INTELLIGENT ANALYSIS OF MANUFACTURING DATA

Galina Setlak, Monika Piróg-Mazur, Łukasz Paśko

Abstract: *The article discusses the data analysis conducted for a company of the glass industry. The first part of the paper introduces the examined production process and presents the descriptive statistics used to analyze the manufacturing data. The dataset was collected from measuring points in the process of quality control of products in a given period. In the second part, another dataset is presented. This data was used as a training set for artificial neural networks. Finally, the paper describes the results of research on the possible use of the neural networks for the automatic classification of defects in finished products.*

Keywords: *artificial neural networks, production process, descriptive statistics.*

ACM Classification Keywords: *I. Computing Methodologies; I.2 Artificial Intelligence; I.2.6 Learning; Connectionism and neural nets. J. Computer Applications; J.2 Physical Sciences and Engineering; Engineering.*

Introduction

Modern companies operating in a market economy are facing the need to improve product quality, increase productivity and reduce costs, as well as to maximize profits. Apart from physical resources, information have become one of the most important assets of every organization. Therefore every firm that would like to be highly competitive collects data on various aspects of processes it conducts. Because large amount of data collected, in order to improve the quality and reduce costs, the proper methods and tools of data analysis have to be used. Very important task is the extraction of knowledge hidden in the large data sets for supporting business activities. It justifies the efforts to develop new approaches in this area and the use of modern advanced methods and tools of artificial intelligence.

Forecasting plays an important role in the functioning of companies and constitutes an integral part of the production process. It reduces uncertainty and helps to eliminate losses through the decision making process improvements. In the forecasting the mathematical and statistical methods, non-mathematical methods, or artificial neural networks can be used, what facilitates the work, reduces the time-to-market and lowers costs.

Statistical analysis of the data allows the formulation of generalizations based on the obtained results. It also allows the prediction of the events evolution, that is to build forecasts. What is more, it provides tools to organize data about the phenomena, and thus the construction of the overall picture. One of the methods category, one can draw conclusions about the entire population based on the random sample is descriptive statistics [Aczel, 2000]. Descriptive statistics deals with problems of statistical surveys, methods of statistical observation as well as methods of preparation and presentation of statistical properties of the total data set. Using descriptive statistics the biggest set of defects occurring on the production line was analyzed. It is called the "SWA" defect, that gathers defects of deposits and lines on bottles.

The production process in the Glassworks defines the tasks of converting raw materials (blank) into finished product, according to the requirements specified in the project. The development of technological process is a very important stage in the production planning and preparation. It is, however, very difficult automation process due to the large number of engineers involved and the know-how used in the design process.

Technologists' experience significantly affects the technological process and its costs. The technological process together with the auxiliary operations (movements of the material) constitute a manufacturing process in which the final product is obtained. In the manufacturing of glass there are 9 main activities related to the transformation of raw materials into finished products (intended for an external consignee). With such great complexity of production process, the occurrence of defects is possible and therefore it's very important for the company to optimize the whole production process [Dejniak et al., 2011], [Piróg-Mazur et al., 2011].

Major production lines, which include dozens of machines linked together, include the measurement points for quality control purposes. Currently in the Glassworks, data from measuring points are collected using specialized software by the PIC - Production Information Computer. The examination parameters set for the individual elements of bottles consist of: the neck characteristics, the thickness of the wall, contoured body and the bottom. The information generated by the PIC software include: a summary of the losses on the production line, a summary statement for the entire steel mill, a summary of the waste equipment for FP (cold end), the losses in the selected line details, recoil defect data (expressed in percentage terms), kickback data defects in the piece, stationary devices report, a summary of the results of all lines and changing the production line to another [Piróg-Mazur et al., 2011].

The article presents the descriptive statistics of the selected defects and test results concerning the possible use of artificial neural networks for the automatic classification of defects in finished products and their causes. The aim of the study was to create a neural classifier, which task is to classify the defects of the products into three classes, where every class expresses the difficulty degree of the defect elimination (low, medium, high).

Descriptive statistics of data

The purpose of descriptive statistics methods is to summarize a set of data and draw some basic conclusions and make generalizations. Descriptive statistics are used as the first and fundamental step in the analysis of the collected data. By entering the analysis of the process we are faced with the choice of an appropriate sample size - a group of representatives. For this reason some statistical concepts can refer to both – the entire population, and the sample (these are called empirical values) [Luszniewicz et al., 2003].

The sample size is 93. It contains data collected during one month, broken down into three shifts. Data describe bottle recoil with characteristics of individual defects (Table 1). Data have been gathered from the measuring points of the production line. In the column headings there are names of defects identified on the production line.

The minimum number of occurrence of defects "SWA" was 952 pieces, and maximally there was made 4241 pieces of recoil in the form of cullet. The arithmetic mean is one of the most intuitive assessment measures of the population. The average of observation set is the sum of all values divided by the number of elements in this set. The average recoil bottles of "SWA" defect was 2168 pieces. The median value in the statistics is a feature value in an ordered set, above and below which there is an equal number of observations values. The median is called the 2-quantile, and the second quartile. The value of median is 2044, what means that half of the observed changes produced no less than 2044 pieces of bottles with "SWA" defect.

