Smart Agriculture and Digital Transformation on Case of Intelligent System for Wine Quality Prediction

D. Oreški*, I. Pihir * and K. Cajzek*
* University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia
{dijana.oreski, igor.pihir, ksenija.cajzek}@foi.unizg.hr

***Abstract* - The use of emerging technologies such as Industry 4.0 in the digital transformation of businesses is growing exponentially in various domains including agriculture. Implementation of digital transformation leads to innovation and automatization in business processes. IoT technologies combined with machine learning approaches achieve significant benefits in agriculture. This paper focuses on wine quality prediction and suggests an intelligent system based on the machine learning methodology of the decision tree. The effectiveness of classification and regression trees has improved a great deal to help in prediction. By using a public data set, the decision tree is developed consisting of red and white wines characteristics with their quality assessed by experts. Based on the developed predictive model intelligent system is constructed to fully automate the process of wine quality detection.**

Keywords* - *smart agriculture; digital transformation; industry 4.0; machine learning; decision tree

I. INTRODUCTION

Intelligent systems and intelligent agents constitute an important research field in various domains. Their decision-making capabilities, are a crucial benefit in achieving the automatization of a business process. Agriculture is no exception [1], [2], [3]. Under the EU funded project Center of Competences for Digital Transformation of the Food Industry in Rural Areas concepts of artificial intelligence (AI), machine learning (ML) and big Internet of Things (IoT) data are being tested for the purpose of digital transformation (DT) in smart agriculture (SA). These concepts are also used by other authors like Ghavate and Joshi in smart farming [4], by Mangiaracina et al.., in wine industry [5], and by Mazzetto et al.. in "precision agriculture" [6]. In this research, a case study of wine quality prediction was presented. Based on the publicly available data set [7], the machine learning approach of a decision tree is applied. Such a predictive model will serve as a basis for intelligent system development what is the final goal in the experimental phase of the project as a minimum viable product (MVP). This approach leads to automatization of wine quality assessment and it is determined to be an experimental case study in favor of digital transformation in wine testing processes.

This paper is structured as follows. After the introduction, a literature review of digital transformation in smart agriculture and wine testing literature is presented. Following, chapter 3 describes the data set used in the research as well as methods used for analyzing data and developing an intelligent system. Chapter 4, gives an overview of the intelligent system development, its testing and describes the stages in the development of the system. The conclusions and directions for future work are given in the final section.

II. DIGITAL TRANSFORMATION IN SMART AGRICULTURE

Digital transformation is present in all industries, whereby the literature research by Kutnjak et al.. [8] presents case studies across industries according to NACE classification. Digital platforms are one of the possible ways to interconnect producers, sellers, and distributors / final users in "smart agriculture", the concept is researched by Tomičić-Pupek et al.. [9]. All these technologies (digital platforms, IoT, Big Data, Machine Learning, …) are employed to add value to the whole supply chain from producers to customers by innovations and possible smart use of platform software as stated in Tomičić Furjan et al. [10]. Emerging technologies could support this way of transformation of product, processes, and business model in many parts of agriculture as already mentioned in Zaripova et al. [1] and Agrawal and Kamboj [3].

One part of "smart agriculture digital transformation" is wine production where producers need to get better decisions and managing the wine production in a "smart way" [11] and smarter use of technologies such as IoT, Big Data, Data Analytics, and Machine Learning helps to support processes such as monitoring vineyards researched by Trilles et al. [12], wine testing by data mining techniques in Cortez et al. [13], quality assessment in Ribeiro et al. [14], quality prediction by Feilhauer [15], measurement by IoT in wine industry [16] and transformation of wine production by use of IoT by Robinson [17].

There are several previous research papers dealing with building predictive models for wine quality prediction. e.g. Nachev and Hogan [18] examined the efficiency of data mining techniques to predict product quality from physicochemical data. Ribeiro et al. [14] investigated the physicochemical properties of wine and tried to link them to the output attribute of wine quality. Authors used various machine learning algorithms to develop the predictive model: linear regression, decision tree, and artificial neural network.

III. Methods and Data

In this chapter, data and appropriate methods will be described briefly to introduce data analysis and development of the intelligent system.

A. Methodology

SDT (Structured Decision Task) methodology was applied in the development of our intelligent systems. SDT methodology was proposed by Attar Software in [19]. The SDT methodology consists of four steps: problem modeling, knowledge structuring, knowledge acquisition, and knowledge testing.

