Substrate Wastage Prediction Analysis of the Heat-shrinkable Labels Manufacturing Process

Dr. Krzysztof Krystosiak

Keywords: substrate wastage, heat-shrinkable labels, prediction methods, data mining, flexographic printing, artificial intelligence, neural networks

Abstract

Substrate, in the form of paper, foil, film, or fabric, is the main part of the total manufacturing costs in the printing industry. It is already known that substrate is a major cost in printing & packaging processes (Levenson, Parsons, 2018), and there are many developed ideas on how to avoid unusual wastage of the substrate. Many methods and tools are used to improve the total effectiveness of printing processes, focusing on substrate wastage. Many printing press manufacturers also develop new printing machines that use less substrate for print setup, which is the main part of the used substrate in the printing process. But still many printers everyday fights with issues that have an impact on increasing substrate wastage.

This paper will show a way to predict the total substrate wastage per work order, based on a few vital factors. Prediction analysis and its effects on the example of the heat-shrinkable labels manufacturing process will be shown. The main steps in the manufacturing process of heat-shrinkable labels are divided into four stages, resulting in a finished product of a wound roll of heat-shrinkable labels. These stages are printing, slitting, seaming, and inspection (Kipphan, 2001; Kit, 2009).
In the manufacturing practices of most printing companies, there are collected data and records of process parameters. Gathering information from this data in the form of developed models and rules, which uses statistical or computational methods from the area of Artificial Intelligence, like Artificial Neural Networks, Boosted Trees, Support Vector Machines, and others are subjects of interdisciplinary fields of science called Data Mining (Krystosiak, 2019).

There will be shown how developed data mining models can be used for the prediction of substrate wastage levels for every new design of printed products. The manufacturing process of every design of a heat-shrinkable label has a lot of

Toronto Metropolitan University

factors and variables. Some of them are more significant, and some of them are not. This paper aims to choose significant factors and compute a model in the learning process by using collected data. In the end, it is important to predict the level of substrate wastage and calculate it in the total cost of each new work order of printed product, to avoid high substrate wastage occurrence during the printing process.

Introduction

Shrink sleeve labels are produced using flexographic printing technology, more specifically it is a rotary printing method that uses a flexible printing form to compensate for the surface irregularities of the substrate. Ink is transferred from the inkpot onto the flexible printing form by an anilox roller and from there directly onto the substrate. Labels printed with this technique are used to perform unstable mass production. The group includes, for example, shrink-sleeve labels (Kipphan, 2001).

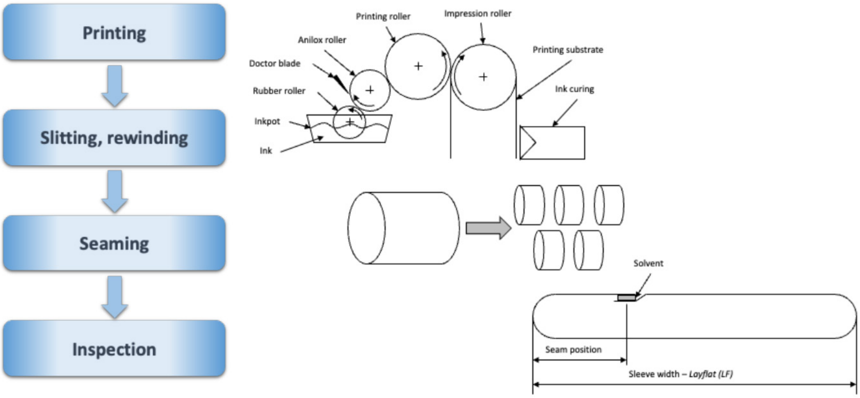


Figure 1. *Heat-shrinkable labels production process*

A label of this type can be printed on its entire surface and due to its thermoplastic properties, does not pose any restrictions as to the shape of the package. This is because the foil sleeve, after shrinking, adheres closely to the walls of the target container, perfectly adapting to its shape. The foil is usually printed on its inside, which makes the print after application resistant to abrasions and scratches that may occur, for example, during transport (AWA, 2014).

