variants, examining possible clustering of characteristics (mostly co-occurring variables), effect size and fit. In summary, Figure 25 below shows the count the different latent variables appeared in the model variants with statistically significant effects.

	M5				M6				M7				M8			
	COT	CET	ACC	TRE	UNC	PNB	SEC	VEN	TMSU	COMP	FIR	GSP	IMG	RDE	OSI	MAI
Perform. expectancy *OPE - 7 variable*	237	0	108		24	176	23	75		716			161	308	213	29
Effort ecpectancy *OEE - 5 variable*	118	304		269	12	281		42	69	9	27	1				
Faciliating conditio. *OFC - 5 variable*	19108	4833	133	1472	718	873	2291	2734	25738	132	846	2	10145	8671	4709	1444

Figure 25: Appearance and constraints of the variables used in the models

Source: Own figure and calculations

For the analysis of the effects on **performance expectancy (PE)**, the queries conducted after the iterative model generation showed that 941 model variants were considered appropriate when including latent variables for the *M5*, *M6*, *M7* and *M8* aspects, compared to an initial 65,518 variants. After reviewing the results, the variant 20D-K2053 was implemented taking into account the order of fit and the number of variables. Table 4 below provides an overview of the standardised path coefficients and the significance levels calculated using bootstrap sampling.

Table 4: The path coefficients for the chosen variant (20D-K2053)

Path	*Path coeff.*	*Path. mean*	*Path. std.*	*t*	*p*
			Negative effect		
MAI → OPE	-0,415	-0,299	0,250	1,660	**0,049**
COT → OPE	-0,291	-0,198	0,183	1,588	**0,056**
RDE → OPE	-0,262	-0,203	0,191	1,371	**0,085**
UNC → OPE	-0,171	-0,145	0,120	1,427	**0,077**
			Positive effect		
COMP → OPE	0,603	0,434	0,247	2,444	**0,007**
VEN → OPE	0,417	0,349	0,232	1,794	**0,036**
IMG → OPE	0,406	0,377	0,225	1,801	**0,036**
SEC → OPE	0,274	0,135	0,216	1,267	**0,100**
PNB → OPE	0,050	0,124	0,195	0,254	0,400

Source: Own calculations

The structure of the model developed to explain the performance expectancy (OPE), the typical paths and modifications are shown in Figure 26 below.

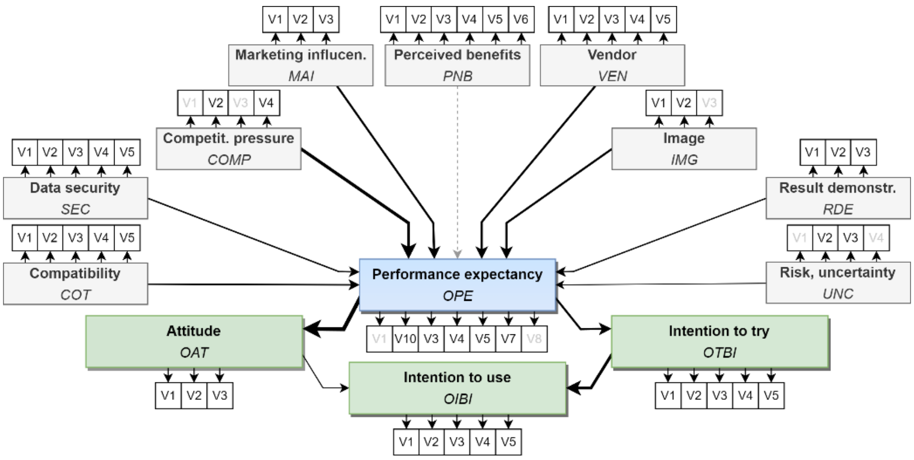

Figure 26: Variant chosen for performance expectancy (20D-K2053)

Source*: Own figure and structure*

For the analysis of the effects on the **effort expectancy (EE)**, the queries suggest that 621 model variants were appropriate compared to the initial 65,518 variants when the latent variables of the model aspects *M5*, *M6*, *M7* and *M8* are taken into account. After sorting the models representing the same aspects by fit and number of variables, model variant *20E-K3030* was selected, which explains the scoring of effort expectancy (OEE) with 7 independent variables. To highlight the main results, Table 5 shows the path coefficients representing the implemented paths and their significance based on bootstrap sampling.

Table 5: Path coefficients for the selected variant (20E-K3636)

Path	*Path coeff.*	*Path. mean*	*Path. std.*	*t*	*p*
			Negative effect		
CET → OEE	-0,343	-0,322	0,154	2,229	**0,013**
FIR → OEE	-0,238	-0,194	0,159	1,497	**0,067**
UNC → OEE	-0,196	-0,199	0,145	1,351	**0,088**
VEN → OEE	-0,001	0,025	0,186	0,004	0,498
			Positive effect		
TRE → OEE	0,375	0,344	0,181	2,067	**0,019**
SEC → OEE	0,280	0,285	0,151	1,853	**0,032**
COMP → OEE	0,143	0,082	0,186	0,769	0,221

Source: Own calculations

The model to support the explanation of the effort expectancy *(OEE)* is developed as shown in Figure 27.

