A R C H I V E S O F M E T A L L U R G Y A N D M A T E R I A L S
Volume 55 2010 Issue 3

J. JAKUBSKI*, ST. M. DOBOSZ*

THE USAGE OF DATA MINING TOOLS FOR GREEN MOULDING SANDS QUALITY CONTROL

ZASTOSOWANIE NARZĘDZI ANALIZY DANYCH W PROCESIE STEROWANIA JAKOŚCIĄ SYNTETYCZNYCH MAS FORMIERSKICH

High competition on the international casting market and customers requirements concerning the casts quality forced foundries to keep introducing more advanced technological, economical and ecological solutions. IT solutions have recently become their integral part. They are usually related to such areas like information flow and logistics. Computer systems allow to define and create processes databases, update data, to follow parameters affecting the quality and use collected data to control current quality.

One of the modern methods for production optimization is using artificial neural networks (ANN). Neural networks have been very popular during last years, because ANN can use collected past data what could be very helpful in solving important industrial problems.

This article presents the comparison of two types of data mining tools for green moulding sands properties analysis, such as artificial neural networks and a naive Bayesian classifier. The tests were performed using collected data sets. An attempt to use artificial neural networks (ANN) for green moulding sands quality control is also presented.

Keywords: data mining, artificial neural networks, green moulding sands

Wysoka konkurencja na międzynarodowym rynku odlewniczym, a także wysokie wymagania klientów odnośnie do jakości odlewów, zmuszają odlewnie do wprowadzania coraz doskonalszych rozwiązań technologicznych, ale także ekonomicznych i ekologicznych. Ich integralną częścią stają się ostatnio także rozwiązania informatyczne. Te ostatnie dotyczą coraz częściej takich obszarów jak przepływy informacji i logistyka. Postęp ten dokonuje się poprzez wdrożenia rozwiązań systemowych. Systemy informatyczne powinny pozwalać na definiowanie i tworzenie baz danych o procesach, śledzić parametry wpływające na jakość, aktualizować bazy danych, a pozyskiwane informacje wykorzystywać do bieżącego sterowania jakością i do jej analiz.

Jedną z nowoczesnych metod optymalizacji produkcji są sztuczne sieci neuronowe. Sztuczne sieci neuronowe wykorzystują bazy zgromadzonych danych i mogą być bardzo pomocne w rozwiązywaniu problemów produkcyjnych.

W artykule zaprezentowano porównanie narzędzi analizy danych, takich jak sztuczne sieci neuronowe i naiwny klasyfikator Bayesa do analizy właściwości syntetycznych mas formierskich. Analizy wykonano wykorzystując zgromadzone dane doświadczalne. Przedstawiono próbę zastosowania sztucznych sieci neuronowych do sterowania jakością syntetycznych mas formierskich.

1. Introduction

A strong competition in the international foundry market as well as high requirements of clients concerning the castings quality force foundry plants to introduce more and more improved technological solutions, and also economic and ecologic ones. Information technology solutions are constituting currently their integral part. They concern such fields as information transfer and logistics. This progress takes place by means of implementations of system solutions. Information technology solutions should allow for defining and formation of data bases on processes, for tracing parameters influencing quality, updating data bases, and utilize the obtained information for the current quality management and its analysing.

Large number of data generated in casting processes is usually not directly measured and recorded, especially automatically. However, even those data which are measured and stored (e.g. chemical compositions of melts determined by spectrometric methods), are not used for the optimisation and computer aided quality management. An access to larger amounts of likelihood data requires purchasing of the appropriate measuring equipment and employing new staff [1].

* FACULTY OF FOUNDRY ENGINEERING, AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, 30-059 KRAKÓW, UL. REYMONTA 23, POLAND

However the knowledge of casting processes, their computer modelling and forecasting of the casting quality, applied statistic methods and neuron networks means already quite a lot. Unfortunately in many cases it is not enough, to assess properly the production anomalies and to prevent them, since there is a lack of the reliable, continuously updated data.