Problem

Determining the level of material or substrate wastage level in the production of labels and packaging is one of the greatest challenges. There is no such thing as one equation for all types of labels or packaging, for different printing techniques and technologies. And substrate wastage level is very important to establish and control because it determines the total yield of each work order.

In many situations, the author of this article has met with different calculations of substrate wastage, like below:

- In kilograms [kg]
- In percent [%]
- In square meters [m^2]
- In running meters [mb]
- In pieces [pcs]

And from the other side, how to measure and control this key metric for the printing industry:

- For the whole work order?
- For a roll?
- For the operator?

There is a double problem here, on the one hand, we do not know exactly how to calculate this material loss Is it in kilograms, in percent, or in square meters, and on the other hand, what key measure to use for a printing company? This research paper used statistical methods so-called soft modeling.

Methods

Data mining as a field of science already has a stable position, even though it is a relatively young branch of science. The term itself was coined at the end of the 1980s. Data mining has been formed in several different stages because of the evolution of such fields of science as classical statistics, artificial intelligence, and machine learning (Flores, 2011; Siddique & Adeli, 2013).

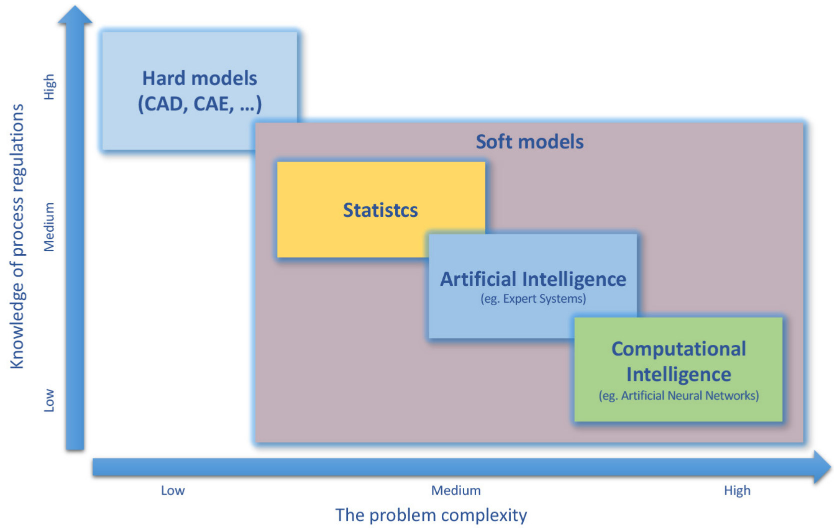


Figure 2. *Choosing the right mathematical model (Tadeusiewicz, 1993)*

Artificial Neural Networks are one of the mathematical modeling tools as a part of soft modeling tools. Artificial neural networks also belong to the elements of artificial intelligence. Models of neural networks are much more primitive than their prototype, i.e., the biological nervous system, yet they allow modeling of poorly structured and algorithmic phenomena and processes. They make it possible to build an effective model without specifying the nature of the relationships between the variables. They also can adapt, and self-realization is acquired in the process of network learning, which is perfectly presented in numerous publications in this field (Nisbet, Elder & Miner, 2009).

Neural networks can be classified based on several distinguishing features, which include the type of input signal, the structure of the network, and the method of their learning. The input signal can be a binary or continuous signal. The first type includes such types of networks as the Hopfield network, Hamming network, or ART1 network. The perceptron and the Kohonen network are representatives of the second type of network (Tadeusiewicz, 1993; Vapnik, 2000).

For this article, statistical analysis used "Statistica" statistical software from the StatSoft company.

Procedure

As a result of the conducted research on the issue of miscalculated substrate wastage, further a thesis was set forward that the use of data mining and machine learning methods in the packaging process of heat-shrinkable labels will allow to reduction production losses and achieve measurable technical, economic, and organizational effects. To verify the above thesis, the following research hypotheses were formulated:

- Using data mining and machine learning tools, it is possible to acquire knowledge, skills, and experience, allowing for the optimal use of the printing process of heat-shrinkable labels.
- The developed predictive models can be helpful in determining the correct substrate wastage of the heat-shrinkable labels.
- There are dependencies in the printing process of heat-shrinkable labels that can be used to build predictive models to use them to properly control substrate wastage.