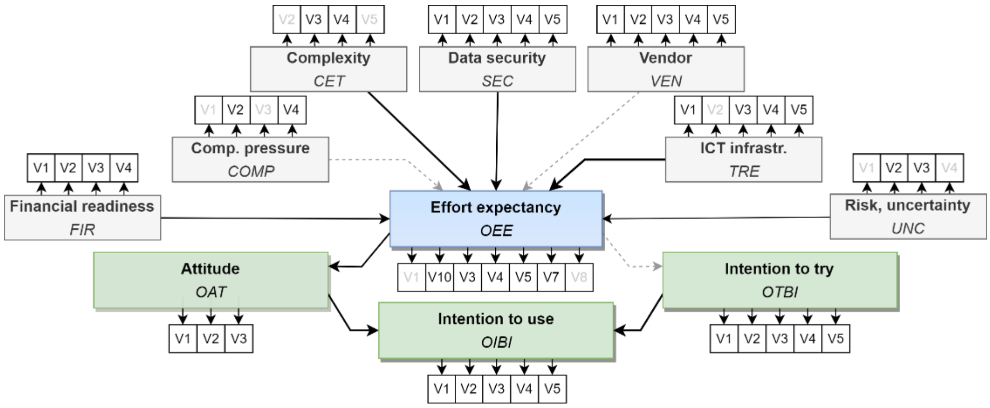

Figure 27: Variant chosen for effort expectancy (20E-K3636)

Source: Own figure and structure

For the analysis of the effects on the **facilitating conditions (OFC)**, 36,053 model variants were appropriate based on the queries, out of an initial 65,518 variants, considering latent variables of the *M5*, *M6*, *M7* and *M8*. After sorting the models representing the same groups of variables by fit and number of variables, the model variant *20F-K1133* was selected to explain the scoring by 8 independent variables. The characteristic path coefficients are presented below, expressing the main effects (Table 7).

Table 6: Path coefficients for the selected model (20F-K1133)

Path	*Path coeff.*	*Path. mean*	*Path. std.*	*t*	*p*
			Negatív hatás		
RDE → OFC	-0,321	-0,307	0,171	1,876	**0,030**
ACC → OFC	-0,229	-0,186	0,139	1,652	**0,049**
UNC → OFC	-0,154	-0,136	0,123	1,245	**0,100**
FIR → OFC	-0,134	-0,138	0,128	1,046	0,148

	Pozitív hatás				
TMSU → OFC	0,693	0,682	0,143	4,842	**<0.001**
IMG → OFC	0,418	0,381	0,182	2,296	**0,011**
SEC → OFC	0,291	0,281	0,167	1,738	**0,041**
COT → OFC	0,211	0,215	0,134	1,568	**0,058**

Source: Own calculation

The structure of the model to aid the explanation of the facilitating conditions *(OFC)* can be described as shown in Figure 28 below.

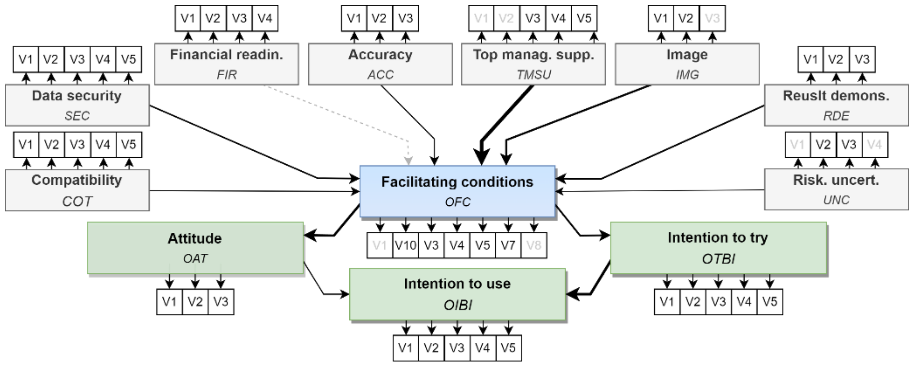

Figure 28: Variant chosen for the facilitating conditions (20F-K1133)

Source: Own figure and structure

The results obtained in the design of the three model variants are varying. In the following, as the summary of the results, based on the tested hypotheses and the developed aspects we are able to explain the scoring of the aggregate variables in question, but the allocation of these variables holds important information as well. The performance expectancy *(OPE)* includes items with a technological focus (aspect of technological assets - *M5* and aspect of technological attitude – *M6*), while the facilitating conditions *(OFC)* include some items concerning the aspect of organisational capabilities *(M7)* and community factors *(M8)* in a balanced way. The effort expectancy includes, in addition to the technology-focused aspects, items related to the aspect of organisational capabilities *(M7)*. Overall, it has been possible to define the items that drive the scoring of aggregate variables that define the model, thus providing a clearer picture of the issues that are considered to be more crucial in case of adoption.