Standard deviation is a basic measure of the variability of observed results. It provides information on results of the "change", i.e. whether the spread of results around the mean is small or big. The coefficient of variation shows how strong is the diversity of the data. The standard deviation value is 699, and the coefficient of variation is 32.23 (32%), what shows the moderate variation of bottles production quality at each change. Kurtosis is a measure of the results concentration. Kurtosis informs us about how our observations results are concentrated around the average. This measure tells us how many of our observations results are close to the average. Kurtosis value for the sample is 0.42.

Table 1. Details of the data set for one production line.

Data: Arkusz1* (18v by 93c)

	1 Number of tests	2 CID	3 SWA	4 BHA	5 FTA_SSG3	6 Blown Collar	7 Cracked Ring	8 Cracking under the Head	9 Micro-vertical cracks in the Head	10 Micro-cracks on the head surface	11 Deviation from the axis	12 Blown Body	13 Oval body. Box label	14 Horizontal cracks in the neck	15 Cracks in the bottom	16 Cracking the bottom/the body	17 Thin upper	18 Thin bottom
	1	1097	2837	897	1029	28	256	1299	591	818	67	0	46	379	94	22	264	489
	2	1507	3037	1150	1492	38	99	624	291	899	104	0	113	241	51	13	219	601
	3	942	3439	1147	1089	13	109	806	308	940	18	0	53	169	48	1	164	620
	4	158	2871	899	1970	24	110	621	521	1386	52	0	60	163	127	13	180	473
	5	2159	2410	1125	2095	33	216	1622	241	973	39	0	56	103	32	22	253	538
	6	1095	2158	931	1128	21	272	969	342	1099	46	5	116	291	71	46	336	216
	7	1195	2654	1879	1628	42	639	1722	346	1147	162	0	166	123	200	0	355	236
	8	1006	2891	1501	2215	26	422	968	308	1138	35	0	398	156	42	0	166	190
	9	2875	4058	1489	1608	16	179	722	163	494	23	0	199	107	32	0	215	282
	10	0	3290	1108	1507	28	244	842	228	687	33	42	119	38	0	0	92	146
	11	0	3791	1318	1109	16	194	739	250	704	87	68	111	35	14	0	202	163
	12	0	3021	1629	1488	21	309	647	273	788	195	45	92	31	0	0	466	195
	13	0	3550	1768	1945	27	321	758	359	985	111	0	12	266	37	0	313	142
	14	0	4241	1492	2389	45	154	855	503	1106	42	0	32	196	56	0	365	305
	15	1160	3484	1440	3000	28	90	631	471	1256	67	1	71	265	82	0	484	162
	16	911	2122	970	1348	21	78	538	265	671	24	0	21	90	68	0	345	106
	17	610	2964	1036	1641	16	57	473	245	836	47	0	22	213	79	0	119	135
	18	0	2300	952	887	33	165	305	230	587	23	0	40	290	20	0	285	254
	19	0	2625	880	1156	18	84	681	298	766	37	0	51	90	45	0	211	225
	20	1385	2098	829	923	24	165	557	280	832	17	0	31	53	34	0	130	176
	21	533	1980	794	933	21	103	312	267	747	24	0	57	178	32	0	78	323
	22	0	2753	866	739	8	77	836	220	637	13	0	64	195	18	0	117	318
	23	2051	2058	714	1313	27	214	1029	267	706	30	0	50	194	22	0	151	189
	24	70	2509	792	1409	15	160	614	245	712	20	0	79	256	26	0	140	147

Below there has been presented a line graph (Figure 1) which is the most common type of statistical graphs. Data are presented with a line, usually broken one. Each point is connected to a line from the first to the last value. This type of chart is used most often for the presentation of the data collected in a given period of time. The x-axis presents a fixed unit of time, while y-axis shows the selected variable – “SWA” defect. With this form, we can determine a variability in the "SWA" in the given period of time. The trend line is always associated with a series of data, but it does not reflect the data in series. The trend line is used to illustrate trends in existing data or to predict future data. An exponential trend line curve is a line used in cases where the values are rising or falling with constantly increasing speed.

One of the basic concepts used in the statistical analysis is stationary variable. Intuitively stationary variable is a variable which properties do not change over time. From the graph in Figure 1, we can see that the graph is non-stationary.

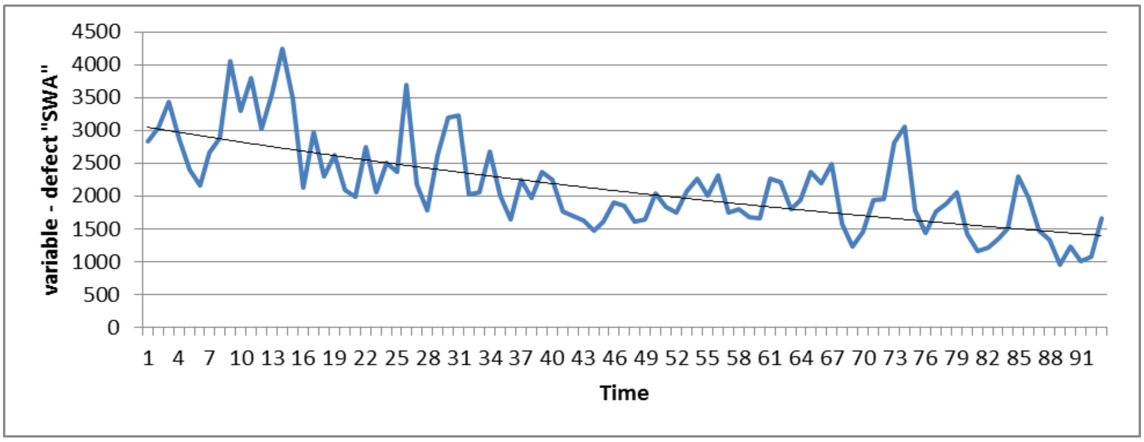

Fig. 1. Line graph of “SWA” defects with the exponential trend line. Source: own work