Large foundry plants are implementing expensive computer programs (e.g. SAP R/3), with an expanded structure, but they are mainly used (according to the list of modules the most often purchased by companies) in finance-economic sectors, production planning, sales, purchases, materials transfers. Widening these systems by the Quality Management module (QM), important from the point of view of a quality control, appears to be especially difficult for implementation in foundry plants, regardless of the fact that this module is offered within the SAP/R3 packet. Simpler systems adapted to the actual needs of foundry plants are easier for implementation and less expensive, than the QM module. They should better agree with the nature of the collected technological data. Then they would be useful in practice of a quality control, which should be based not only on the data recorded in actual time but also on the historical data, ordered and available easily and fast. The modern Data Mining requires the most often the preparation – from the very beginning – of such systems of the appropriate architecture of data basis.

To this end computer tool presented in papers [1,5,6] – KonMas-final system – is quite interesting since it was prepared in accordance with the given above guidelines. It simplifies decision taking on the grounds of the recorded data as well as enables their preparation as bases for analysis. This is a complementary tool in relation to the basic modules of the SAP R/3 system. It constitutes the first of the auxiliary systems for the assurance quality (AQ) and fulfilling functions expected from the QM module [1,4].

Another interesting solution is the system of analysis of casting defects presented in paper [7]. A wide variety of defects occurring in castings results from the very nature of the casting production technology, consisting of several operations such as designing and performing the casting mould and the liquid metal melting technology. Qualifications of employees, production character, technical equipment of foundry plant, are also important. When analysing the reasons of casting defects occurrence the so-called constant knowledge of reasons of their occurrence and methods of prevention as well as the current information from the process should be taken into account. The informative structure of analytical system of casting defects should be adequately put in order [3]. Out of seven basic tools of quality management the most important are Pareto-Lorenz diagram, based on empirically found correctness. The Ishikawa diagram enables the graphical presentation of relations between causes of the given problem and their hierarchy. Due to its shape, this diagram is often called a "fishbone" scheme. The Ishikawa diagram has a hierarchic structure: main causes are the nearest to a core, while intermediate causes, directly related to the main ones, constitute their expansion. The rule "from the whole to detail" is applied when preparing the diagram, which means that in the first place the main causes are determined, followed by the intermediate causes: of the second order, and when there is such necessity the causes of successive orders. The method of making the Ishikawa diagram is described in paper [7].

Some interesting solutions on application of data analysis and artificial intelligence to support the technological processes are also presented in the work of Kusiak [8], Perzyk [9, 10] and Kluska-Nawarecka [11, 12].

A main difficulty in the rebounding of moulding sands process (this process is the basis of green moulding sands quality control process) is the time factor. The final rebounding of moulding sands process is made just before and during the mould mixing. General trend in foundry is to use rotary mixers, what shortens the time of mixing to 90 s. The modern mixers are mostly equipped with complex control systems such as Michenfelder and others. These systems are based on measurements of moulding sands properties, but only those which, in the working moisture range, have a linear dependency on moisture. Compactibility is the most commonly used moulding sand property for such systems. The aim of the present work is finding moulding sand parameters which best characterize their moulding capability and which are the simplest in rapid implementation.

The results of analysis of the selected examination methods, which were instrumental in prediction the moulding sand quality on the basis of experimental data concerning sand compactibility and friability, are presented in the hereby paper.

2. Investigation methods

The results of data analysis performed by two various methods: neural networks and the Naive Bayesian Classifier are presented. The Statistica 8.0 program was used in the analysis.

2.1. Neural networks

Out of several statistic methods applied for aiding technological processes the artificial neural networks stand out. They can be applied in a very broad range of problems since they are capable of reflecting complex functions. Especially their non-linearity should be emphasised. They are gaining wider and wider application in the foundry industry, among others, for controlling melting processes in cupolas and arc furnaces, designing castings and supply systems, controlling the moulding sands treatments, the prediction of properties of cast alloys as well as the selection of die casting parameters.

Neural networks belong to modern self-training systems. The values of constants determining the significance of input data (network weights) are determined on the basis of experimental results (training examples) and by means of successive corrections, in such a way as to have the output data nearing the actual values. This is, the so-called supervised training (in other words: with a trainer), which is the most often used. Neural networks can realise several types of tasks depending on the kind of the problem, which is to be solved.

Network training is being done by solving the problem of optimisation of several variables function. From the mathematical point of view, we are aiming to find such weight values to get the smallest root-mean square error from all network outputs in relation to the experimental observations.

Corrections of network weights are performed several times for the whole training set. One cycle containing an error calculation and a weight modification is called an epoch. The end of training occurs the most often when the root-mean square error starts to increase for testing data. This is related to the possibility of the network over-training, which means an excessive fitting to the trained data without the ability to generalising expectations for other data.