The scope of the intelligent system is to define the form of a problem-solving model. Our intelligent system was designed for predicting wine quality level based on the characteristics of red and white wines.

The model was converted into a structured hierarchy of decision-making tasks based on the knowledge gathered by IoT data and as a result of objective tests, conducted in the laboratory. The obtained knowledge was represented in the form of decision rules in the intelligent system. Decision rules are obtained by means of a decision tree developed by the CRISP-DM process [20]. Other methods of machine learning could be suggested such as by Oreški et al.l [21], Cortez et al. [13]. The decision tree was selected following guidelines by Oreški et al. [22].

Finally, knowledge testing was performed by means of data exchange between the system and its users. The machine learning approach of the decision tree was used to create a structured hierarchy of tasks. The decision tree recursively partitions data according to a relationship between the predictors and output variables. Such an approach develops a predictive model in form of a decision tree. The decision tree algorithm searches all possible splits of predictors to best predict the output. These splits are done recursively to form a tree of decision rules tool.

B. Data description

Data sets used in this research were created by using and studying samples of red and white wines. The data set was created by laboratory testing such as PH value and IoT sensors. The output variable is constructed as subjective research that came up as a median of at least three assessments given by specific wine experts. Each of the experts included in the research evaluated the quality of the wine with a number, from the scale in the range from 0 to 10, where 0 represents - very bad and 10 - excellent.

Research data are related to Vinho Verde which represents a range of Portuguese wines that have a very distinctive taste. As their name suggests, it is a "green wine" and it does not mean its color but the exceptional youth and immaturity, the peculiarity of which is the recommended period of consumption within a year from harvest. The wine has a fresh taste that targets many consumers, with an alcohol content of up to 11%. The wine comes from the Minho region in the north of Portugal.

The data set consists of 4898 instances for white wines and 1599 instances of red wines. There were 11 input variables and one output variable. Input variables are based on physicochemical tests. The output variable is a subjective evaluation of wine experts. Each variable with its respective ranges is presented in Table 1.

TABLE I. Variable Description

Variable	Range
Fixed acidity	2 - 6
Volatile acidity	0 - 1
Citric acid	0 - 1
Residual sugar	0 - 150
Chlorides	0 - 1
Free sulfur dioxide	0 - 70
Total sulfur dioxide	0 - 250
Density	0,95 - 1,1
pH	0 - 14
Sulphates	0 - 3,7
Alcohol	0 – 25
Quality score	1 - 10

(Source: Authors)

C. Data understanding

In order to gain insight into associations between variables, initial correlation analysis was performed before data modeling. Correlation between variables indicates if one variable changes in value, does the other variable tends to change in a specific direction. Understanding correlation can be very useful since indicates can we use the value of one variable to predict the value of the other variable. Such correlation analysis serves as a basis for predictive models' development and consequently, intelligent system deployment.

TABLE II. Correlation Analysis

Variable	Variable	R red wine	p	R white wine	p
quality	alcohol	0,4762	<,0001	0,4356	<,0001
quality	sulphates	0,2514	<,0001	0,0537	0,0002
quality	citric acid	0,2264	<,0001	-0,0092	0,5193
quality	fixed acidity	0,1241	<,0001	-0,1137	<,0001
quality	residual sugar	0,0137	0,5832	-0,0976	<,0001
quality	free sulfur dioxide	-0,0507	0,0428	0,0082	0,5681
quality	pH	-0,0577	0,0210	0,0994	<,0001
quality	chlorides	-0,1289	<,0001	-0,2099	<,0001
quality	density	-0,1749	<,0001	-0,3071	<,0001
quality	total sulfur dioxide	-0,1851	<,0001	-0,1747	<,0001
quality	volatile acidity	-0,3906	<,0001	-0,1947	<,0001

(Source: Authors)

Wine quality has the highest correlation with alcohol level, for both red and white wines. A positive correlation indicates higher quality for wines with a higher level of alcohol. The level of sulphates has the second-highest positive correlation with wine quality for red wines. Volatile acidity has the highest negative correlation for red wines, indicating a decrease in wine quality level when volatile acidity increases. Chlorides have the highest negative correlation with wine quality for white wines. All of those correlations are statistically significant.