To this approach, a problem was defined: substrate wastage prediction and its analysis. So, substrate wastage is a dependent variable for this research, and there will be computing a predictive model based on the set of independent variables – discrete and continuous type – listed below:

- Order type [New/Repetable] - Discrete
- Substrate used [m] - Continuous
- Number of rolls [pcs] - Discrete
- Roll amount [m] - Continuous
- Colors/Units [pcs] - Discrete
- Substrate manufacturer [A, B, …] - Discrete
- Printing press [P_1/P_2/…] - Discrete

In the next step will be presented results of the mathematical modeling with the above parameters set. Also, will be presented a different analysis of the data, which was gathered from the production process of the printing company.

Results

As a first step was calculated a Spearman rank order correlation between variables. This type of correlation was used because of the non-parametric data.

Variable	Substrate used [m]	Number of rolls [pcs]	Roll amount [m]	Colors/ Units [pcs]
Substrate used [m]	1.000000	0.667392	0.162484	-0.156023
Number of rolls [pcs]	**0.667392**	1.000000	**-0.576868**	-0.149638
Roll amount [m]	0.162484	**-0.576868**	1.000000	0.034440
Colors/Units [pcs]	-0.156023	-0.149638	0.034440	1.000000

***Table 1.** Spearman rank order correlation*

As we can observe from the above table (Table 1), the number of rolls has a maximum correlation with the substrate used, which is 0.67 and it’s very high. Also, the roll amount variable has a high correlation to the number of rolls, which is 0.58.

This research was performed on the data collected from a flexographic printing company, where basic printing processes were made in a standard process for flexographic printing which is from roll to roll. So, this printing process relay of the number of rolls and every roll change had a significant impact on the substrate wastage, because always it took from 50 to 200 line meters after each roll change to achieve good print quality.

The next step was using a standard Pearson correlation for two continuous variables: wastage and substrate used (Figure 3). It’s obvious to all printing specialists that total wastage is very highly correlated with the substrate used, or in other words –

the work order amount. In this example, the correlation coefficient was 0.80 which is a very high correlation between the two variables.

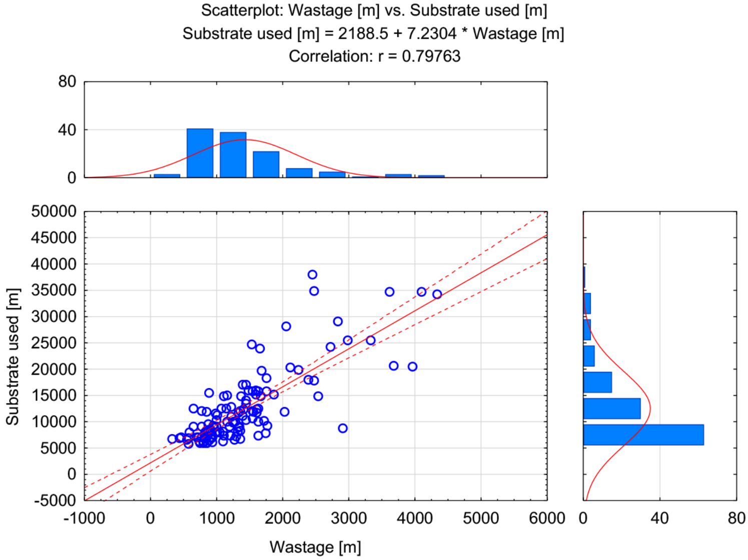


Figure 3. *Total wastage per work order vs substrate used*

The below graph (Figure 4) shows the correlation coefficient for the substrate used and roll wastage. In this analysis, there is no significant correlation because it's only -0.13.