3.3.9. Investigating the feasibility of estimation using a neural network

In the assessment of general effects, rudimentary attempts were made to estimate the dependent model variables, but in the absence of implementation of independent latent variables, the results were variable, usually with poor accuracy. Accuracy could be improved in some cases by gradient boosting based optimization, but not in all cases and then typically only marginally. However, the subsequently developed PLS-SEM models provided a new perspective on estimation. Since the method is less typically used for estimation with new data, an alternative solution was implemented. To perform the estimations, a number of different neural networks were used, according to the models developed. In order to increase the efficiency, hyperparameter optimization (grid-search) was implemented in addition to the change of the structure. With respect to dependent variables, the latent variables considered relevant were those included in the M3 model aspect. The results of the calculations and their comparison with previous results are summarised in Table 7 below.

Table 7: Accuracy of estimates based on different methods

Estimated variable	*Train RMSE*	***Test RMSE***	*Train acc.*	***Test acc.***
OAT (Boost)	0,800	0,953	0,559	0,468
OAT (NN)	0,389	**0,405**	0,697	**0,682-0,989**
OTBI (Boost)	0,986	1,111	0,495	0,390
OTBI (NN)	1,024	**1,046**	0,506	**0,544-0,986**
OIBI (Boost)	0,882	1,062	0,535	0,425
OIBI (NN)	0,841	**0,962**	0,590	**0,566-0,989**

Source: Own calculation

4. NEW OR NOVEL RESULTS OF THE THESIS

To summarise the findings relevant to the thesis, *research questions* were formulated, followed by the review of related *theses* and main hypotheses.

4.1.1. Results of the first phase of the research

Research question 1: The stages of the process involved a quantitative review of the literature and community entries. The preliminary research was a key step in identifying topics of interest to the sector, which formed the basis of the research framework for measurement and evaluation. Five theses were formulated within this question.

Research question 1: In what form do relevant **tools** and **solutions** appear within the field of agricultural production, regarding to their **relationships**, **associations** and **opinions**, based on literature and public entries?			
Thesis 1: The network of keywords in the literature can be used to identify the main **tools**, **solutions** and **themes**	**Thesis 2:** Based on keywords in the literature, clear **groups** (domains) can be established.	**Thesis 3:** The keywords in the literature can be used to identify the **emerging fields**. The key factors that characterise them are identified using the keywors.	**Thesis 4:** Definable **aspects** can be determined in public entries expressing positive and negative opinions.
No hypothesis was formulated for this question.			

Figure 29: The first research question and theses

Source: Own data

Overall, the first stage of the research was to identify and summarise the relevant **tools** and **methods**, as well as the **doubts** that were prominent in the variable selection process, by analysing the literature and public entries.

4.1.2. Results of the second phase of the research

An important task of the second half of the research was the implementation of the three-step variable selection and the transformation of the selected general and model variables into aspects for the summary conclusions, followed by the implementation of the survey.

Research question 2: Since a large number of variables were used, the question concerned the assumed redundancy, which provided an answer to simplify the model to be developed. Within the question, a thesis and a hypothesis were formulated.

Research question 2: Do participants have **different opinions** on the fields related to relevant tools and solutions?	
Thesis 5: There is a difference in opinion on the possibilities of **data acquisition** and **management** (based on model variables), i.e. the target group sees the fields as a single subject that otherwise can be differentiated functionally in a professional sense.	
Hypothesis 1: In the assessment of each of the factors in the relevant fields, the conditions for differences were met for all the model variables considered.	Accepted

Figure 30: Second research question, related theses and hypothesis

Source: Own data

The formulated hypothesis was accepted according to the results of the invariance analysis, whereby there was no detectable difference between the scoring of data acquisition and data management aspects, thus allowing for simplification in the model design.

Research question 3: The latent variables were reduced in the previous step by excluding the M4 model aspect, thus the assumption of possible externalities was formulated regarding the M3 model aspect, as a continuation. For the present question, a single thesis and three hypotheses were formulated based on the aspects of the general variables, using the method presented in the thesis.

Research question 3: Is it possible to determine the impact of the **general variables** on the assessment of the latent factors that are to be included in the model?	
Thesis 6: Each combination of **external factors** (general variables) influences the evolution of the scores of the different latent variables (model variables), thus influencing the outcome of the models constructed on the basis of these variables by segmenting the sample.	
Hypothesis 2: One item of the personal aspect (A1) was relevant in the analysis of externalities.	Accepted
Hypothesis 3: One item of the technological aspect (A2, A4 és A5) was relevant in the analysis of externalities.	Accepted
Hypothesis 4: One item of the organisational aspect (A3) was relevant in the analysis of externalities.	Accepted

Figure 31: The third research question, related theses and hypotheses

Source: Own data

The related hypotheses were accepted. The research question is marked in yellow through partial response as the results were not used as moderator variables or clustering factors in the PLS-SEM model due to limitations.

Research question 4: The base model developed has been extended in two directions, primary and secondary, to adapt it to the specificities of the field and the sector. The hypotheses can be classified into three groups. The hypotheses were formulated in an aggregated way, taking into account the primary and then secondary model aspects, taking into account the developed aspects. In total, three theses were formulated. One hypothesis was formulated for the base model design and three hypotheses were formulated for the primary model extension and the secondary model extension.