IV. Intelligent System Development

Intelligent systems comprise the knowledge and are capable of making conclusions based on the available data. Knowledge is a basic component of intelligent systems. The goal of knowledge representation, is to represent knowledge in a way that enables reasoning (i.e., making inferences) on its basis. Rules, semantic networks, and decision trees are among the most commonly used knowledge representation techniques in intelligent systems. Therefore, fundamental steps of intelligent systems development are focused on knowledge acquisition and knowledge structuring.

By following SDT methodology steps, intelligent system development is explained in the following subsections.

A. *Problem modelling*

In the problem-modeling phase, the problem is defined. According to the SDT methodology, our intelligent system belongs to the "classification" and "subjective estimation" groups of intelligent systems, based on the task they perform, since the operation of this intelligent system relies on assessment based on some parameters that refer to subjective estimation of the wine experts.

B. *Knowledge acquisition*

In order to gain the knowledge necessary for our system development, a publicly available data set was used available here [6] and produced by [13]. Description of data set characteristics is given in the previous chapter.

C. *Knowledge structuring*

Knowledge structuring is the stage where the model is transformed into a structured hierarchy. In our case, knowledge and processes are divided into several nodes built-in within decision trees represent in Figure 1 for red wines and Figure 2 for white wines.

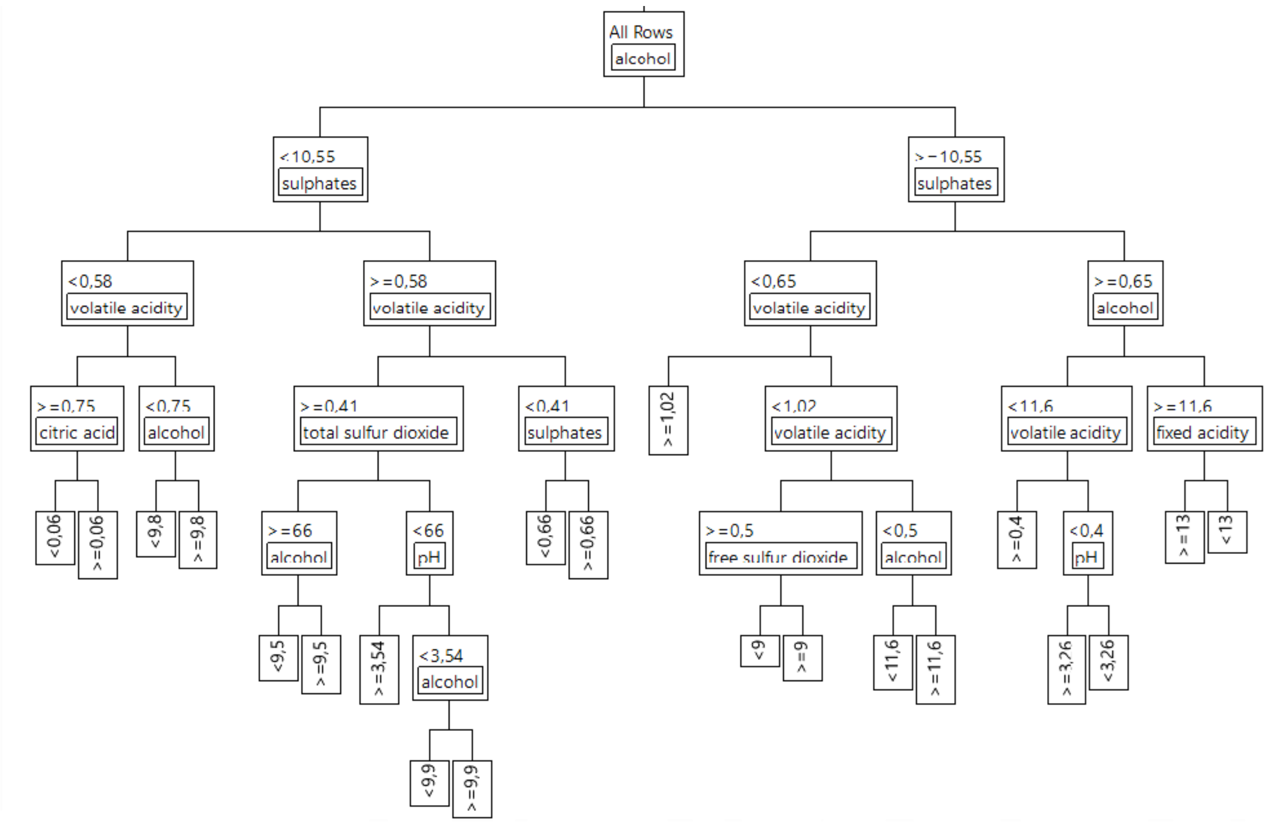

Figure 1. Decision tree for red wine quality prediction

In both cases, splitting is performed several times to achieve quality models. The resulting tree-shaped graphs give us models of possible prediction paths. Flow structure in each inner node represents a test on the attribute, each branch represents the test result and each leaf represents a class value (wine quality). The root-to-leaf path demonstrates the classification rules. For instance, first path in Figure 1 states: IF (alcohol<10,55 AND sulphates <0,58 AND volatile acidity >= 0,75 AND citric acid < 0,06 THEN quality=4,5). As seen, its graphic representation is intuitive what facilitates process automation. Inflow structure for both, red and white wines, the alcohol level is in the root node, indicating that this attribute gives the most of information for quality prediction. This is confirmed by the results of sensitivity analysis (Figure 4.)

