DOI: 10.24326/fmpmsa.2017.69

APPLE TREE YIELD ANALYSIS USING DATA MINING

Jędrzej TRAJER, Iwona PIETRZYCKA, Ewa PIOTROWSKA, Ewa GOLISZ
Faculty of Production Engineering, Warsaw University of Life Sciences – SGGW, POLAND
E-mail of corresponding author: jedrzej_trajer@sggw.pl

Keywords: apple tree yielding, Data Mining, sustainable gardening production

ABSTRACT

The paper analyses an orcharding farm that specializes in apple trees production. Based on the data for the period of 2008-2014, the authors analysed the main factors that might have impact on apple yield. A computer system for assessment of apple trees cultivation efficiency that aids in making appropriate decisions allowing for obtaining the highest yield, was proposed. The system was developed using selected Data Mining techniques such as cluster analysis and Kohonen networks. The system may be useful for decision support in sustainable horticulture production, and thus contributes to the development of sustainable agriculture. Although its quality is acceptable it still requires improvement using a bigger dataset.

INTRODUCTION

Orcharding is one of horticulture production divisions, characterized by unique requirements, considerable obligations as well as both thorough and specialized knowledge. Fruit farmers must meet extremely high quality norms set by consumers' and food processing industry's requirements, related to taste, fruit firmness, abundance in vitamins and minerals, and appealing look encouraging the consumer to buy and eat particular fruits. To obtain yields of high quality and quantity, it is important to possess universal soil, climatic and cultivar-related interpretation, which determines the quality and quantity of yields. In order to efficiently produce safe, high quality food, and simultaneously protect and improve natural environment, as well as achieve farmer's social and economic living conditions (sustainable agriculture), it is important to comply with the highest standards set by experience and practice, supported by specialist knowledge in areas related to orchard maintenance, Zaliwski (2007). Running an orchard is an extremely costly and long-term undertaking. Numerous decisions related to selection of the area, cultivars, storage facilities and sale. Every year, many hectares of new orchards are grown, and since it is permanent crop, all the above factors that have impact on positive or negative results of these actions ought to be rationally analysed, which is best done with the help of appropriate decision support system. Such a computer system requires a regularly updated database to work properly. Appropriate combination of these elements (information + a processing system) may, in effect, bring satisfactory results, Bielecki (2001). This work presents a conception of such a computer system to support decision-making related to obtaining the best yields in sustainable apple production.

APPLE TREE YIELDING

In order to reach the highest production goals, it is necessary to correlate the knowledge of parameters that play a decisive role and have impact on their effectiveness. The above knowledge allows for taking effective actions, which result in high yields. The most crucial factors that have impact on apple yielding include: rootstock, cultivar, tree age, spacing distance between trees as well as soil class and irrigation, Pietrzycka (2016). Other important yielding factors comprise weather and climatic conditions, selection of appropriate plant protectants and fertilization. Thorough knowledge and appropriate use of these factors may contribute to sustainable apple trees cultivation.

Rootstock – is a plant used for shield budding of cultivars. Starting a modern orchard requires selection of an appropriate rootstock. Rootstocks for apple trees can be divided into four categories: dwarf, semi-dwarf, semi-standard and standard, with the standard being the largest and the dwarf being the smallest, Czynczyk (2012). In Poland, soils are mostly luvisols.

Therefore, dwarf rootstocks such as M9, P22 or P59, Rejman (1997) are not appropriate. Better results on lighter soils, with lower content of humus are obtained with semi-dwarf rootstocks, i.e. M7, M26, P60 or P14 – they are appropriate for orchards grown in lower lying areas.

Cultivar – most species of orcharding plants have numerous cultivars, which differ in yield, frost- and disease-resistance, ripening time, and utility value, Pieniążek (1995). Production value has largely depended on the market and consumers, and these factors have a significant influence on the selection of cultivars by farmers, Sobiczewski (2013). Only cultivars, for which the demand is stable and high, are profitable. They include such varieties as Alwa, Boskoop, Gala, Idared, Lobo, Szampion, Ligol, Gloster, Decosta, Empire, Mutsu, Golden Delicious, Jonagold, Red Prince.

Irrigation – providing trees with appropriate amount of water, especially in critical phases of development allows for high and good quality yields. It is estimated that for optimal growth and yielding, orchard plants require yearly precipitation of 700-850 mm. However, due to the fact that such values are rarely reached, it is necessary to consider the possibility of installation of artificial watering system when starting and orchard.

Spacing between trees – the best spacing is 1500-3000 dwarf trees or 1000-1500 semi-dwarf trees per hectare, Sobiczewski (2013). Decreasing the spacing between trees by planting over 3000 per hectare significantly increases the cost of planting material, and may also result in fruit quality deterioration, and make pest and disease protection of trees more difficult. In a popular row system of planting, dwarf apple trees are planted with the spacing of 1.0-2.0 m between the trees and 3.5 m spacing between rows, while for semi-dwarf apple trees, the spacing between the rows should be 4.0 m, and the spacing between trees in a row – 1.5-2.5 m.