Research question 4: Is it possible to develop **theoretical models** that can explain the factors affecting the acceptance and use of **relevant tools and solutions** based on the selected latent variables?			
Thesis 7: The base model developed by extending the **dependent factors** of the UTAUT2 theoretical framework explains attitudes, willingness to try and, through these, willingness to use.	**Thesis 8:** It is possible to adapt to the area and sector and improve the performance of the model by extending the **independent factors** of the base model.	**Thesis 9:** The **independent factors** in the UTAUT2 theoretical framework can be adequately explained by the selected secondary combinations of secondary model variables.	
Hypothesis 5: In designing the base model, a difference was observed between the effect on theimplemented levels, i.e. attitude and intention to try.			Accepted
Hypothesis 6: The aspect of digital attitude (M1) was relevant for the primary model extension.			Accepted
Hypothesis 7: The aspect of organisational personel (M2) was relevant for the priamry model extension.			Accepted
Hypothesis 8: The aspects of technology (M5 és M6) were relevant for the secondary model extension.			Accepted
Hypothesis 9: The aspect of organisational opportunities (M7) was relevant for the secondary model extension.			Partial. accepted
Hypothesis 10: The aspect of community factors (M8) was relevant for the secondary model extension.			Partial. accepted

Figure 32: The fourth research question, related theses and hypotheses

Source: Own data

The majority of the listed hypotheses were accepted, however, as not all elements of the model aspects of organisational capabilities *(M7)* and community factors *(M8)* were applied during the secondary model extension, resulting in a partially accepted hypothesis.

Research question 5: During the development of PLS-SEM models, the use of its structure as a neural network for estimation was considered. A thesis was formulated to assess the possibility.

Research question 5: Can factors affecting the acceptance and use of tools and solutions be **estimated**?
Thesis 10: It is possible to **estimate** the main **dependent factors** (model variables) using the general variables and some model variables based on the model structure developed.
No hypothesis was formulated for this question.

Figure 33: The fifth research question and associated thesis

Source: Own data

Although the constructed network yielded good results in several cases based on cross-validation, a larger volume of data would be needed for generalization and to train the network structure properly, so although the model was constructed and the parameters optimized, I consider it as a partial result due to the limitations.

4.1.3. Summary of results

The results cover two main themes, involving the *development of the supporting applications* and the steps taken to achieve the *research objectives*, which are summarised in Figure 34. Grey highlight indicates the supporting applications and blue highlight the objectives, and the associated research question is indicated.

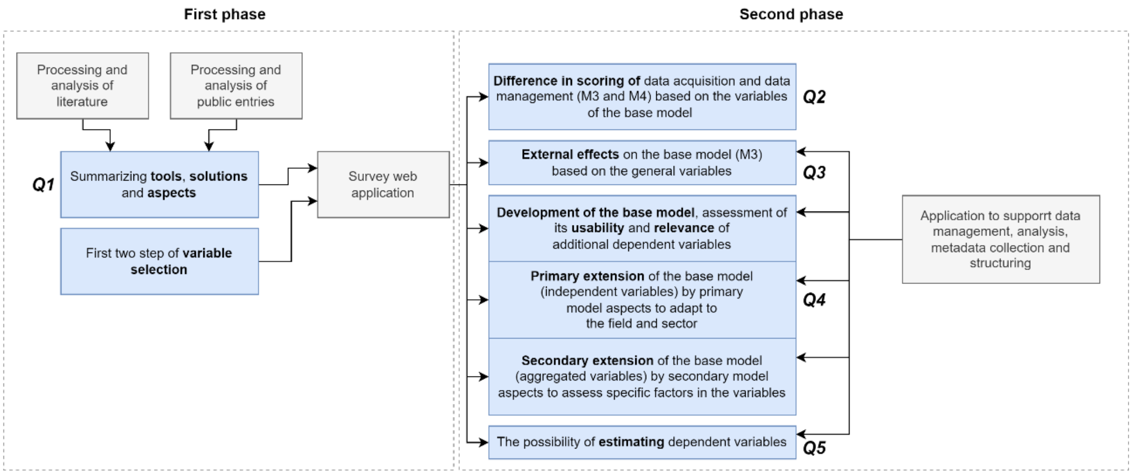

Figure 34: Summary of key findings

Source: Own data

On the one hand, in order to achieve the objectives, it was necessary to **develop a number of supporting applications**, which include, without being exhaustive:

- The **creation of a workflow and application** to support the management and analysis of **literature data** and **public entries**,

- the **creation of an application for the administration of the questionnaire** to display the main features (dynamic questions, additional information),
- the **preparation of a workflow and application** that performs the **three-step variable selection**, **performs out the calculations** as planned and structures and compares the results,
- finally, the **design and implementation of the system concept** and **open-source system** to assess the impact of practical application on processes and users.

Variable selection and the development of model structures adapted to the specificities of the sector and fields played a key role in the results. In the absence of a theoretical framework, it was necessary to develop a workflow that would allow the confirmation of the research points. Consequently, the **research consists of two main phases**.