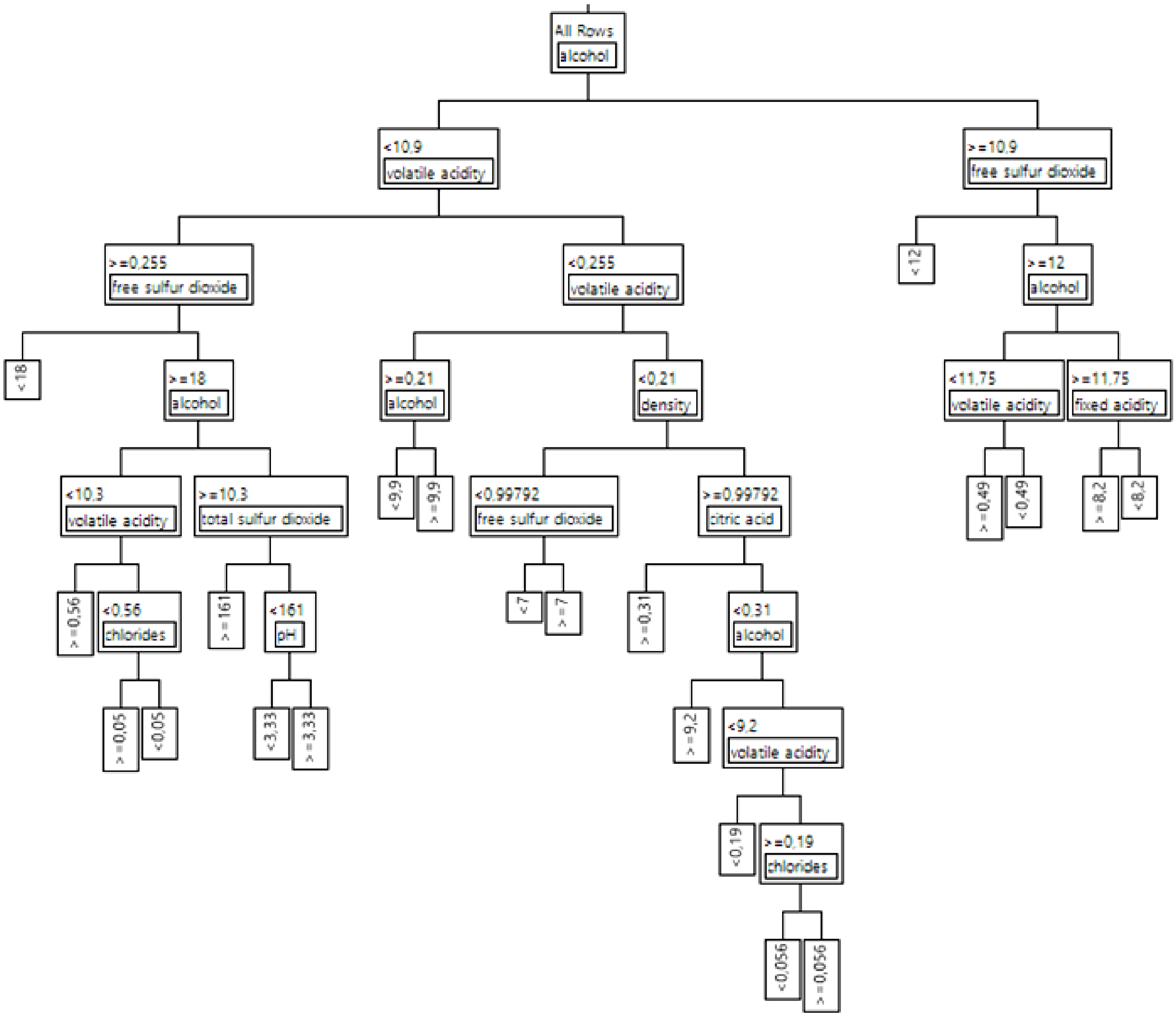

Figure 2. Decision tree for white wine quality prediction

To evaluate the quality of obtaining models model evaluation was performed through 5-fold cross-validation and using RSquare and RMSE as quality metrics (see Table 3).

TABLE III. MODEL EVALUATION

Quality metrics	Red wine	White wine
Model reliability (RSquare)	0,536	0,450
Model error (RMSE)	0,345	0,423

(Source: Authors)

The root means square error indicates differences between the response and p: the fitted probability for the actual wine quality. Smaller values indicate a better fit. RSquare gives a proportion of the variance in the dependent variable wine quality that is contained in independent variables. Results for our models are shown in Table 3. In both cases, red and white wines, models could be generalizable on data beyond the used dataset.

D. Knowledge Testing

Once the system was developed, testing for bugs was performed, following by over a hundred tests running to test each task as well as parts of tasks separately. Knowledge testing was performed in two ways: first, by inserting data to see if the evaluations are correct based on the raw data, and secondly, fully through the screens. The testing example was made by 10 runs through the decision tree can be seen in Figure 3 as one example. Users can input their wine sensor or laboratory values and calculate quality. In the lower part of the screen results and accuracy can be seen. On the right part of the screen, there is a number of test rows and results. User can also see correct predictions and their percentage.