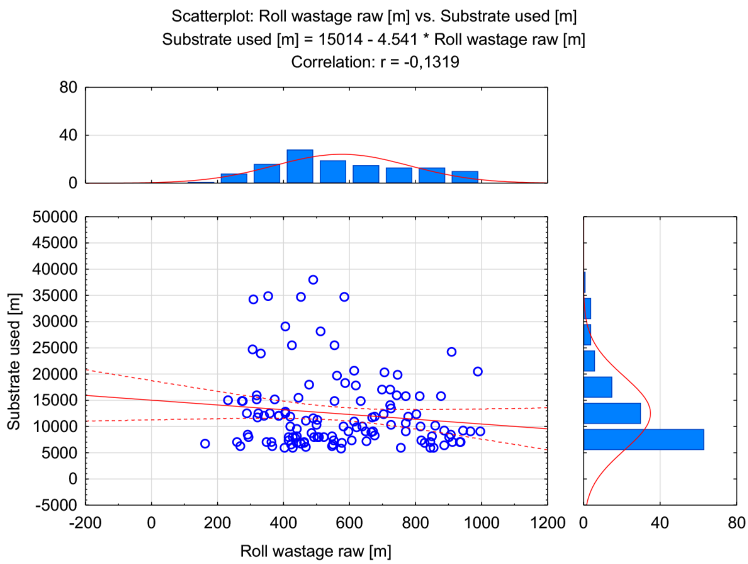


Figure 4. *Roll wastage vs substrate used*

Roll wastage is independent of the substrate used or work order amount, so it is a better Key Performance Indicator (KPI) for printing production and using this metric it is possible to compare many different work orders, large and small.

Next step we can see the results from the analysis of the best predictors of all Data Miner Recipes. Data Miner Recipes is a tool in Statistica StatSoft software to check all prediction tools at once. From this analysis (Table 2) is clearly seen that the first three variables have met the requirements of the probability test and originally are highlighted in red.

Variable	**F – value**	**p – value**
Number of rolls [pcs]	51.76322	0.000000
Substrate used [m]	37.03607	0.000000
Order type	5.01278	0.026989
Substrate manufacturer	2.34261	0.076612
Printing press	2.16188	0.077495
Roll amount [m]	2.12288	0.039072
Colours/Units [pcs]	0.72613	0.668138

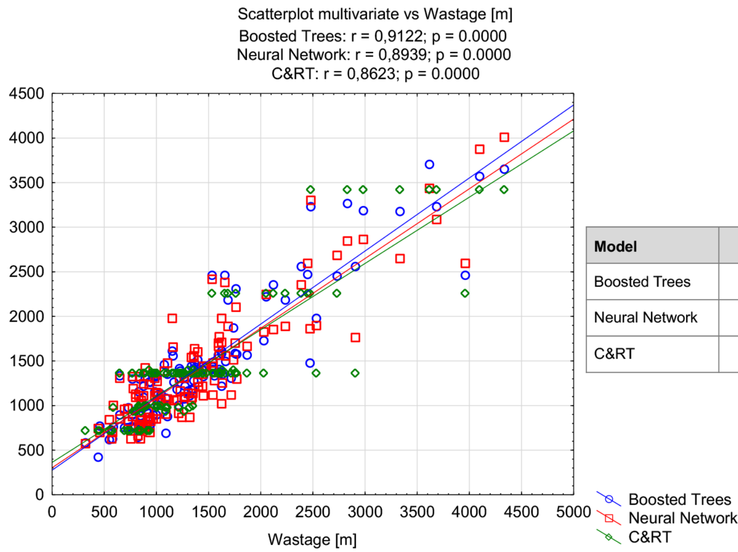
Table 2. *Best predictors for the variable Wastage – all Data Miner Recipes*

Then similar tests were conducted for Classification & regression trees (Table 3) and those results vary from the above test (Table 2) and that's clear that different tools use different algorithms to calculate the best predictors.