2.2. Naive Bayesian Classifier

The Naive Bayesian Classifier is another statistic method being used in processes supporting the modern technologies. It is one of the machine trained method applied for solving problems of sorting and classification. The task of the Bayesian Classifier is to assign a new case to one of the decision class, while the whole set of decision classes must be finished and a priori defined. The Naive Bayesian Classifier is a statistic classifier based on the Bayes Rule.

As far as its effectiveness is concerned, this Classifier is comparable to classification algorithms by the Decision Tree induction method and to the classification based on neural networks. It is characterised by a high accuracy and scaling even for large data volumes. The Naïve Bayesian Classifier assumes that attributed values in classes are independent. This assumption is called the class conditional independence. Sometimes the class conditional independence assumption at given classes is called the Naive Bayesian Assumption.

3. Investigation results

Data concerning moulding sands properties used in the article do not come from a foundry, but from laboratory measurements. These results are not "disturbed" by products of moulding sands thermal destruction, contained in moulding sands, for example inactive binder. To network preparation the database containing 460 records for each properties was used. Data were divided into learning and testing set in proportion 80% – 20%.

The results obtained for two networks giving the best prediction of the experimental data for the sand compactibility are presented in Fig. 1 and 2. The best network is RBF 1-47-1, where the training (learning) quality equals 0.919712. The quality of the network is, by Statistica, the correlation coefficient between the output variable and its prediction done by the network. The correlation coefficient takes values between -1 and 1, when 1 indicates perfect consistency. When the correlation coefficient is very high, there is a possibility of network overlearning. Therefore, when choosing a network, its quality for a test and validation sets should be considered. So, in order to check the quality of the learning

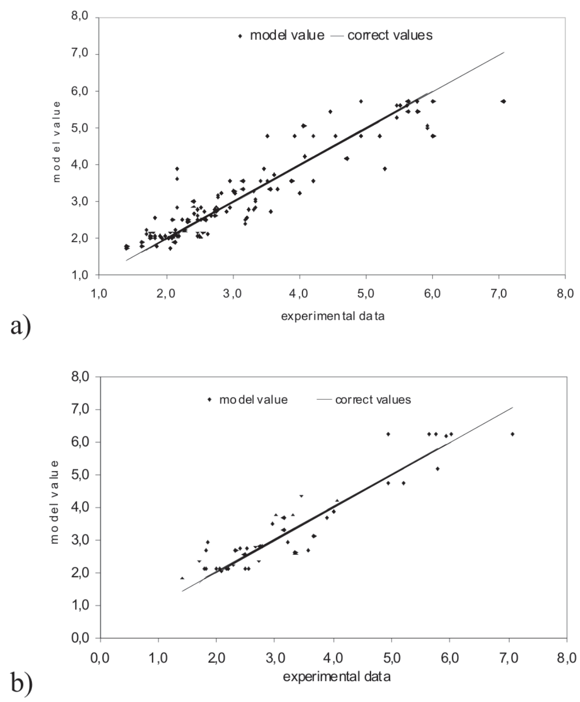


Fig. 1. Comparison of the distribution of data generated by the network and experimental data for moulding sand compactibility, model of the RBF 1-47-1 network, a) learning data set, b) test data set

process, the network verification has been made on the test sample. Testing quality for that network equals 0.90616. Using compactibility in order to determine the accurate moisture value generates values very similar to the experimental ones.