Soil class **–** arable lands are divided into 9 valuation classes, with the following symbols: I, II, III a, III b, IV a, IV b, V, VI and VI Rz. When choosing the plot of land to start an orchard, it is necessary to consider land configuration, soil fertility, and underground water level.

Trees age – when planning to start an orchard it is necessary to consider the fact that full tree productivity is reached in a specified time interval. For apple trees, the full yield stage falls for years 5-14. In extremely favourable conditions, the trees may be kept until they are 20 years old but it must be remembered that 15-20-year old trees the yield is lower.

MATERIALS AND METHODS

The dataset contains 175 cases obtained from observations at the turn of 2008-2014, Pietrzycka (2016). The most important factors determining the quantity and the quality of yields, namely, cultivar, rootstock, number of trees per hectare, soil class, irrigation, and trees age, were selected for analysis. The study involved 16 different cultivars of apple trees. In order to simplify the model, based on cluster analysis, which uses EM (*Estimation Maximization*) algorithm Witten and Frank (2000), the cultivars were divided into two groups. The cultivar was described by the following characteristics: fruit size, sensitivity to frost and diseases, harvest date and bloom dates, table 1.

Table 1. Results of classification of apple cultivars

Apple cultivar	Fruit size	Sensitivity to diseases	Sensitivity to frost	Harvest date	Bloom date	Classification results
Idared	3	2	1	1	1	**2**
Gala	1	2	3	1	2	**1**
Szampion	3	3	3	3	2	**1**
Ligol	3	2	1	2	3	**2**

Apple cultivar	Fruit size	Sensitivity to diseases	Sensitivity to frost	Harvest date	Bloom date	Classification results
Gloster	3	2	1	3	3	**2**
Jonagored	3	2	2	2	2	**1**
Decosta	2	2	2	2	1	**1**
Mutsu	3	1	3	3	2	**1**
Lobo	2	3	1	1	3	**2**
Alwa	1	3	1	3	3	**2**
Empire	2	1	2	3	2	**1**
Golden D	2	2	3	3	3	**2**
Boskoop	3	1	1	3	2	**1**
Red Prince	3	3	3	2	1	**1**
Rubinstar	3	2	2	2	1	**1**
Elize	1	2	2	3	2	**1**

The following rootstocks are used in the plantation: M26 labelled 1 and M9 – 2. Irrigation can be assigned two values in the model, namely 0 and 1, 0 denotes no irrigation, and 1 – irrigation exists. Soil classes were divided into 5 categories, see table 2.

Table 2. Soil classification

Assigned value of soil class	Soil class	Description of soil classes
1	IIa	Good soils
2	IIIa, IIIb	Quite good soils
3	IIIb, IVa	Medium quality soils, better
4	IVa, IVb	Medium quality soils, worse
5	IVb, V	Poor soils

Kohonen networks (*Self Organizing Map*) were used to analyse the yield, Kohonen (1982). The input variables include the cultivar, rootstock, number of trees, soil class, irrigation and trees age, and the output layer consists of 169 (13x13) neurons that make up a topological map, Fig. 1.

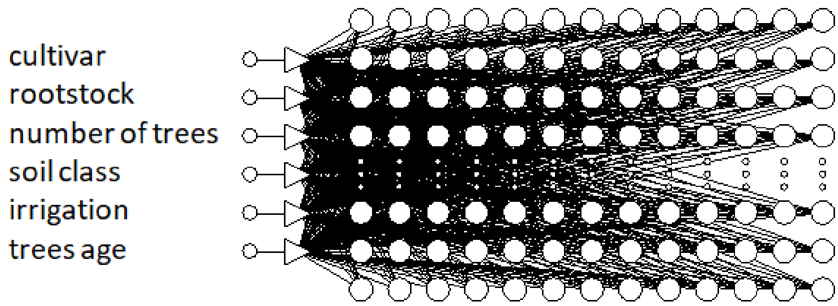

Fig.1. Kohonen network for the analysis of apple yields

When developing a topological map, the number of neurons should be approximately the same as the number of the cases being analysed (here 175), but it should not be greater than this number as the resultant 'empty' cases might have a negative impact on classification quality.

RESULTS AND DISCUSSION

The system was developed using Kohonen network and data analysis software *Statistica 12.5* StatSoft, Poland. The network's ability to activate neighbouring neurons acquired during learning process when similar input data is input, allows for interpretation of input signals set as a map mapping topological relations between input data in the space of input signals. The topological map (fig. 2) shows winner neurons (number of i-th case denoted as 'vi').