The histograms (Fig. 2) are the graphic representations of the quantity distribution of the "SWA" variable on which the columns (bars) are plotted over the class intervals, and the height of the column is proportional to the size of classes. The chart facilitates the evaluation of empirical normality (description of the values taken by the characteristic statistical sample using their frequency), because the histogram is applied to fit the curve of the normal distribution. It also allows qualitative evaluation of various aspects of the distribution [Internetowy podręcznik statystyki]. The distribution in this case is unimodal (mode equals 1 - it has one peak), and the frequency of mode equals 2.

The chart of normality for the "SWA" variable from the analyzed data set is given below (Fig. 2). If the points lie close to the straight line graph, and they are uniformly distributed at both sides (e.g., alternately), the data come from a normal distribution.

The box-and-whisker charts (Fig. 2) are developed based on the descriptive statistics, so their use is limited to the numerical characteristics. Most frequently developed charts are those containing median, quartiles, minimum, and maximum. The length of the rectangle represents the middle 50% of observations values. The box is separated by a vertical line (or dot), which determines the value of the median. It divides the quartered section into two areas in which there is 25% of the observations. Whiskers combine a box with the largest and smallest value of the test variable. The first section is 25% of the observations with values below the lower quartile and the second with 25% of the observations with values greater than the upper quartile. The position of the box with respect to the number line shows the position of the distribution. The dot indicates the central tendency of the distribution. The length of the rectangle and the entire chart shows the dispersion of the characteristics in the data set. The proportions on both sides of the vertical line defined by the median value indicate the type of skewness of characteristics distribution (whether it is right- or left sided) [Luszniewicz et al., 2003].

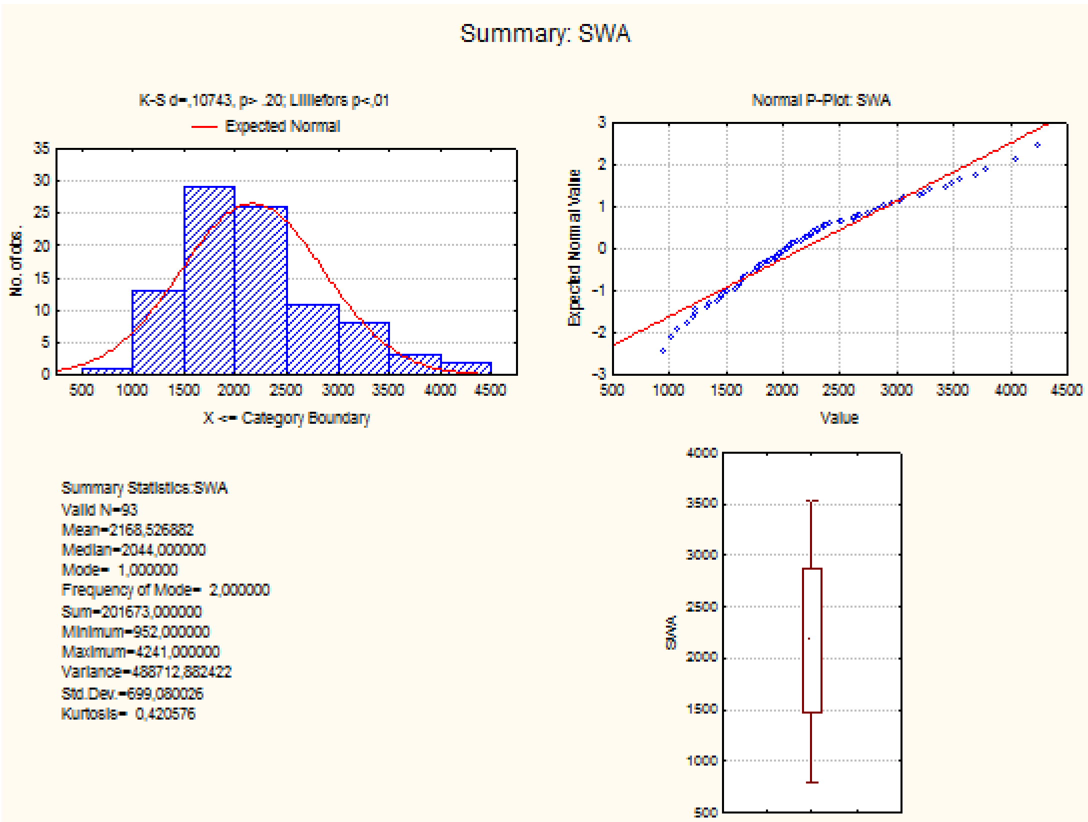

Fig. 2. Descriptive statistics of "SWA" defects. Source: own work

In the histogram can be seen numbers of observations (Y- axis, vertical), in the interval (x-axis, horizontal). The size of the compartments is equal. That is, the first pistil describes a number of changes to produce bottles of defect "SWA" from 500 to 1,000. Width of the ranges in the histogram is the same. The difference is the amount of columns (number of observations). In the analyzed example can be seen that most changes have occurred which produced defective bottle in the number of units from 1500 to 2000 and from 2000 to 2500. In chart normality there are shown the minor deviations in the case of points at the top and at the bottom - these points lie further from the straight line than the other points. However, the deviation is so small that the Shapiro-Wilk test does not indicate deviations from a normal distribution. The analyzed box-and-whisker graph shows that the distribution is symmetrical, without outliers.