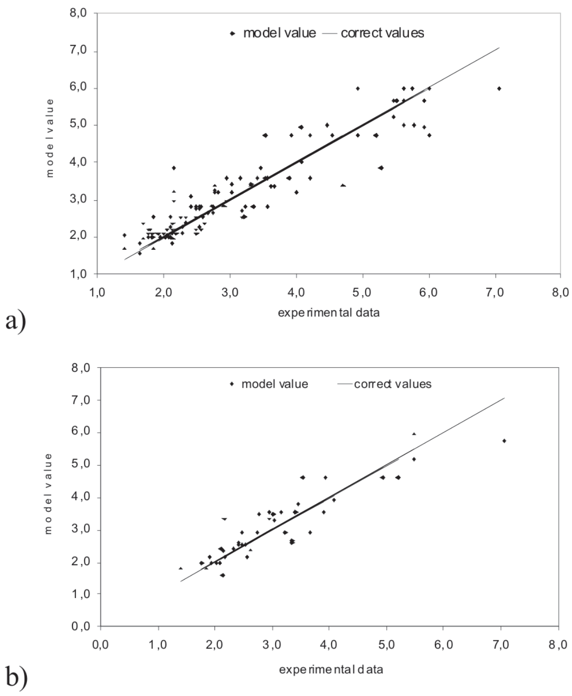


Fig. 2. Comparison of the distribution of data generated by the network and experimental data for moulding sand compactibility, model of the RBF 1-37-1 network, a) learning data set, b) test data set

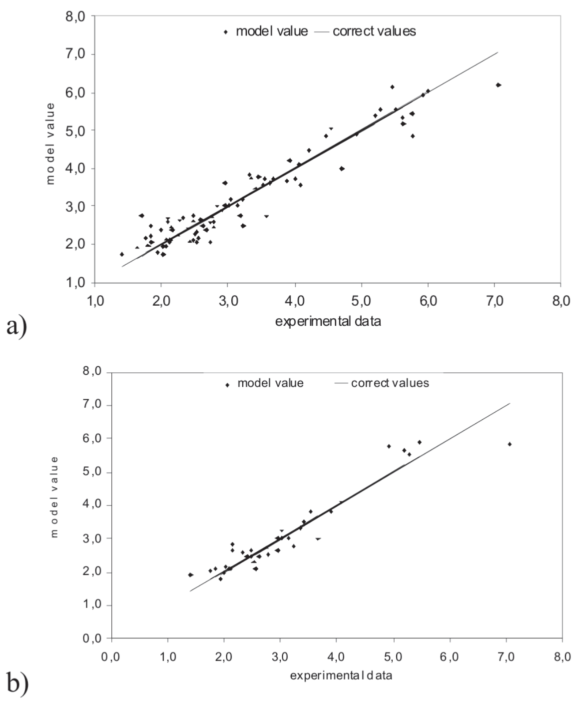


Fig. 3. Comparison of the distribution of data generated by the network and experimental data for moulding sand friability, model of the RBF 1-38-1 network, a) learning data set, b) test data set

The results obtained for two networks, in which the best prediction in relation to the moulding sand friability experimental data was achieved, are presented in Fig. 3 and 4. Solid line present actual moisture values, while points are output data for individual network models.

The best prediction is in the RBF 1-44-1 network diagram (Fig. 4). Using the moulding sand friability values in order to determine the accurate moisture value provides the network generated results very similar to the experimental data.

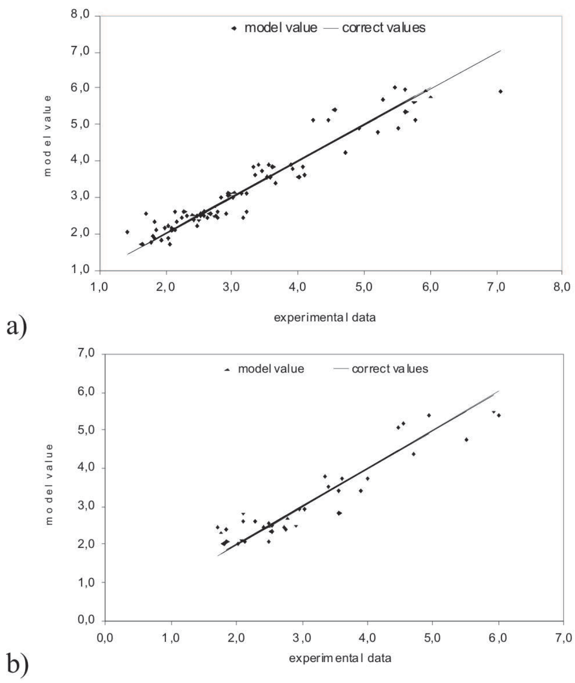
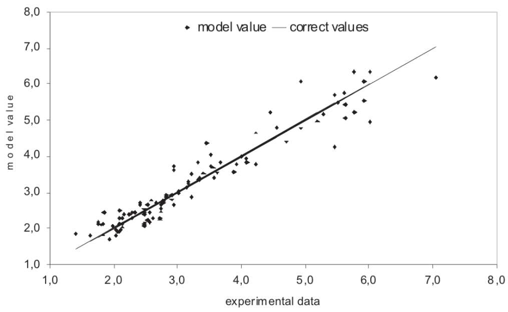