Phase 1: Summary of key factors, survey design

- Based on the literature and community contributions, the main **tools and methods have been identified**, as well as the most **important aspects**, highlighting the doubts.
- A three-stage variable selection was started by **collecting, sorting and grouping latent frequently used latent variables and PLS-SEM model variants**, based on the literature to support the survey.
- Finally, the **survey was compiled** on the basis of the results, followed by a survey of small and medium-sized enterprises engaged in arable crop production.

Stage 2: Summary of key factors, survey design

- As a first step, it was found that based on the sample, **there is no difference between of the scoring of the base model variables** (based on invariance analysis), describing data acquisition and data management, thus providing an opportunity to simplify the model.
- The **main externalities were then assessed** (based on ordinal logistic regression, supported by iterative process) on the basis of the general variables, to support the subsequent modelling process by knowing the biasing effects.
- The modelling process resulted in the **development of UTAUT2 technology acceptance model variant** (based on PLS-SEM modelling), which, after the

extension of the dependent factors, was now able to test for different levels of acceptance. The model proved to be suitable for measuring the basic effects but was characterised by over-generalisability.

- Therefore, during the **primary extension, it was possible to improve the performance of the model** (based on PLS-SEM modelling, supported by iterative process) by implementing additional variables corresponding to the primary model aspects during the optimization.
- In the **secondary extension, the interpretability of the independent variables** in the UTAUT2 model with aggregated information content **was investigated** by considering combinations of the effects on them (based on PLS-SEM modelling, supported by iterative process), according to the factors included in the secondary model aspects.
- Finally, on the basis of the model structure developed, the **estimation of dependent variables** (based on neural networks, supported by iterative process), including the various levels of intention to use, were reviewed.

5. PRACTICAL APPLICABILITY OF THE RESULTS

In the case of the research, the possibility of practical application is mainly summarised in some alternatives for continuation. It is possible to aid the factors that are perceived as specific benefits and to mitigate the factors that are perceived as barriers in a targeted way when developing certain system concepts. In the case of the actual application of practical systems designed with this information in mind, it is possible, on the one hand, to assess the **impact on attitudes** (before and after) by reapplying the model that has been developed, and, on the other hand, to examine the **impact of the relevant systems on the processes**.

5.1. Factors affecting implementation

The factors can be divided into three groups, according to personal interpretation, based on their influenceability, the first of which includes factors that can be **influenced directly**, the second can be **influenced indirectly**, while the third **cannot be influenced** (or influenced only with substantial effort). In the context of the present research, the indirectly influenceable aspects are considered to be the most important, which include the latent variables of perceived need *(NTC)*, compatibility *(COT)*, complexity *(CET)*, accuracy *(ACC)*, ICT infrastructure *(TRE)*, financial readiness *(FIR)* and data security *(SEC)*. The analysis of the observed effects for the selected models also overlaps with the present items, which are explained in Table 8 below up to a maximum of two levels.

Table 8: Indirect impacts for directly influenced factors

Path	*Path coeff.*	*Path. mean*	*Path. std.*	*t*	*p*
ACC → OFC → OAT	-0,108	-0,092	0,074	1,454	**0,073**
ACC → OFC → OTBI	-0,080	-0,069	0,059	1,351	**0,088**
CET → OEE→ OAT	-0,097	-0,089	0,063	1,553	**0,060**
COT → OPE→ OAT	-0,224	-0,154	0,142	1,573	**0,058**
COT → OPE→ OTBI	-0,201	-0,138	0,127	1,585	**0,057**
FIR → OEE → OAT	-0,067	-0,056	0,054	1,227	**0,098**
TRE → OEE → OAT	0,107	0,104	0,081	1,312	**0,095**

Source: Own calculations

In order to influence each factor, I have tried to formulate feasible alternatives within the focus of this research, which are presented in Figure 35, with the corresponding categories and latent variables indicated.