Variable	**Variable rank**	**Importance**
Substrate used [m]	100	1.000000
Number of rolls [pcs]	97	0.965813
Roll amount [m]	25	0.250023
Printing press	14	0.142128
Colours/Units [pcs]	11	0.111862
Order type	11	0.112771
Substrate manufacturer	10	0.101369

Table 3. *C&RT model best predictors for variable Wastage*

Finally, all results using tools Boosted Trees, Neural Networks, and C&RT were obtained and shown in the below graph (Figure 5).



Model	Prediction
Boosted Trees	0.912184
Neural Network	0.893857
C&RT	0.862326

Figure 5. *Prediction quality of selected models*

The prediction quality of selected models was calculated on the correlation of the results gained from developed prediction models to the historical data taken from the learning process of developed models. In this example was calculated substrate wastage, so on one side was wastage from the historical data and the results of developed prediction models.

The best results were obtained by using Boosted Trees prediction models and the prediction quality was 0.91 which is a very high correlation. But other results using Neural Networks (0.89) and C&RT (0.86) models were very good too.

At this point, all three prediction models present very high prediction quality and can be used to predict future substrate wastage based on selected predictors.

In the next steps were calculated some statistical analyses like:

- ANOVA (ANalysis Of VAriances) using Kruskal-Wallis rank tests
- ANOVA P-value parameter for multiple comparisons
- Basic statistics in the form of graphs.

Data gathered for this research purpose were non-parametric, so this type of test mentioned above was selected the basic statistics were used for the analysis of medians, quartiles, and number of cases.

As first analyzed the printing press ANOVA. The result based on the Kruskal-Wallis test reaches p = 0.02 (Table 5) which is less than 0.05 and according to the hypothesis test, it can be stated that there are significant differences between variables – which are printing presses.

ANOVA Kruskal-Wallis rank; Wastage [m] (Data) Independent Variable (grouping): Printing Press Kruskal-Wallis test: H (4, N = 123) = 11,65760 p = 0,0201				
Printing Press	**Code**	**N**	**Sum**	**Average**
P_5	1	21	904.000	43.04762
P_2	2	31	2115.000	68.22581
P_1	3	4	184.000	46.00000
P_4	4	42	3022.000	71.95238
P_3	5	25	1401.000	56.04000

Table 4. *ANOVA Kruskal-Wallis rank of Wastage vs Printing Press*

The above analysis shows that there are statistically significant differences in wastage levels on selected printing presses, but still, we do not know which ones differ. In order to check which printing presses generated a higher substrate wastage, the next analysis of the P-value for multiple comparisons should be performed.

P value for multiple (double) comparisons: Wastage [m] (Data) Independent variable (grouping): Printing Press Kruskal-Wallis test: H (4, N = 123) = 11,65760 p = 0,0201					
	P_5	**P_2**	**P_1**	**P_4**	**P_3**
P_5		0.124596	1.000000	0.024164	1.000000
P_2	0.124596		1.000000	1.000000	1.000000
P_1	1.000000	1.000000		1.000000	1.000000
P_4	0.024164	1.000000	1.000000		0.772398
P_3	1.000000	1.000000	1.000000	0.772398	

Table 5. *P value for multiple comparisons: Wastage vs Printing Press*

From the above analysis (Table 5) we can see that there are significant differences based on P-value result 0.02 between printing press P_4 and P_5.

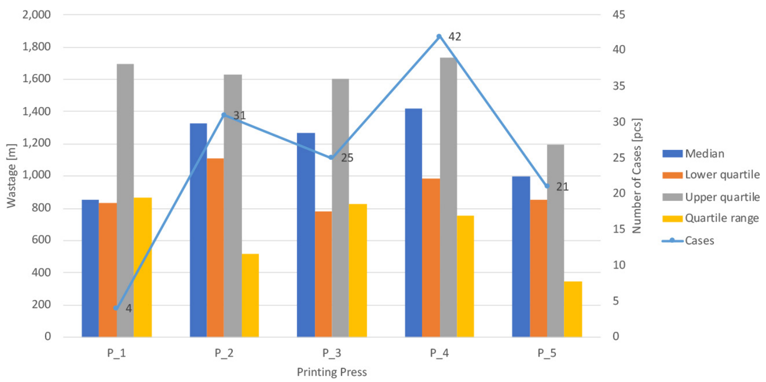


Figure 6. *Basic statistics for the Printing Press*

Basic statistics of the median, quartiles, and number of cases are shown in the above graph (Figure 6). Presented results can be used to see based statistics for the analysis of wastage on selected printing presses.