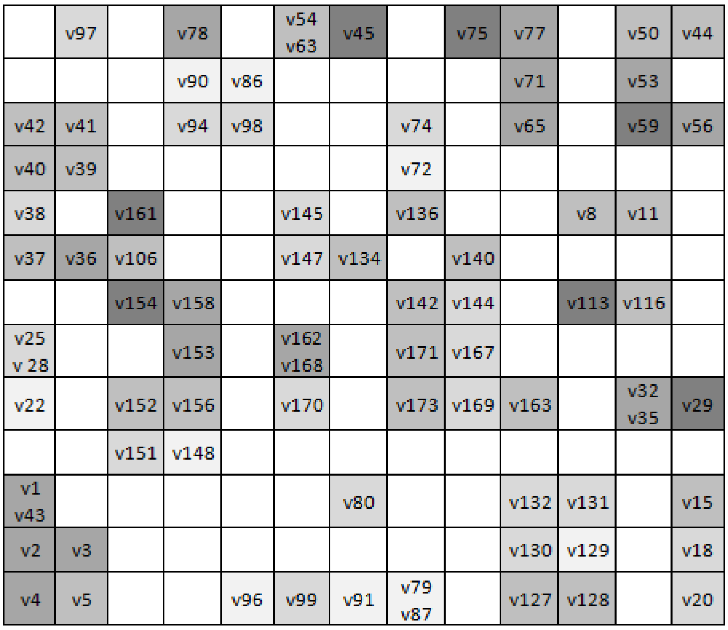

Fig.2. A topological map that shows apple trees yielding classification results.

Individual regions of the map were shaded depending of the quantity of yield, table 3.

Table 3. Yield quantity denotations

Yield quantity [1000 kg]	Yield rank	Colour shade on the map
(0-20)	very low	
<20-35)	low	
<35-50)	average	
<50-65)	satisfactory	
more than 65	high	

The results of sensitivity analysis for this model indicating the rank of individual input variables are presented in table 4.

Table 4. Sensitivity analysis for the apple trees yielding classification model

Predictor:	Cultivar	Rootstock	Number of	Soil class	Irrigation	Age
Rank	2	1	4	5	3	6
Error	0.418	0.427	0.306	0.204	0.308	0.147
Quotient	16.85	17.186	12.309	8.219	12.42	5.908

The functioning of the developed system for supporting decisions regarding the assessment of efficiency of apple trees cultivation and yield may be considered correct although its quality requires refining. The growth index p, Trajer et al. (2012), used for the assessment of the classification, has the value of p = 0.526, which indicates that the system requires refining. Growth is defined as a percentage of positive hits in the set of positive classifications PTK to the percentage of positive hits in the whole set of data PTD, formula (1):

$$p = \frac{PTK}{PTD}, \tag{1}$$

The efficiency of the system requires improvement such as the use of larger dataset and selection of better predictors. Based on sensitivity analysis it was determined that the importance of factors that have impact on yield quantity is as follows (in descending order of importance): rootstock, cultivar, number of trees per area unit, soil class, irrigation and trees age.

CONCLUSIONS

The developed computer system, based on selected decision variables, allows for satisfactory support of decisions related to obtaining the best yields in sustainable apple production, which was confirmed by both the results of logical and empirical verification. The decision variable, i.e. apple cultivar, requires refining as grouping the cultivars in the study by using cluster analysis into two groups only, poorly describes the variability of this parameter. The improvement of the system efficiency might be obtained by analysing a greater number of cases, which would also consider other factors significant for this orcharding production. Numerous empty fields on the topological map indicate that the quality of the system might also be improved by optimizing the size of the map. Sensitivity analysis indicates that the most significant factors that have impact on yield quantity include rootstock and apple tree variety, which is confirmed by information contained in relevant literature and empirical observations.

REFERENCES

Bielecki, W. (2001). Informatyzacja zarządzania. *PWN*. Warszawa.

Czynczyk, A. (2012). Szkółkarstwo ogrodnicze. *Powszechne Wydawnictwo Rolnicze i Leśne*. Warszawa.

Kohonen, T. (1982). Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics, 43*.

Pieniążek, S. (1995). Sadownictwo. *Państwowe Wydawnictwo Rolnicze i Leśne*. Warszawa.

Pietrzycka, I. (2016). Analiza plonowania jabłoni z wykorzystaniem Data Mining. *Praca magisterska*. WIP SGGW. Warszawa.

Rejman, A. (1997). Szkółkarstwo roślin sadowniczych. *Państwowe Wydawnictwo Rolnicze i Leśne*. Warszawa.

Sobiczewski, P. (2013). Metodyka integrowanej ochrony jabłoni dla producentów. *Instytut Ogrodnictwa*. Skierniewice.

StatSoft Polska. www.statsoft.pl

Trajer, J., Paszek, A., Iwan, S. (2012). Zarządzanie wiedzą. *PWE*. Warszawa.

Witten, I.H., Frank, E. (2000). Data Mining. *Practical Machine Learning Tools and Techniques*. New York. Morgan Kaufmann.

Zaliwski, A.S. (2007). Systemy wspomagania decyzji w nowoczesnej produkcji roślinnej. System doradztwa w zakresie zrównoważonej produkcji roślinnej. *IUNG-PIB.* Puławy.