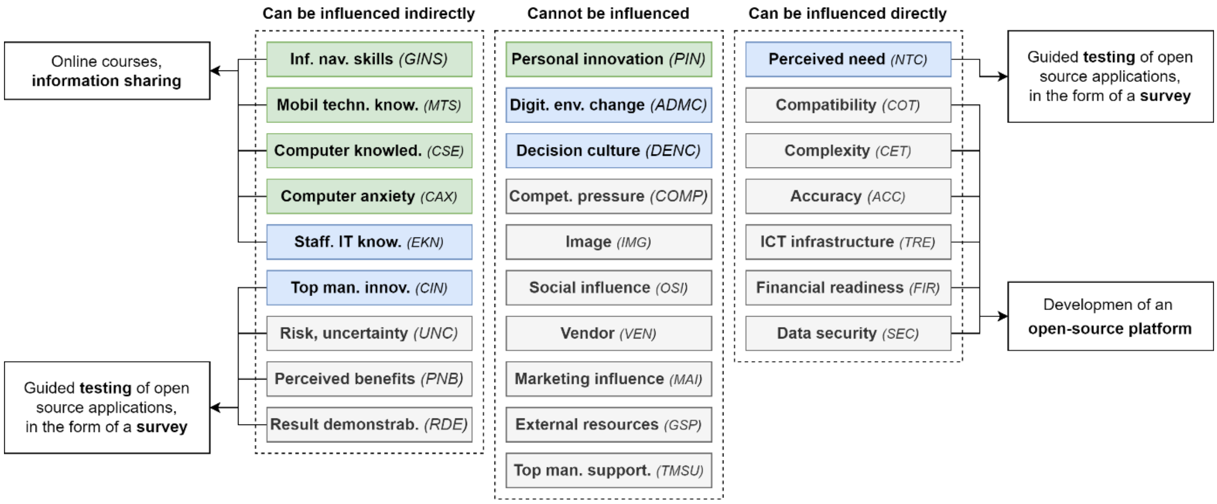

Figure 35: Possibilities to influence the effects represented by each latent variable

Source: Own figure

A further relevant issue of adaptation and, at the same time, of this test environment, involves the preferred way of purchasing and operating and the way of acquiring the knowledge needed to use and adapt (Figure 36). The participants' responses suggest that the application is primarily relevant in the form of individual services.

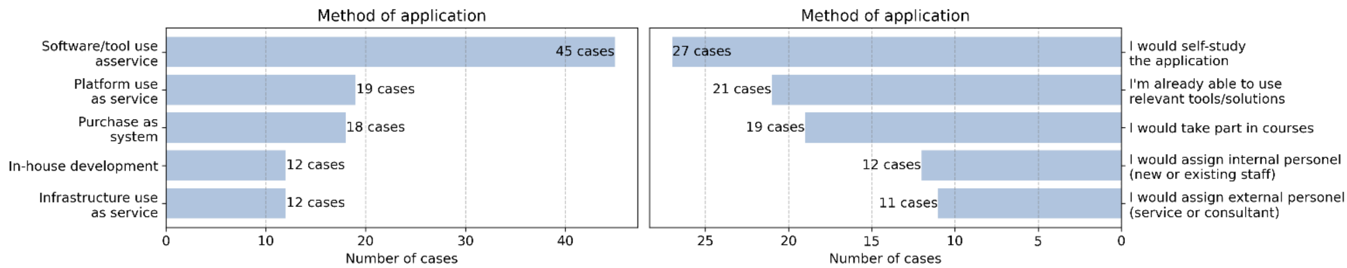

Figure 36: Method of implementation and application

Source: Own figure and data

The use of tools and methods might cause the issue of compatibility, which can be overcome by the development of standards. Among the options, I have set the objective of producing a system design to reduce the knowledge gap and identify information needs through feasibility studies and practical surveys that can be carried out in the future. Several experiments (environmental data collection, hardware and software systems, and data management frameworks) were also conducted during the research, the results of which are part of the test platform supporting the demonstration of options.

5.2. Develop a system model to support the assessment of practical impact

The primary objective of the testbed is to use the information gathered from the survey to test the practical possibilities for data collection, data management and data analysis in terms of professional decision support among the target group of participants, thus testing the changes in opinion and acceptance resulting from practical application and the actual information needs of decision-makers. The data model, developed according to the needs identified through the experience and the survey, is based on a proprietary platform used in practice, but the aim was to generalise it as much as possible to develop open source (transparent and cost-effective) alternatives. The components of the platform include an information system (web and desktop application), external and internal databases (Big Data concept), an API (application programming interface) for integration and sensor networks (Internet of Things concept).

As the target group preferred to use it as a service during the needs assessment, decentralisation was an important consideration, with the creation of server-side application programming interfaces as opposed to servers maintained in the organisation. Two databases are required to store the data. For raw data, the relevance of the Big Data concept arises. The associated Hadoop ecosystem is characterized by horizontal scalability, which provides a cost-effective solution. In addition to storing the data, I felt it necessary to create a metadatabase capable of storing descriptive data about the components (devices and data) (Figure 37).

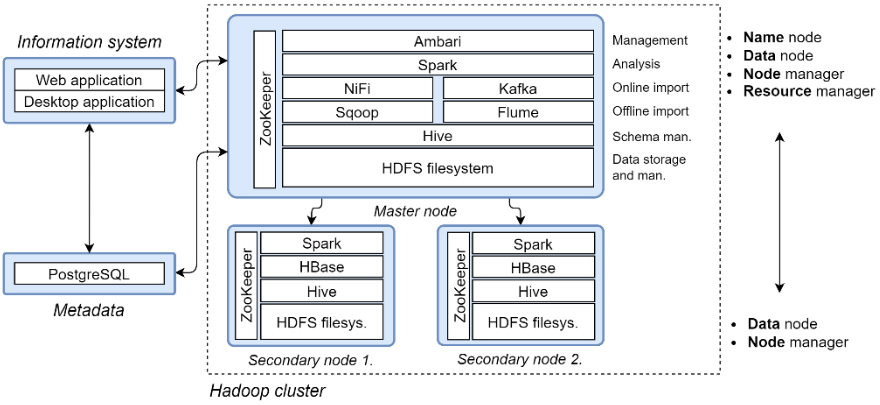

Figure 37: Schematic overview of the system to demonstrate practical options

Source: Own design

Several data sources have been integrated to build the dataset used to perform the tests, mimicking real-world usage conditions. In the case of environmental data, the sensor network of the previously mentioned system was used. For economic data, several structured and semi-structured data sources available online were used. In order to collect and integrate semi-structured, freely available economic data, a number of websites (CSO, FADN, FAO, Eurostat, etc.) have been mirrored, providing data series such as crop area, producer and market prices, commodity options, production costs, input use, market prices of related products, import, export data and weather data, with different resolutions. To integrate the data, open-source alternatives to the relevant procedures have been prepared, supporting research and their application in decision support (Figure 38).