The number of rolls used for the printing process was analyzed in the same way. Results are presented in the below table (Table 6) and there are significant differences between variables, thus the P-value is p=0.00.

ANOVA Kruskal-Wallis rank; Wastage [m] (Data) Independent Variable (grouping): Number of Rolls [pcs] Kruskal-Wallis test: H (8, N = 123) = 65,22316 p = 0,0000				
Number of Rolls	**Code**	**N**	**Sum**	**Average**
1	1	22	508.500	23.1136
2	2	55	3001.500	54.5727
3	3	17	1245.000	73.2353
4	4	10	909.000	90.9000
5	5	10	900.000	90.0000
6	6	3	354.000	118.0000
7	7	4	466.000	116.5000
8	8	1	119.000	119.0000
14	14	1	123.000	123.0000

Table 6. *ANOVA Kruskal-Wallis rank of Wastage vs Number of Rolls*

In the next step we can see the results of the P-value for multiple comparisons.

P value for multiple (double) comparisons: Wastage [m] (Data) Independent variable (grouping): Number of Rolls [pcs] Kruskal-Wallis test: H (8, N = 123) = 65,22316 p = 0,0000									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**14**
1		0.016875	0.000482	0.000022	0.000031	0.000550	0.000052	0.306971	0.221053
2	0.016875		1.000000	0.109305	0.138418	0.096950	0.028650	1.000000	1.000000
3	0.000482	1.000000		1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
4	0.000022	0.109305	1.000000		1.000000	1.000000	1.000000	1.000000	1.000000
5	0.000031	0.138418	1.000000	1.000000		1.000000	1.000000	1.000000	1.000000
6	0.000550	0.096950	1.000000	1.000000	1.000000		1.000000	1.000000	1.000000
7	0.000052	0.028650	1.000000	1.000000	1.000000	1.000000		1.000000	1.000000
8	0.306971	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000		1.000000
14	0.221053	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	

Table 7. *P value for multiple comparisons: Wastage vs Number of Rolls*

In the above table (Table 7) we can see the result of the P-value for multiple comparisons analysis. There are many statistically significant differences between the number of rolls used for the printing process, which proves that the number of rolls used for the printing process has a big impact on final substrate wastage.

But in this case, it will be more telling to analyze the chart with basic statistics (Figure 7). It's clearly visible that by using more rolls for the work orders, substrate wastage will be increasing. Apart from the fact that in the case under study there were most work orders where two rolls were used, analysis of the median and quartiles clearly shows that the number of rolls used for an order has the greatest impact on the substrate wastage.

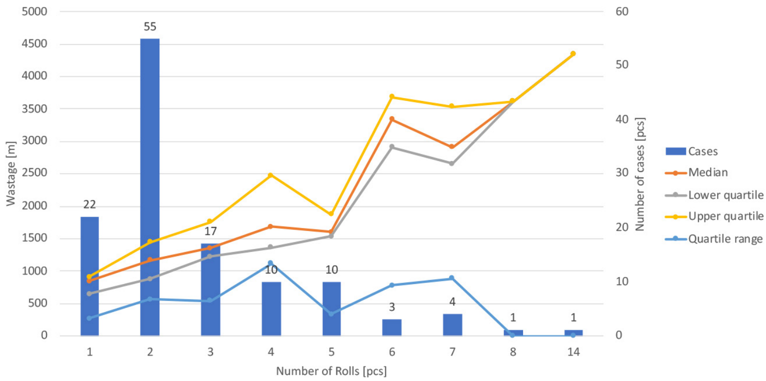


Figure 7. *Basic statistics for the Number of Rolls*

Next analysis was conducted to research if the number of colors/units has a statistically significant impact on substrate wastage.