These statistical graphs are the visualizations of previously conducted statistical analyzes (for example, group data or descriptive statistics).

In the next sections, artificial neural networks are described. The networks have been used to classify defects in finished products in terms of the difficulty degree of the defect elimination.

Artificial neural networks

Artificial neural networks are often used as classifiers, what has been presented in the following papers: [Adamczak, 2001], [Jang et al., 1997], [Moon et al., 1998], [Osowski, 2013], [Setlak, 2000], [Setlak, 2004], [Setlak, Paśko, 2012], [Stąpor, 2011], [Zieliński, 2000]. Neural networks are considered as one of the data mining techniques. The objective of data mining is to find some hidden patterns and relationships that occur in the large sets of data. Discovered patterns can provide a hitherto unknown knowledge, crucial from the point of view of analyzed production process. Data mining is also called intelligent data analysis because it can use the methods and techniques of artificial intelligence [Hand et al., 2005], [Larose, 2006].

In the experiment there have been used only those networks which are fully connected (each neuron of the preceding layer is connected to all the neurons of the next layer). The analyzed networks included one-way connections, without feedback. To train the networks, the supervised learning method has been used.

The classes of artificial neural networks may differ with regard to definition of their activation functions $f(net)$. During the experiment, the following types of neural networks have been examined: linear networks, multilayer perceptrons (MLP) with three kinds of activation function, networks with radial basis functions (RBF), and probabilistic neural networks (PNN).

Linear networks

Linear networks have two layers of neurons: the input and output. The neurons in the input layer serve only to provide input data to the output neurons, without performing any transformations at the same time. Each neuron contained in the output layer, designated as out, comprises a linear activation function, which for the k-th output neuron is described with formulae:

$$f_k^{out}\left(net_k^{out}\right)=net_k^{out}, \tag{1}$$

where net is the sum of neuron's weighted inputs. The function described by the formulae (1) is sometimes called an identity function. Networks only formed from linear neurons have limited capabilities and can be used to solve the simplest problems [Witkowska, 2002].

MLP networks

Multilayer perceptrons (MLP) are the most universal networks commonly used in the classification problems. Such network has at least three layers: input, hidden and output. Like the linear-type network, the input layer does not perform the computational functions. Perceptron can have more than one hidden layer, but in practice it is assumed that most of the problems can be solved using perceptron with one or two hidden layers.

In the performed study it has been analyzed the MLP network, which hidden and output neurons have been equipped with the following activation functions:

- unipolar threshold function:

$$f(net) = \begin{cases} 1 \text{ where } net \geq 0 \\ 0 \text{ where } net < 0 \end{cases}, \tag{2}$$

- logistic function (or unipolar sigmoidal function):

$$f(net) = \frac{1}{1 + \exp(-\beta\, net)}, \tag{3}$$

- hyperbolic tangent function (or bipolar sigmoidal function):

$$f(net) = \frac{\exp(\beta\, net) - \exp(-\beta\, net)}{\exp(\beta\, net) + \exp(-\beta\, net)}. \tag{4}$$

In the formulas (3) and (4) β is a function parameter - the higher the value is the graph of the function is steeper, close to the threshold function [Nałęcz, 2000], [Osowski, 2013], [Tadeusiewicz, 1993].

RBF networks

Radial basis function networks (RBF) are used especially for nonlinear approximation of numerical variables, but they can be also used in classification tasks, where they describe the probability density function of the input variables. RBF networks typically have three layers: input, hidden, and output. Input neurons transfer the signals to hidden neurons that are equipped with radial basis functions. The radial functions are a class of functions which values are decreasing or increasing monotonically with distance from the center of the neuron. Therefore, in contrast to multilayer perceptrons, the RBF's hidden neurons are arranged as centers in the data space. The most frequently used radial function is the Gaussian function. The output layer implements a linear summation of signals from the hidden layer [Nałęcz, 2000].

PNN networks

Probabilistic neural networks (PNN) are used as classifiers. During a training process, PNN networks learn to estimate the probability density function, which is represented by the training data. In the base case, this network contains an input, a radial, and an output layer. The input neurons do not transform the input data. In the radial layer, the number of neurons corresponds to the number of training cases. Each radial neuron is equipped with the Gaussian function centered at the data space over the corresponding training case. The task of the output neurons is to sum the signals coming from the radial layer. After normalization, the calculated sums estimate the probability of belonging of the training case to the output classes [Żurada et al., 1996].

The experiment and the results

The second analyzed data set contained 409 cases. Each case is described using four input variables and one output variable that are presented in Table 2.

Due to the qualitative character of the input variables, each of them has been re-coded by the *one of n* method. Such coding is based on the mapping of a qualitative variable having *n* values, using the *n* input neurons. Each neuron corresponds to a different value of the variable. In the case of occurrence a given value, a corresponding neuron transmits to the network the value 1, and the remaining neurons take the values 0.