Fig. 4. Comparison of the distribution of data generated by the network and experimental data for moulding sand friability, model of the RBF 1-44-1 network, a) learning data set, b) test data set

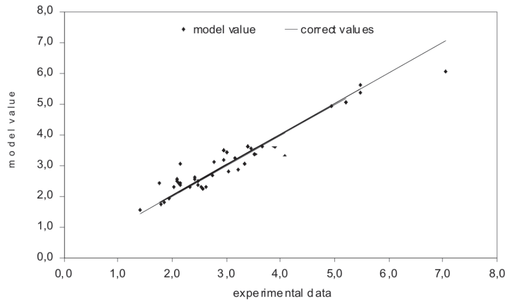
Two best models of complex networks, taking into consideration simultaneously experimental data of moulding sand compactibility and friability, are presented in Fig. 5 and 6.

The best prediction was obtained for the RBF 2-46-1 network, where the training value equals 0.963809, which is a slightly better result than the results obtained for the network of single parameters. Testing quality for that network equals 0.938295.

In the case of using the Naive Bayesian Classifier, 5 classes corresponding to the moulding sand moisture were assumed. Due to the program requirements the data for analysis were specially prepared. The obtained experimental results as well as the expected ones are presented in Fig. 7.

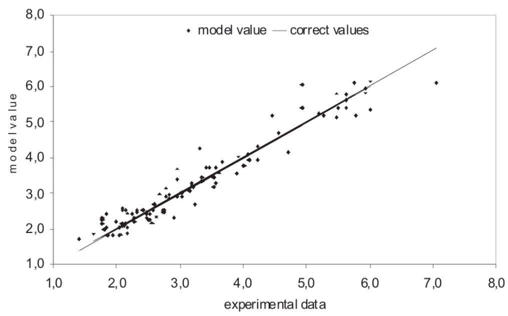


a)

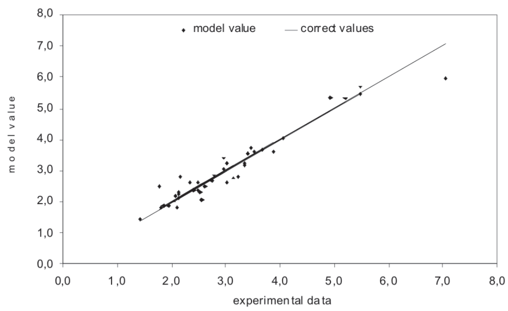


b)

Fig. 5. Comparison of data generated by the network and experimental data for the complex networks models with moulding sand compactibility and friability parameter model of the RBF 2-46-1 network, a) learning data set, b) test data set



a)



b)

Fig. 6. Comparison of data generated by the network and experimental data for the complex networks models with moulding sand compactibility and friability parameter model of the RBF 2-38-1 network, a) learning data set, b) test data set

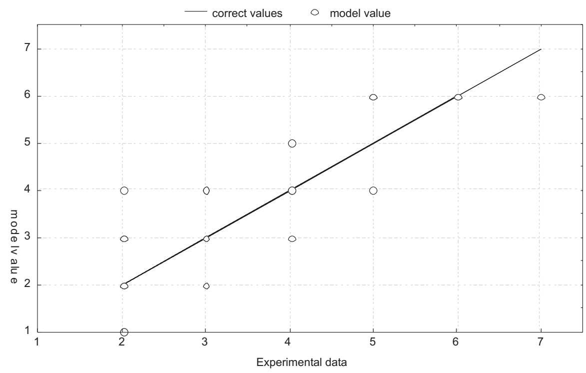


Fig. 7. Comparison of the expected data and the experimental ones (application of the Naïve Bayesian Classifier)

The summation of the analysis results is given in Table 1. The first column contains the class names (intervals of sand moisture). In next columns successively are placed: number of recording, number of right assignments to the given class, number of erroneous prediction and percentage fractions of right and wrong answers. The highest number of data contain classes from 1 to 3, which corresponds to the sand moisture interval in the same range. Generally, the accuracy of the model prediction for all classes equals 0.83.