ANOVA Kruskal-Wallis rank; Wastage [m] (Data) Independent Variable (grouping): Colors/Units [pcs] Kruskal-Wallis test: H (9, N = 123) = 6,931301 p = 0,6443				
Number of Colors/Units	**Code**	**N**	**Sum**	**Average**
0	0	1	67.000	67.00000
2	2	3	180.000	60.00000
3	3	5	376.000	75.20000
4	4	3	293.000	97.66667
5	5	8	604.000	75.50000
6	6	2	64.000	32.00000
7	7	7	449.000	64.14286
8	8	48	2787.000	58.06250
9	9	40	2445.000	61.12500
10	10	6	361.000	60.16667

Table 8. *ANOVA Kruskal-Wallis rank of Wastage vs Number of Colors/Units*

In this example, the P-value has reached the value of p=0.64 and according to the hypothesis tests rules, there are no statistically significant differences between those two variables, which are substrate wastage and number of colors/units. That proves that the numbers of colors or units, or separation have no impact on wastage. It's important to add at this part that data for this research was gathered from the printing company which has developed an ink management system for many years and all color inks were prepared by the specialized ink mixing station. That could be the answer that several colors have no impact on substrate wastage, which is obvious that more colors can be more difficult to properly print for the printer. Currently, the P value for multiple comparisons analysis has no sense because there are no statistically significant differences between the variables.

Also verifying the basic statistics graph (Figure 8) we can clearly see that there are no significant differences between the number of colors – especially looking at work orders with 7, 8, 9, or 10 colors.

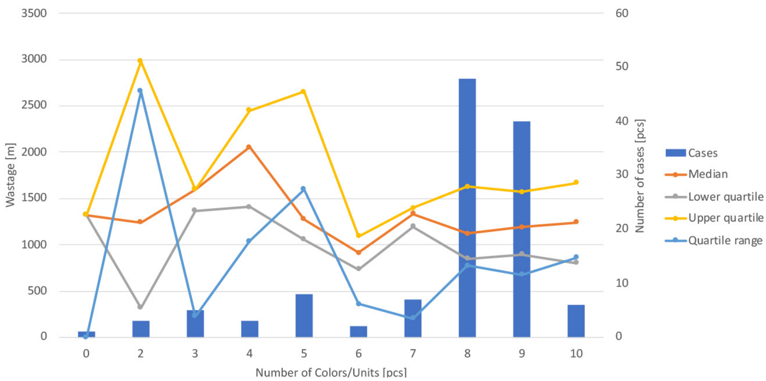


Figure 8. *Basic statistics for the Number of Colors*

Order type was the last type of analysis performed to check any differences between variables. As we know, for this research data has only two types of order type: New and Repeatable. Below (Table 9) are shown results of this analysis.

ANOVA Kruskal-Wallis rank; Wastage [m] (Data) Independent Variable (grouping): Order type Kruskal-Wallis test: H (1, N = 123) = 5,977149 p = 0,0145				
Order type	**Code**	**N**	**Sum**	**Average**
Repeatable	1	98	5687.000	58.03061
New	2	25	1939.000	77.56000

Table 9. *ANOVA Kruskal-Wallis rank of Wastage vs Order type*

The P-value shows p = 0.01 which meant that there are statistically significant differences between order type and substrate wastage. As we know that this analysis has only two variables for the Order type (New and Repeatable), there is no reason to perform the P value for multiple comparisons analysis.