The number of neurons in the input layer is determined by the number of variables that are taken into account by the network. The maximum number of neurons occurs when the network uses all four input variables. Taking into account the above-described encoding *one of n*, the number of input neurons in this case is 742.

Table 2. Variables in the second data set. Source: own work

Designation	Description	Type
VAR1	the type of the defect	input
VAR2	the group of defects	input
VAR3	the main cause of the defect	input
VAR4	the way to eliminate the defect	input
VAR5	the difficulty degree of the defect elimination	output

Output layer determines the final outcome of the network; hence the number of neurons in this layer is also closely conditioned by the nature of the problem under consideration. In this experiment, the output variable is the difficulty degree of the defect elimination. The difficulty degree can have one of the following three values:

- low (designation **L**) – the defect is easy to elimination, production line downtime is minimal,
- medium (**M**) – the elimination of the defect requires a longer stop of production process and a greater usage of company's resources,
- high (**H**) – the defect is very difficult to eliminate, the removal of the defect is the most time-consuming.

The difficulty degrees are treated as classes during the classification task. Therefore, each of the examined neural networks includes three output neurons corresponding to three different classes. The object is classified to the class for which the corresponding output neuron reaches the highest value of the activation function.

During the experiment, the following parameters of network architecture were determined:

- the number of hidden layers of networks (none, one or two),
- number of neurons in the hidden layers,
- activation functions of hidden and output neurons.

The above parameters were determined in two ways:

- by using the *Intelligent Problem Solver* options (*IPS*), used for their automatic determination,
- independently modifying parameters until satisfactory results of the network were obtained (only in the case of MLP networks).

IPS option, in addition to creating network architecture, also allows to automatically training the network, and adopting parameters of the learning process selected by the software. In contrast, self-design of MLP network required the selection of an appropriate algorithm and learning parameters that were determined experimentally based on the numerous repetitions of the training process:

- learning method: backpropagation algorithm,

- learning coefficient: 0.1,
- momentum: 0.3,
- number of epochs: 100.

STATISTICA Neural Networks software generated, using the *IPS* option, seven linear types of network, some characteristics of which have been shown in Table 3. Each of them was trained by pseudo-inverse method.

By analyzing the input variables used by the program it can be observed that the most important inputs for the linear network are VAR3 and VAR4. "Knowledge" only of VAR3 variable can correctly classify over 98% of the cases, which shows the LIN1/3 network. The use of only VAR4 results in almost 96% of correct classifications. Using both of these variables in the network LIN2 does not improve the results, even increases the number of misclassifications. To significantly reduce the number of misclassifications, variables VAR1 and VAR2 should be taken into account, what is shown by the results of LIN3 and LIN4 networks.

In addition, Table 3 illustrates another significant correlation. The percentage of correct classifications in LIN3 and LIN4 is very similar, despite the fact that these networks take into consideration the different number of input variables. Comparing these networks it can be concluded that if the variables VAR1, VAR3, VAR4 are available for the linear neural network, the additional use of VAR2 does not significantly affect the outcome of the classification task.

Table 3. Selected parameters of the linear networks. Source: own work

Designation	**The input variables**	**Percentage of correct classifications**	**Number of misclassifications for each class**			
			L	**M**	**H**	**total**
LIN3	VAR1, VAR3, VAR4	99,51%	0	2	0	2
LIN4	VAR1, VAR2, VAR3, VAR4	99,27%	0	3	0	3
LIN2	VAR3, VAR4	94,62%	16	5	1	22
LIN1/1	VAR1	49,14%	41	88	79	208
LIN1/2	VAR2	42,54%	35	106	94	235
LIN1/3	VAR3	98,04%	0	4	4	8
LIN1/4	VAR4	95,60%	1	9	8	18

The second group of examined networks is multilayer perceptrons with one hidden layer. Networks created using *IPS* options differ not only in the number of inputs, but also in the number of hidden neurons. Features of selected perceptrons have been shown in Table 4. The use of perceptrons with one hidden layer allowed increasing the percentage of correct classifications. The best of them make one misclassification. However, as in the case of the linear networks, the largest number of errors concerns the M class.

Looking at the first four networks it can be seen that the number of misclassifications for the perceptrons with three inputs was the same as in the case of the networks with all input variables. This confirms earlier observations noted on the example of the linear networks. Besides, when comparing the MLP 3-42 (three input neurons, 42 hidden neurons) and MLP 3-26 it can be seen that increasing the number of hidden neurons does not change the results of the classification. A similar relationship exists in the network with four inputs: MLP 4-42 and 4-28.

Table 4. Selected parameters of the MLP networks with one hidden layer. Source: own work

Designation	The input variables	Percentage of correct classifications	Number of misclassifications for each class			
			L	M	H	total
MLP 3-42	VAR1, VAR3, VAR4	99,76%	0	1	0	1
MLP 4-42	VAR1, VAR2, VAR3, VAR4	99,76%	0	1	0	1
MLP 3-26	VAR1, VAR3, VAR4	99,76%	0	1	0	1
MLP 4-28	VAR1, VAR2, VAR3, VAR4	99,76%	0	1	0	1
MLP 2-14	VAR3, VAR4	99,27%	1	2	0	3
MLP 1-29	VAR3	98,04%	3	3	2	8
MLP 1-8	VAR1	49,14%	66	80	62	208

Considering the importance of the variables, the program most often selects VAR3 input. All perceptrons (except the last) take into account this variable. Looking at the networks, which use a one variable, it can be seen that the results of the perceptrons are identical as the linear networks. Examples of this are MLP 1-29 and LIN1/3 giving 8 errors, or MLP 1-8 and LIN1/1 with 208 errors. Although the total number of misclassifications is the same, the compared pairs of networks differ in the distribution of errors in the classes.