The results of investigation presented above indicate, that in the discussed problem of predicting sand moisture the better prediction is given by neural networks. It is the best at creating the network models both when taking into account individual parameters and when considering several parameters. Therefore in the next stage of investigations the analysis results when using additional data concerning sand permeability are presented. These results are presented in Fig. 8 and 9.

The best prediction was obtained for the MLP 3-8-1 network where the training quality was 0.959703. Testing quality for that networks equals 0.950197. Results for quality all presented networks are shown in Table 2. Using additional data did not improve the network quality in relation to models, which were predicting on the basis of the moulding sand compactibility and friability data only.

TABLE 1

The summation of the Naive Bayesian Classifier analysis

Class name	Number of recording	Number of right assignments to the given class	Number of erroneous prediction	Right answers(%)	Wrong answers(%)
1	118	100	18	84.745	15.254
2	87	69	18	79.310	20.689
3	28	26	2	92.857	7.143
4	13	11	2	84.615	15.385
5	24	18	6	75	25

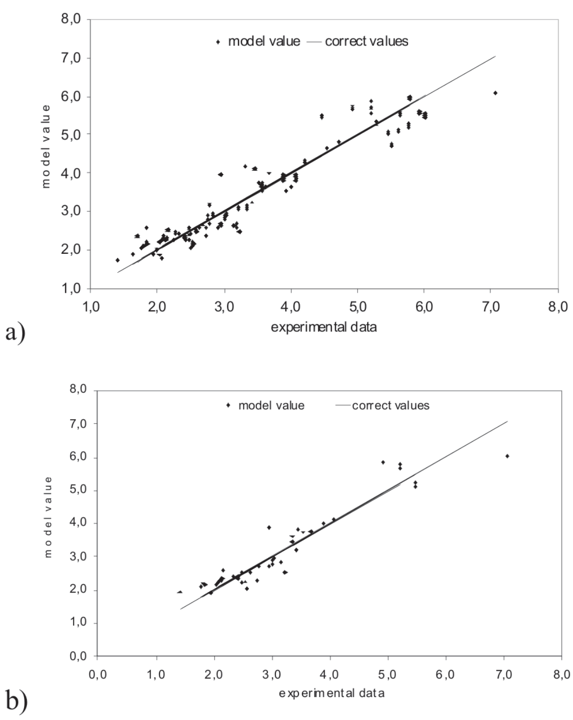


Fig. 8. Comparison of the distribution of data generated by the network and the experimental data for sand permeability, compactibility and friability model of the MLP 3-7-1 network, a) learning data set, b) test data set

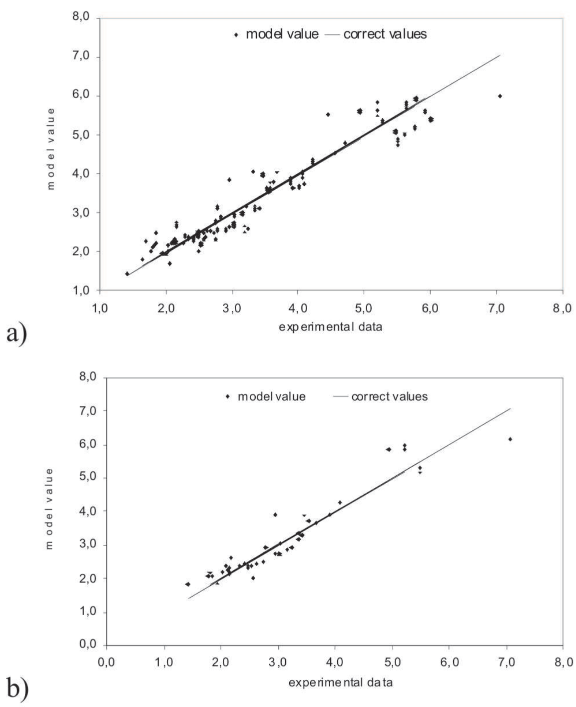


Fig. 9. Comparison of the distribution of data generated by the network and the experimental data for sand permeability, grindability and friability, model of the MLP 3-8-1 network, a) learning data set, b) test data set

TABLE 2

The quality of all presented networks, a) one input variable, b) two input variables, c) three input variables

Network name		Learning quality	Testing quality
a)	RBF 1-47-1	0.919712	0.90616
	RBF 1-37-1	0.918044	0.904577
	RBF 1-38-1	0.960040	0.934367
	RBF 1-44-1	0.965923	0.937756
b)	RBF 2-46-1	0.963809	0.938295
	RBF 2-38-1	0.971515	0.947976
c)	MLP 3-8-1	0.959703	0.950197
	MLP 3-7-1	0.955963	0.945612

4. Conclusions

1. Presented results of investigation show, that the best moulding sand parameters for prediction of the moulding sand moisture are compactibility and friability. The best results are achived for neural network built for both parameters.