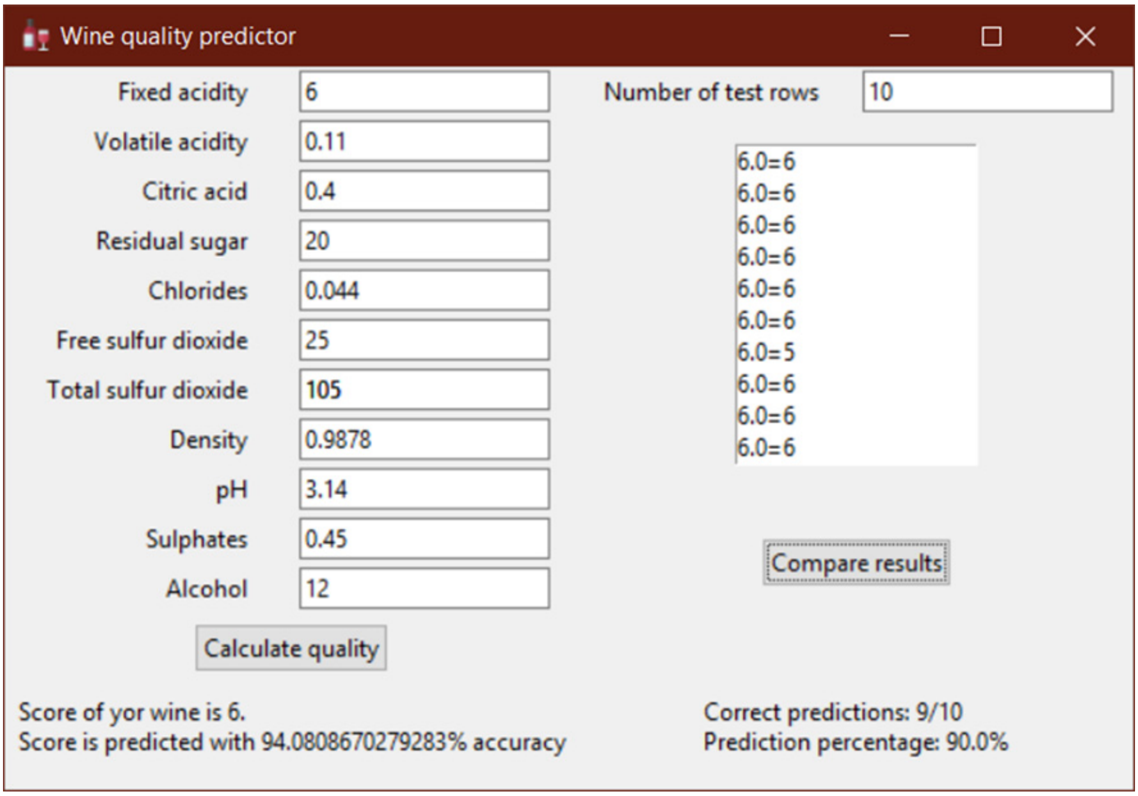

Figure 3. Wine quality predictors

The last activity in the testing phase was performing sensitivity analysis. Sensitivity analysis refers to, the process of varying parameters among a model's input variables space to understand the extent to which different variables affect model outcomes and identify input variables that are most important for the prediction of a given task.

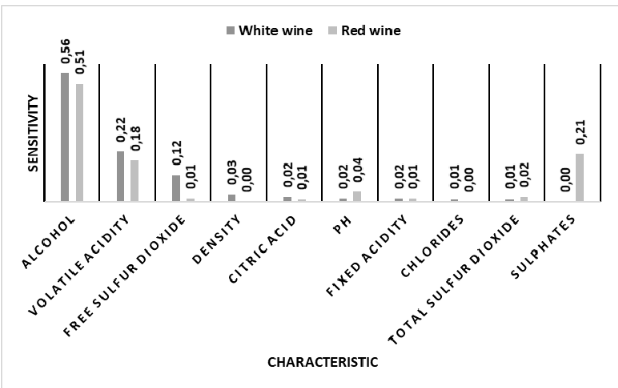

Figure 4. Sensitivity analysis

Hereinafter, Figure 4 indicates the results of sensitivity analysis performed for both, red and white wines. In both cases, the percentage of alcohol shown to be the strongest predictor of wine quality. However, there are differences between red and white wines: volatile acidity stands out as the second most important predictor for white wines, whereas the level of sulphates in wine seems to be the second most important predictor for red wines.

V. Conclusion

In this article, we proposed an intelligent system to support the automatization of wine quality assessment as a case study for digital transformation and smart agriculture. The application of modern technologies, their lower cost, and wide availability make it possible to modernize the traditional ways of wine quality assessment and develop new features that lead to the successful realization of the assessment. Even if this intelligent system was the only MVP as an experiment in a research project the major contribution of this research is therefore reflected in the novel approach to the wine quality assessment which is based on artificial intelligence technologies.