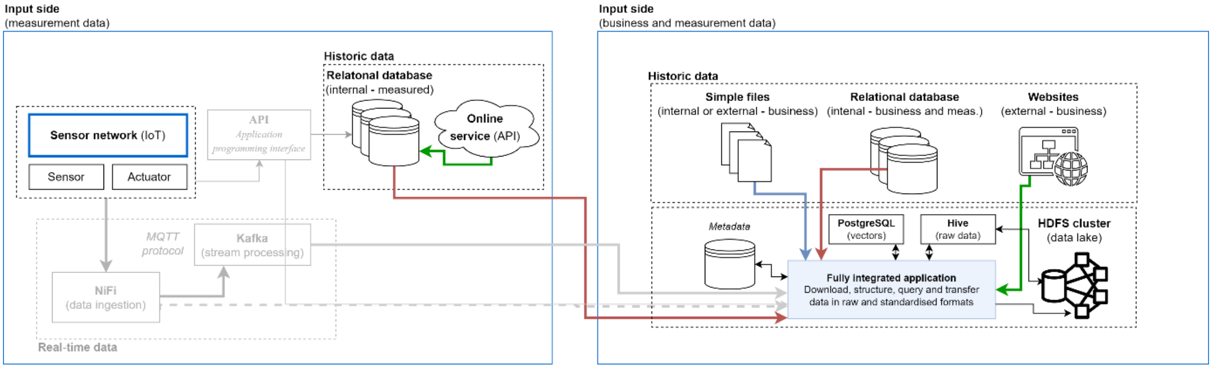

Figure 38: Importing economic and sensor data from different sources

Source: Own design

Regardless of the shape of the data and the overlaps, the algorithm reviews the variants before querying and then concatenates them into an output vector. After integration, both literature and decision makers' views can be considered to develop decision support models, thus adapting to the requirements. Several models have been taken from the previously developed system for presentation purposes. In the present version, a module supporting temporal and spatiotemporal estimation based on the managerial aspect of the concepts outlined has been developed for testing with the economic data collected and is able to estimate purchase prices and quantities based on the ConvLSTM architecture chosen for optimisation, thus presenting some options in a limited way.

In terms of the potential for practical application, the post-research objectives include the formulation and integration of as many options as possible into the information system to provide a basis for possible continuation. For the factors listed, we have seen attempts to overcome some of the barriers, including reducing complexity through automated operations, increasing compatibility through standardised processes, reducing costs through open-source components, increasing accuracy through selective models, adapting to scale through scalability, and meeting needs through needs assessment by decision makers.

6. PUBLICATIONS RELATED TO THE THESIS

Scientific literature in foreign languages

1. **TÓTH M.**, DÉR D., BOTOS SZ., SZILÁGYI R. (2021): Computer vision in agriculture, application development using open source tools and systems. Journal of Agricultural Informatics. Vol. 10, No. 2. pp. 37-47. DOI: 10.17700/jai.2019.10.2.545
2. **TÓTH M.**, FELFÖLDI J., SZILÁGYI R. (2019): Possibilities of IoT management system in greenhouses. Georgikon for Agriculture Vol. 23, No. 3. pp. 43-62.
3. **TÓTH M.**, SZILÁGYI R. (2017)**:** Development and testing experiences of a management supporting data acquisition system. Journal of Agricultural Informatics. Vol. 8, No. 2. pp. 55-70. DOI: 10.17700/jai.2017.8.2.382

Scientific literature in Hungarian with abstract in foreign language

4. **TÓTH M.**, FELFÖLDI J., SZILÁGYI R. (2018): IoT eszközök alkalmazása a döntéshozatal támogatására. International Journal of Engineering and Management Sciences (IJEMS) Vol. 3. (2018). No. 4. pp. 125-141. DOI: 10.21791/IJEMS.2018.4.12

Scientific book/chapter in a foreign language

5. **TÓTH M.**, FELFÖLDI J., VÁRALLYAI L., SZILÁGYI R. (2022): Application possibilities of IoT based management systems in agriculture. Springer

REFERENCES

Agovino, M. – Casaccia, M. – Ciommi, M. – Ferrara, M. – Marchesano, K. (2019): Agriculture, climate change and sustainability: The case of EU-28. Ecological Indicators, v.105(April), pp. 525–543. doi:10.1016/j.ecolind.2018.04.064