The third group of the networks consists of perceptrons with two hidden layers of neurons. Characteristics of these perceptrons are described in Table 5. When analyzing the percentage of correct classification it can be seen that the addition of the second hidden layer did not result in reduction in the number of classification errors. Each network uses at least the three inputs committed one error. This error cannot be eliminated by the extension of network architecture. Network MLP 2-6-9 (2 input neurons, 6 in the first hidden layer, 9 in the second hidden layer) can be compared to MLP 2-14. Both perceptrons have variables VAR3 and VAR4. Despite the differences in their architecture, the use of these classifiers gives 3 incorrect classifications.

Comparing the outcomes of MLP 1-4-14 with the results of previous networks using only the VAR1 (MLP 1-8, LIN1/1), it can be concluded that adding more layers and hidden neurons does also not enable to achieve better results. Using only the variable VAR1, results in efficiency of about 50% correct classification. Similarly in the case of networks, which use the VAR3 (MLP 1-19-20, MLP 1-29, LIN1/3) – each of them commits the same number of misclassifications. As is shown, the networks with two hidden layers (Table 5) have identical percentage of correct classifications as the corresponding networks with one hidden layer (Table 4). The difference is only the distribution of errors in each class of the output variable.

Table 5. Selected parameters of the MLP networks with two hidden layers. Source: own work

Designation	The input variables	Percentage of correct classifications	Number of misclassifications for each class			
			L	M	H	total
MLP 3-25-20	VAR1, VAR3, VAR4	99,76%	0	1	0	1
MLP 4-43-36	VAR1, VAR2, VAR3, VAR4	99,76%	0	1	0	1
MLP 3-11-14	VAR1, VAR3, VAR4	99,76%	1	0	0	1
MLP 4-21-21	VAR1, VAR2, VAR3, VAR4	99,76%	1	0	0	1
MLP 2-6-9	VAR3, VAR4	99,27%	0	3	0	3
MLP 1-19-20	VAR3	98,04%	2	4	2	8
MLP 1-4-14	VAR1	49,14%	57	75	76	208

All MLP networks generated by the *Intelligent Problem Solver* contained neurons with logistic activation function. To compare the effects of other activation function, there has been used the opportunity of independent design and training of the networks.

Firstly, 12 perceptrons have been created, which characteristics and results of classification are summarized in Table 6. For analyzing four different network architecture have been selected that varied in the number of input variables, hidden layers, and hidden neurons. Next, in each of these architectures three activation functions have been used: logistic, hyperbolic tangent and threshold. These functions are applied in all hidden and output neurons.

Table 6. Selected parameters of the self-created MLP networks. Source: own work

Designation	Number of neurons in hidden layers		The input variables	The activation function	Number of misclassifications
	first	second			
MLP 3-42	42	-	VAR1, VAR3, VAR4	Logistic	1
MLP 3-42	42	-	VAR1, VAR3, VAR4	Hyperbolic	1
MLP 3-42	42	-	VAR1, VAR3, VAR4	Step	7
MLP 4-42	42	-	VAR1, VAR2, VAR3, VAR4	Logistic	1
MLP 4-42	42	-	VAR1, VAR2, VAR3, VAR4	Hyperbolic	1
MLP 4-42	42	-	VAR1, VAR2, VAR3, VAR4	Step	205
MLP 3-25-20	25	20	VAR1, VAR3, VAR4	Logistic	1
MLP 3-25-20	25	20	VAR1, VAR3, VAR4	Hyperbolic	1
MLP 3-25-20	25	20	VAR1, VAR3, VAR4	Step	272
MLP 4-43-36	43	36	VAR1, VAR2, VAR3, VAR4	Logistic	1
MLP 4-43-36	43	36	VAR1, VAR2, VAR3, VAR4	Hyperbolic	1
MLP 4-43-36	43	36	VAR1, VAR2, VAR3, VAR4	Step	272

Analyzing the results expressed with number of misclassifications it can be concluded that the used activation function was primarily responsible for the results. All networks using the logistic function and hyperbolic tangent function committed one error. The network with threshold activation function behaved differently. Perceptron MLP 3-42 incorrectly classified 7 cases, however, expanding the network architecture, it also increased the number of errors, which reached up to 272 false classifications of networks with two hidden layers (MLP 3-25-20).

Comparing the number of inputs, which were taken into consideration during learning process, a different relationship can be observed. All perceptrons classify cases identically, regardless of whether they used all of inputs or only three of them. The exception was the mentioned perceptrons with the threshold activation function. Using the three input variables results in 7 classification errors (MLP 3-42), while taking into account all inputs, the number of errors increases to 205 (MLP 4-42).