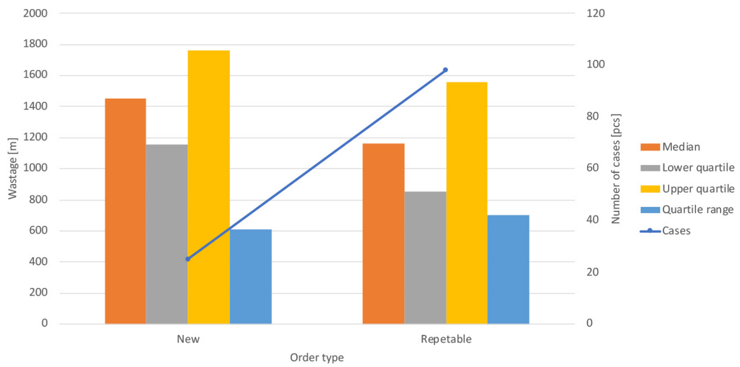


Figure 9. *Basic statistics for Order type*

Basic statistics show that the median is higher for the new order type.

In the last step, the model of the Artificial Neural Network was developed by changing the topology of the Neural Network to achieve the best prediction.

A standard sampling method was used, i.e., the number of random samples: training - 70%, testing, and validation - 15% each. The proposed settings regarding the number of neurons hidden between 3 - 11 were also selected, and the type of network - MLP network, i.e., a multi-layer perceptron (MLP). All possible activation functions were used for both hidden and output neurons. Finally, a network named MLP 32-4-1 was used for further research. MLP stands for multi-layer perceptron, 32 means that this network has 32 neurons in the input layer (but only seven input variables), 4 neurons in the hidden layer, and 1 neuron in the output layer – which is the substrate wastage. Prediction quality was calculated on different groups of data:

- Learning 0.93
- Testing 0.92
- Validation 0.92

The BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm found this neural network at the 14th iteration, the error function was SOS (Sum Of Squares), the exponential activation function was used for hidden layer neurons, and Tanh for the output layer.

Figure 10. *Prediction quality of developed Artificial Neural Network MLP 32-4-1*

The above graph (Figure 10) shows the correlation between predicted substrate wastage and substrate wastage from the historical data. All points on the graph are close to the regression curve, which proves that the developed model represents good quality – as we know it reaches a 0.93 correlation coefficient.

Conclusions

The developed model using the Data Miner Recipes found dependencies in the process, showing very good results of prediction:

- Boosted Trees: 0.91
- Artificial Neural Network: 0.89
- C&RT: 0.86

However own developed model of artificial neural network topology could bring even better prediction quality, like in the example of the MLP 32-4-1 model, which reached 0.93 for the learning quality.

What is more important when working with data and developing such mathematical models is properly learning the raw data, because it is essential for proper prediction models. The poor quality of data gathered could vary in further analyses and prediction model development.

Choosing vital variables is essential when preparing the prediction models.
Developed prediction models could be useful for the proper prediction of substrate wastage before planning the production of work orders.

References

Flores J., A., *Mathematics Research Developments: Focus on Artificial Neural Networks*, Nova Science Publishers, Inc. New York 2011

Heat Shrink Sleeve Label Technical Manual & Test Methods, Technical Publication, AWA Alexander Watson Associates, Amsterdam 2014

Kipphan H., *Handbook of print media. Technologies and Production Methods,* Springer Germany 2001

Kit L. Y., *The Wiley Encyclopedia of Packaging Technology, 3rd Edition*, John Wiley and Sons 2009

Krystosiak K., *Application of Data Mining Tools in Shrink Sleeve Labels Converting Process*, Monography "Studies in Systems, Decision & Control. Soft Modelling in Industrial Manufacturing", vol. 183, p. 89-106, Springer Nature Switzerland AG 2019

Levenson H., Parsons J., *Introduction to Graphic Communication*, IntuIdeas LLC, 2nd edition, 2018

Negnevitsky M., *Artificial Intelligence. A Guide to Intelligent Systems*, Addison-Wesley, Edinburg 2005

Nisbet R., Elder J., *Miner G., Handbook of Statistical Analysis & Data Mining Applications*, Academic Press/Elsevier 2009

Siddique N., *Adeli H., Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks Intelligent Systems and Applications*, John Wiley & Sons, Inc., 2013

Tadeusiewicz R., *Neural Networks*, The Academic Publishing House, Warsaw 1993

Vapnik V., *The Nature of Statistical Learning Theory*, II, Springer-Verlag, New York 2000