Aria, M. – Cuccurullo, C. (2017): bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, v.11(4), pp. 959–975. doi:10.1016/j.joi.2017.08.007

Batagelj, V. – Cerinšek, M. (2013): On bibliographic networks. Scientometrics, v.96(3), pp. 845–864. doi:10.1007/s11192-012-0940-1

Benotsmane, R. – Kovács, G. – Dudás, L. (2019): Economic, social impacts and operation of smart factories in Industry 4.0 focusing on simulation and artificial intelligence of collaborating robots. Social Sciences, v.8(5). doi:10.3390/socsci8050143

Cobo, M. J. – López-Herrera, A. G. – Herrera-Viedma, E. – Herrera, F. (2011): An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. Journal of Informetrics, v.5(1), pp. 146–166. doi:10.1016/j.joi.2010.10.002

Conroy, R. (2015): Sample size A rough guide (pp. 1–30)pp. 1–30. Elérés forrás https://pdfs.semanticscholar.org/4781/878153e13322c028c7d8970e7f52fbaa102a.pdf

De Marneffe, M. C. – Dozat, T. – Silveira, N. – Haverinen, K. – Ginter, F. – Nivre, J. – Manning, C. D. (2014): Universal stanford dependencies: A cross-linguistic typology. Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014, pp. 4585–4592.

Devlin, J. – Chang, M. W. – Lee, K. – Toutanova, K. (2019): BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, v.1(Mlm), pp. 4171–4186.

Dimitrov, D. M. (2010): Testing for factorial invariance in the context of construct Validation. Measurement and Evaluation in Counseling and Development, v.43(2), pp. 121–149. doi:10.1177/0748175610373459

Faul, F. – Erdfelder, E. – Lang, A.-G. – Buchner, A. (2007): G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, v.39(2), pp. 175–191. doi:10.3758/BF03193146

Holden, R. J. – Karsh, B.-T. T. (2010): The Technology Acceptance Model: Its past and its future in health care. Journal of Biomedical Informatics, v.43(1), pp. 159–172. doi:10.1016/j.jbi.2009.07.002

Jeffery, M. (2010): Data-driven marketing: The 15 metrics everyone in marketing should know. John Wiley.

Kamada, T. – Kawai, S. (1989): An algorithm for drawing general undirected graphs. Information Processing Letters, v.31(1), pp. 7–15. doi:10.1016/0020-0190(89)90102-6

KSH (2020a): Agrárdigitalizáció. . Elérés 2022. június 20., forrás Agrárcenzus 2020 website: https://www.ksh.hu/docs/hun/xftp/ac2020/agrardigitalizacio/index.html

KSH (2020b): Gazdaságtipológia. . Elérés 2022. június 20., forrás Agrárcenzus 2020 website: https://www.ksh.hu/docs/hun/xftp/ac2020/gazdasagtipologia/index.html

KSH (2020c): Mezőgazdasági munkaerő, generációváltás. . Elérés 2022. június 20., forrás Agrárcenzus 2020 website: ksh.hu/docs/hun/xftp/ac2020/mezogazdasagi_munkaero_generaciovaltas/index.html

Rejikumar, G. – Aswathy Asokan, A. – Sreedharan, V. R. (2020): Impact of data-driven decision-making in Lean Six Sigma: an empirical analysis. Total Quality Management and Business Excellence, v.31(3–4), pp. 279–296. doi:10.1080/14783363.2018.1426452

Sarker, M. N. I. – Wu, M. – Alam, G. M. M. – Islam, M. S. (2019): Role of climate smart agriculture in promoting sustainable agriculture: A systematic literature review. International Journal of Agricultural Resources, Governance and Ecology, v.15(4), pp. 323–337. doi:10.1504/IJARGE.2019.104199

Tóth, M. – Dér, D. – Botos, S. B. – Szilágyi, R. (2019): Computer vision in agriculture, application development using open source tools and systems. Journal of Agricultural Informatics, v.10(2), pp. 37–47. doi:10.17700/jai.2019.10.2.545

Tóth, M. – Felföldi, J. – Szilágyi, R. (2019): Possibilities of IoT based management system in greenhouses. Georgikon for Agriculture, v.23(3), pp. 43–62.

Tóth, M. – Szilágyi, R. (2017): Development and testing experiences of a management supporting data acquisition system. Journal of Agricultural Informatics, v.8(2), pp. 55–70. doi:10.17700/jai.2017.8.2.382

Venkatesh – Morris – Davis – Davis (2003): User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, v.27(3), pp. 425. doi:10.2307/30036540

Zambon, I. – Cecchini, M. – Egidi, G. – Saporito, M. G. – Colantoni, A. (2019): Revolution 4.0: Industry vs. Agriculture in a Future Development for SMEs. Processes, v.7(1), pp. 36. doi:10.3390/pr7010036

ABBREVIATIONS

MCA	Multiple cossrespondence analysis
PLS	Partial least square
SEM	Structural equation modeling
STÉ	Standard production value – Standard termelési érték
TAM	Technology Acceptance Model
TOE	Technology, Organization, Environment
UTAUT	Unified Theory of Acceptance and Use of Technology