2. Presented analysis proves that artificial neural networks are very useful application to support the process of rebonding of green moulding sand.

3. The comparison of the two proposed statistic methods indicates, that better results concerning the determination of the moulding sand moisture are achieved by artificial neural networks, which – on the basis of measurements of the selected sand properties – provide results very similar to the expected values.

4. In the case of the Naive Bayesian Classifier a number of erroneous classifications is too big.

5. The obtained results indicate the purposefulness of conducting further research on the application of artificial neural networks in foundry.

REFERENCES

[1] Z. Ignaszak, R. Sika, System do eksploracji wybranych danych produkcyjnych oraz jego testowanie w odlewni. Archiwum Technologii Maszyn i Automatyzacji **28**, 1, 61-72 (2008).

[2] Z. Ignaszak, J. Ciesiołka i in, Kompleksowe zastosowanie metod badań nieniszczących do optymalizacji technologii i kosztów wytwarzania odlewów, w aspekcie wzrostu efektywności wykorzystania komputerowych systemów symulacyjnych, raport końcowy projektu celowego nr 6 T08 2003 C 06228, Poznań – Śrem 2007 (maszynopis).

[3] I. Rojek-Mikołajczak, Integracyjna rola baz danych w przedsiębiorstwie, w: Computer Integration in Manufacturing, Poznań (1997).

[4] K. Bramczewski, M. Idee, S. Szwajkowski, System pomiaru i rejestracji temperatury zalewania form, instrukcja obsługi programu RTO PC Soft s.c, Piła (1996).

[5] R. Sika, Studium nad strukturą systemu SAP R/3 i możliwości jego dostosowania do zarządzania oraz sterowania jakością w Odlewni Żeliwa ŚREM S.A., praca dyplomowa pod kierunkiem Z. Ignaszaka, Politechnika Poznańska, Wydział Budowy Maszyn i Zarządzania (2006).

[6] R. Sika, Z. Ignaszak, Po wdrożeniu programu KonMas-final – jego wykorzystanie do analizy procesu produkcji odlewów na wydziale W6 – Odlewni Żeliwa ŚREM S.A., w: XI International Symposium – Modeling of casting and foundry processes, 26^{TM} 27 October 2006, Poznań-Śrem (2006).

[7] W. Łybacki, K. Zawadzka, Wspomaganie diagnostyki wad odlewów narzędziami zarządzania jakością, Archiwum Technologii Maszyn i Automatyzacji **28**, 1, 89-101 (2008).

[8] J. Kusiak, A. Żmudzki, A. Danielewska-Tułecka, Optimization of materials processing using a hybrid technique based on artificial neural networks, Archives of Metallurgy and Materials **50**, 3, 609-620 (2005).

[9] M. Perzyk, R. Biernacki, J. Kozłowski, Data mining in manufacturing: significance analysis of process parameters, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture **222**, 11, 1503-1516, (2008).

[10] M. Perzyk, J. Kozłowski, Zastosowanie metod inteligencji obliczeniowej do diagnostyki i optymalizacji procesów odlewniczych, Tendencje optymalizacji systemu produkcyjnego w odlewniach – Monografia, Komisja Odlewnictwa PAN Oddział Katowice, 143-156 (2010).

[11] S. Kluska-Nawarecka, Z. Górny, D. Wilk-Kołodziejczyk, A. Smolarek-Grzyb, The Logic of Plausible Reasoning in the Diagnostics of Castings Defects, Archives of Metallurgy and Materials **52**, 3, 375-380 (2007).

[12] S. Kluska-Nawarecka, Z. Górny, Metody sztucznej inteligencji w doskonaleniu procesów produkcyjnych w odlewnictwie, Tendencje optymalizacji systemu produkcyjnego w odlewniach – Monografia, Komisja Odlewnictwa PAN Oddział Katowice, 187-212 (2010).

Received: 10 July 2010.