During training the networks there was also measured the duration of this process. Of course, the results of the measurements also depend on hardware capabilities, and moreover the duration of training with considered amount of data is not a significant parameter, therefore it is not taken into consideration in the analysis. Before starting learning it was assumed that the maximum number of epochs of backpropagation algorithm would amount to 100. For each perceptron algorithm ended learning before it reached that number. However, it is hard to find any relationship between the number of epochs and other network parameters considered in this study, thus the analysis of the number of epochs has also been omitted.

The last group of networks consists of RBF and PNN, whose characteristics are shown in Table 7. By using the IPS option, several RBF networks with different numbers of hidden neurons have been generated (from 68 to 393). Each of these RBF used all input variables to classify. Three selected networks are presented in Table 7.

In case of the RBF networks it can be observed the following relationship: when the number of hidden neurons increases, the number of misclassifications decreases. Such situations did not exist for MLP networks – an increase in the number of neurons did not result in a decrease in classification errors. In addition, the best of established RBF networks committed 8 errors. This result is worse in comparison with the linear and MLP networks.

Table 7. Selected parameters of the RBF and PNN networks. Source: own work

Designation	The input variables	Percentage of correct classifications	Number of misclassifications for each class			
			L	M	H	total
RBF 4-393	VAR1, VAR2, VAR3, VAR4	98,04%	0	8	0	8
RBF 4-237	VAR1, VAR2, VAR3, VAR4	86,06%	6	38	13	57
RBF 4-68	VAR1, VAR2, VAR3, VAR4	55,99%	60	62	58	180
PNN 4-409	VAR1, VAR2, VAR3, VAR4	99,76%	0	1	0	1
PNN 3-409	VAR1, VAR3, VAR4	99,76%	0	1	0	1
PNN 2-409	VAR3, VAR4	99,27%	0	3	0	3
PNN 1-409	VAR3	98,04%	0	5	3	8

In contrast, each PNN network is equipped with 409 hidden neurons, which is consistent with the number of training cases. The *IPS* option created four PNN networks with different number of inputs. The results of PNN are identical to the MLP with one or two hidden layers. The only difference is the distribution of errors in the output classes.

To better compare the behavior of the neural networks, confusion matrices have been prepared (Table 8).

Table 8. Confusion matrices of the selected networks. Source: own work

		Predicted class								
		L	M	H	L	M	H	L	M	H
Actual class	L	159	0	0	159	0	0	159	0	0
	M	2	111	0	1	112	0	8	105	0
	H	0	0	137	0	0	137	0	0	137
Network's designation		LIN3			MLP 3-42 MLP 3-25-20 PNN 4-409			RBF 4-393		

The matrices were created for all tested types of networks, taking into account the classifiers committing the least number of errors. Rows of the matrices correspond to correct classes, while the columns represent classes predicted by the classifier. The matrices have been established for five neural networks. For three of them (MLP 3-42, MLP 3-25-20, PNN 4-409), the matrices are the same. The results show that the misclassifications refer only the cases that belong to the class M. These cases are incorrectly assigned to the class L.

Conclusion

In a manufacturing company the decision-making process related to quality control constitutes an essential and crucial part of the whole production process. Therefore the data collected in a manufacturing process that can support decision-making, must be properly interpreted. At this point, the data mining techniques can be used, which are designed to extract information from the data.

The analysis of data is the starting point for daily discussions about the number of defects generated during production process, causes of the defects, methods to eliminate the defects, and the difficulty degree of the defect elimination.

In the paper a part of the data has been presented, and the descriptive statistics on sample dataset have been calculated, which is the starting point for further research. Finally, artificial neural networks have been used to classify the causes of defects in finished products in terms of the difficulty degree of the defect elimination.

The purpose of the descriptive statistics was to investigate the distribution of the features as well as to estimate the characteristics of this distribution. For the statistical analyzes, the *STATISTICA* software (module *Basic Statistics and Tables*) has been used. To perform the neural networks experiments, the *STATISTICA Neural Networks* software has been applied.

The dataset used for training the neural networks has not be extended with new cases from the future. Therefore, the study did not analyze the ability of generalization of the networks. The analysis focuses only on testing the ability to learn the dataset with qualitative variables.

The best of the established networks has made one misclassification. This error could not be avoided by changing the type of the network, the use of different activation functions, and by modifying the network's architecture.

Comparing types of neural networks used it can be concluded that the problem was solved the best by MLP and PNN networks. Trying to select the network with the lowest number of misclassification and the simplest architecture, the MLP 3-26 should be chosen. Among the examined activation functions in MLP, the logistic function and hyperbolic tangent function allowed to achieve good classification results. The threshold function is not suitable to solve the considered problem.

The obtained results can be treated as analysis of significance of the variables, which indicates the most important inputs for the classification and allows determining the optimal subset of input variables. The most significant inputs for the networks are VAR3 and VAR4 variables. However, the use of only these inputs does not give the best classification. The best results of neural networks have been achieved by taking into account all input variables as well as using only three variables: VAR1, VAR3, and VAR4. These observations are confirmed in the case of networks created with automated *IPS* option and networks generated and learned independently using experimental method. Thus, having only three mentioned inputs, the neural networks are able to classify with almost absolute certainty.

In the further research the attempts to explore knowledge from data using another data mining techniques, such as neuro-fuzzy systems, decision trees and association rules will be undertaken.

Acknowledgements

Research financed by the European Social Fund under the project entitled "Podkarpacki scholarship fund for graduate students".