However, this research has some limitations. Variables included in the research and examined by some wine testing experts only as starting point in wine quality assessment by human organoleptic testers. According to experts, there are also other factors influencing wine quality that will be examined in future research. Furthermore, our model was built on a dataset consisting of Portuguese wines Vinho Verde, and our results could not be generalized to other types of wine. But, this kind of approach can be applied to other wine types.

Future research, will include other wine types and also examine the capability of other machine learning approaches, such as for ensembles of classifiers to develop more accurate models.

Acknowledgment

This research has been conducted as a part of the wider research in the project Competence Centre for Digital Transformation of the Food Industry in Rural Areas, which is co-funded by European Union through the European Regional Development Fund (ERDF).

References

[1] R. Zaripova, M. Tyurina, O. Rocheva, A. Chupaev and A. Sharifullina, "Major Trends in the Digital Transformation of Agriculture," Advances in Economics, Business and Management Research, vol. 131, 2020, pp271-275.

[2] N. Kadoić, K. Tomičić-Pupek and N. Vrček, "Decision making on Digital Platforms in Agriculture" In Proceedings of the 43rd International Convention on Information and Communication Technology, Electronics and Microelectronics MIPRO 2020, K. Scala, Eds. Opatija: Croatian Society for Information, Communication and Electronic Technology – MIPRO, 2020, pp 1731-1736.

[3] K. Agrawal and N. Kamboj, "Smart agriculture using IOT: A futuristic approach," In International Journal of Information Dissemination and Technology, vol. 9 (4), 2019, pp. 186-190.

[4] S. Ghavate and H. Joshi, "Smart Farming using IOT and Machine Learning with Image Processing," In International Journal of Current Engineering and Technology, Special Issue-8, pp 16-19, February 2021.