Bibliography

[Adamczak, 2001] Adamczak R.: Zastosowanie sieci neuronowych do klasyfikacji danych doświadczalnych, PhD thesis, Nicolaus Copernicus University, Toruń, 2001.

[Aczel, 2000] Aczel A.D.: Statystyka w zarządzaniu, PWN, Warszawa, 2000.

[Dejniak et al., 2011] Dejniak D., Piróg-Mazur M.: Elements of forecasting and time series analysis in the process of the glass industry, 6th National Scientific Conference INFORMATION SYSTEMS IN MANAGEMENT, 2011.

[Hand et al., 2005] Hand D., Mannila H., Smyth P.: Eksploracja danych, WNT, Warszawa, 2005.

[Internetowy podręcznik statystyki] Internetowy podręcznik statystyki, www.statsoft.pl/textbook/, © Copyright StatSoft, Inc., 1984-2011.

[Larose, 2006] Larose D.T.: Odkrywanie wiedzy z danych, PWN, Warszawa, 2006.

[Luszniewicz et al., 2003] Luszniewicz A., Słaby T.: Statystyka. Teoria i zastosowania, Wydawnictwo C.H.Beck, Warszawa, 2003.

[Moon et al., 1998] Moon Y.B., Divers C.K., Kim H.-J.: AEWS: An Integrated Knowledge-based System with Neural Network for Reliability Prediction. Computers in Industry, Vol.35, No. 2, 1998, pp. 312-344.

[Nałęcz, 2000] Nałęcz T. (Ed.): Biocybernetyka i Inżynieria Biomedyczna 2000, Tom 6: Sieci neuronowe, Warszawa, 2000.

[Osowski, 2013] Osowski S.: Metody i narzędzia eksploracji danych, Wyd. BTC, Legionowo, 2013.

[Piróg-Mazur et al., 2011] Piróg-Mazur M., Setlak G.: Budowa bazy danych oraz bazy wiedzy dla przedsiębiorstwa produkcyjnego w przemyśle szklarskim, VII Krajowa Konferencja Bazy Danych: Aplikacje i Systemy BDAS'2011, Studia Informatica, Ustroń 2011.

[Piróg-Mazur et al., 2011] Piróg-Mazur M., Setlak G.: Database and Knowledge Base as Integral Part of the Intelligent Decision Support System, Created for Manufacturing Companies. In: Business and Engineering Applications of Intelligent and Information Systems, G. Setlak, K. Markov (Eds.), pp. 202-210, Rzeszów, 2011, ITHEA.

[Setlak, 2004] Setlak G.: Intelligent Decision Support System, LOGOS, Kiev, 2004, (in Rus.).

[Setlak, 2000] Setlak G.: Neural Networks in the Intelligent Information Systems of Production Control, Journal of Automation and Information Sciences, Begell House Inc. Publishers, ISSN Print: 1064-2315, Vol. 32, 2000 Is. 2, pp. 88-93.

[Setlak, Paśko, 2012] Setlak G., Paśko Ł.: Intelligent Analysis of Marketing Data. In: Artificial Intelligence Methods and Techniques for Business and Engineering Applications, Galina Setlak, Mikhail Alexandrov, Krassimir Markov (Eds.), ITHEA®, Rzeszów, Sofia, 2012, pp. 254-275.

[Stąpor, 2011] Stąpor K.: Metody klasyfikacji obiektów w wizji komputerowej, PWN, Warszawa, 2011.

[Zieliński, 2000] Zieliński J.: Inteligentne systemy w zarządzaniu – teoria i praktyka, PWN, Warszawa, 2000.

[Witkowska, 2002] Witkowska D.: Sztuczne sieci neuronowe i metody statystyczne, C.H. Beck Press, Warszawa, 2002.

[Tadeusiewicz, 1993] Tadeusiewicz R.: Sieci neuronowe, Academic Publishing House, Warszawa, 1993.

[Żurada et al., 1996] Żurada J., Barski M., Jędruch W.: Sztuczne sieci neuronowe, PWN, Warszawa, 1996.

Authors' Information

Galina Setlak – D.Sc, Ph.D., Associate Professor, Rzeszow University of Technology, Department of Computer Science, W. Pola 2 Rzeszow 35-959, Poland, and The State Professional High School, Czarnieckiego 16, Jarosław, Poland, e-mail: gsetlak@prz.edu.pl

Major Fields of Scientific Research: decision-making in intelligent manufacturing systems, knowledge and process modeling, artificial Intelligence, neural networks, fuzzy logic, evolutionary computing, *soft computing.*

Monika Piróg-Mazur, **M.Phil.**, Institute of Technical Engineering, The Bronislaw Markiewicz State School of Technology and Economics in Jaroslaw, Czarniecki Street 16, 37-500 Jaroslaw, Poland; e-mail: *m_pirog@pwste.edu.pl*

Major Fields of Scientific Research: knowledge representation, decision support system, project management.

Łukasz Paśko, **M.Phil., Eng.** – Rzeszow University of Technology, Department of Computer Science, The Faculty of Mechanical Engineering and Aeronautics, Powstancow Warszawy ave. 8, 35-959 Rzeszow, Poland; e-mail: *lukasz.pasko48@gmail.com*

Major Fields of Scientific Research: artificial intelligence, decision support systems, data mining.