[5] R. Mangiaracina, G. Miragliotta. A. Perego, G. Salvadori and A. Tumino, "Adopting the Internet of Things technologies in Agriculture: an analysis of the benefits in the organic wine industry," In XXI Summer School Francesco Turco - Industrial Systems Engineering, 2016, pp 48-53.

[6] F. Mazzetto, R. Gallo and P. Sacco, "Reflections and Methodological Proposals to Treat the Concept of "Information Precision" in Smart Agriculture Practices", In Sensors, vol. 20, 2020, pp. 1-27.

[7] UCI Machine learning Repository (2009.) Wine Quality Data Set. Available from: https://archive.ics.uci.edu/ml/datasets/wine+quality

[8] A. Kutnjak, I. Pihir and M. Tomičić Furjan, "Digital Transformation Case Studies across Industries – Literature Review," In Proceedings of the 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics MIPRO 2019, K. Skala, Eds, Rijeka: Croatian Society for Information and Communication Technology, Electronics and Microelectronics - MIPRO, 2019, pp. 1498-1503.

[9] K. Tomičić-Pupek, I. Pihir and M. Tomičić Furjan, "The role of perception in the adoption of digital platforms in agriculture," In Proceedings of the 43rd International Convention on Information and Communication Technology, Electronics and Micro-electronics MIPRO 2020, K. Scala, Eds. Opatija: Croatian Society for Information, Communication and Electronic Technology – MIPRO, 2020, pp. 1429-1434.

[10] M. Tomičić Furjan, L. Hrustek and I. Pihir, "Value of innovation platforms in Agriculture," In Proceedings of the 43rd International Convention on Information and Communication Technology, Electronics and Microelectronics MIPRO 2020, K. Scala, Eds. Opatija: Croatian Society for Information, Communication and Electronic Technology – MIPRO, 2020, pp. 1451-1456.

[11] Digital Transformation Monitor. (2017). Smart vineyard: management and decision making support for wine producers. Available from: https://ati.ec.europa.eu/sites/default/files/2020-06/Smart%20vineyard-%20management%20and%20decision-making%20support%20for%20wine%20producers%20%28v1%29.pdf

[12] S. Trilles Oliver, A. González-Pérez and J. Huerta Guijarro, "An IoT proposal for monitoring vineyards called SEnviro for agriculture," In *Proceedings of the 8th International Conference on the Internet of Things (IOT '18)*, 2018, pp. 1–4.

[13] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, vol. 47 (4), 2009, pp. 547-553.

[14] J. Ribeiro, J. Neves, J. Sanchez, M. Delgado, J. Machado and P. Novais, "Wine vinification prediction using data mining tools," In *ECC'09 Proceedings of the 3rd international conference on European computing conference*, 2009, pp. 78-85.

[15] P. Feilhauer,"Prediction of wine quality," 2013 Available from: https://www.yumpu.com/en/document/read/20592556/prediction-of-wine-quality

[16] G. Masetti, F. Marazzi, L. Di Cecilia and L. Rovati, "IoT-Based Measurement System for Wine Industry," In Workshop on Metrology for Industry 4.0 and IoT, Brescia, 2018, pp. 163-168.

[17] D. Robinson, "Technology in the wine industry: How IoT is transforming vineyards," In NS Agriculture, 2019, Available from: https://www.nsagriculture.com/news/technology-wine-industry-iot-vineyards/

[18] Nachev A., Hogan M. (2013). Using data mining techniques to predict product quality from physicochemical data. In Proceedings on the International Conference on Artificial Intelligence.

[19] A. Al-Attar, "Rule induction from mythology to methodology, Research and Developments," In Expert Systems VIII, London, 1991.

[20] D. Oreški, I. Pihir and M. Konecki, "CRISP-DM process model in educational setting," In 20th International Scientific Conference on Economic and Social Development: Book of Proceedings, Prague, Czech Republic, VADEA, 2017, pp. 19-28.

[21] D. Oreški, M. Konecki and I. Pihir, "Predictive Modelling of Academic Performance by Means of Bayesian Networks," In *47th International Scientific Conference on Economic and Social Development Economic and Social Development: Book of Proceedings*; Prague, Czech Republic, VADEA, 2019, pp. 435-441.

[22] D. Oreški and N. Begičević Ređep, "Data-driven decision-making in classification algorithm selection," In Journal of Decision Systems, vol . 27, 2018, pp. 248-255.