

## **Techniki zgłębiania danych w prognozowaniu wytrzymałości na ściskanie betonu z dodatkiem popiołu lotnego**

### **Prediction of compressive strength of concrete containing fly ash using data mining techniques**

#### **1. Wstęp**

Wytrzymałość betonu na ściskanie jest powszechnie wykorzystywaną właściwością w projektowaniu mieszanin betonowych. Prognozowanie wytrzymałości betonu przed jego zastosowaniem pozwala udoskonalić projektowanie i kontrolę jakości. Ponadto poprawne prognozowanie wytrzymałości betonu oszczędza czas, przyspieszając realizację, głównie dużych budowli z kilkoma etapami betonowania. Jednak względy środowiskowe i potrzeba stosowania wysokowartościowych betonów musi prowadzić do ich bardziej złożonego składu, w porównaniu z tradycyjnymi. W obecnych czasach dostępność przydatnych do wykorzystania odpadów przemysłowych powoduje stałe stosowanie dodatków mineralnych do produkcji betonu, a szczególnie popiołu lotnego z elektrowni. Dlatego czynniki, które wpływają na wytrzymałość betonu na ściskanie są coraz liczniejsze i bardziej złożone, prowadząc do nieliniowych zależności. Zależności te pochodzą zwykle z danych doświadczalnych lub są uzyskiwane metodą regresji ze względu na brak do tej pory ogólnych reguł opisujących te zjawiska i ten wysoce złożony układ. Pod tym względem metody tradycyjne, polegające na uogólnieniu zdobytego już doświadczenia nie są wystarczająco dokładne do formułowania zadawalających zależności. Z tego powodu modele inteligentne, z możliwością uczenia się przy wykorzystaniu przykładów, zastosowano w prognozowaniu wytrzymałości betonu na ściskanie (Gupta, 2007, Kim *et al.*, 2002, Lai i Serra, 1997, Saridemir, 2009 i Topçu i Saridemir, 2008).

W niniejszym artykule zastosowano do prognozowania wytrzymałości betonu technikę zgłębiania danych, przy uwzględnieniu właściwości najszerzej stosowanego dodatku mineralnego do betonu, czyli popiołu lotnego. Techniki te są ważnymi, inteligentnymi narzędziami, które uczą się za pomocą przykładów i doświadczeń, a były już z powodzeniem stosowane w prognozowaniu wytrzymałości betonu przez innych autorów (Gupta, 2007, Kim *et al.*, 2002, Lai i Serra, 1997, Saridemir 2009 i Topçu i Saridemir, 2008). Prognozy te jednak nie uwzględniały jakości stosowanego popiołu lotnego. W rzeczywistości, zgodnie z posiadanymi przez

#### **1. Introduction**

The concrete compressive strength is the most used mechanical property in the designing of concrete structures. The prediction of concrete strength before its placement allows to improve planning and quality control. Moreover, a well defined concrete strength prediction can save time, accelerating the overall construction process, mainly in large constructions with a lot of concreting stages. However, environmental problems and the need of high performance concrete application have led to more complex mixes, than the traditional ones. Nowadays, the economic problems as well as the necessity to diminish the CO<sub>2</sub> emission, caused the systematic use of mineral additions in concrete, very frequently fly ash from electrical power plants. Therefore, the factors that affect the concrete compressive strength have increased in number and complexity leading to a nonlinear relationships. These relationships are usually empirical or based on correlation analysis due to the fact that until today there is no general law describing the phenomena and explaining this hugely complex system. In this context the traditional methods based on generalization of previous experience are not enough accurate to provide satisfactory relationships. That is why intelligent models with the capability of learning with examples have been applied in the prediction of the concrete compressive strength (Gupta, 2007, Kim *et al.*, 2002, Lai and Serra, 1997, Saridemir, 2009 and Topçu and Saridemir, 2008).

In this study the forecasting of concrete strength was carried on using Data Mining (DM) techniques, taking into account the properties of the most used mineral addition in concrete i.e. fly ash (FA). These techniques are powerful intelligent tools that learn with examples and experiences and were applied successfully to predict concrete strength by other authors (Gupta, 2007, Kim *et al.*, 2002, Lai and Serra, 1997, Saridemir 2009 and Topçu and Saridemir, 2008). However, these predictions did not take into account the quality of the FA used. In fact, to the best of our knowledge, the physical and chemical parameters of the FA have never been used before in these predictions.

nas wiadomościami, właściwości fizykochemiczne popiołu lotnego nie były dotychczas wykorzystywane w takim prognozowaniu.

W tej pracy porównano możliwości prognostyczne kilku technik zgłębiania danych, przy uwzględnieniu fizykochemicznych właściwości popiołu lotnego w celu liczbowego określenia jego aktywności i wpływu na wytrzymałość betonu na ściskanie. Najważniejsze właściwości popiołu lotnego zazwyczaj oznaczane i dostępne w literaturze zostały potraktowane jako zmienne wejściowe w modelach zgłębiania danych.

Na początku artykułu zamieszczono krótki opis właściwości popiołu lotnego, wpływających na wytrzymałość betonu na ściskanie. Następnie omówiono definicje technik zgłębiania danych i ich zastosowanie do prognozowania właściwości betonu. Wyjaśniono także jak oceniać różne algorytmy zgłębiania danych. Na koniec przedstawiono uzyskane wyniki wraz z ich omówieniem oraz z wnioskami.

## 2. Wpływ właściwości popiołu lotnego na wytrzymałość betonu na ściskanie

Stosowanie popiołu lotnego do produkcji betonu nie jest pomysłem nowym. Po raz pierwszy popiół lotny znalazł takie zastosowanie w latach trzydziestych dwudziestego wieku w USA i wpływ popiołu lotnego na właściwości betonu jest dostatecznie rozpoznany i dobrze udokumentowany (Wesche, 1991, Malhotra i Ramezaniapour, 1994, ACI Committee 232, 1996, Joshi i Lohtia, 1997 i Camões, 2002). Dlatego w niniejszym rozdziale zamierzamy tylko przedstawić główne zagadnienia bezpośrednio związane z przeprowadzonymi badaniami, dotyczącymi wpływu fizykochemicznych właściwości popiołu lotnego na wytrzymałość na ściskanie betonu.

Fizykochemiczne właściwości popiołu lotnego mogą zmieniać się znacznie w zależności od jego pochodzenia, a nawet w różnych partiach pochodzących z tej samej elektrowni, jednak niektóre właściwości można uważać za typowe i stałe.

### 2.1. Właściwości fizyczne

Ziarna popiołu lotnego są kształtu kulistego i zawierają głównie szkło (80%) i mają dużą miąłkość. Niektóre popioły lotne zawierają ziarna nieregularne i graniaste. Większość ziaren ma średnice zawarte w przedziale od 1 do 150  $\mu\text{m}$ , bardzo zbliżone do uziarnienia cementu. Średnia średnica ziaren popiołu lotnego zazwyczaj jest podobna do średnicy ziaren cementu zawierającej się w przedziale od 7  $\mu\text{m}$  do 12  $\mu\text{m}$  (Joshi i Lohtia, 1997). Często powierzchnia właściwa Blaine'a zawarta jest w przedziale 250  $\text{m}^2/\text{kg}$  do 550  $\text{m}^2/\text{kg}$  (Alonso i Wesche, 1991).

Z pewnym przybliżeniem można powiedzieć, że właściwości fizyczne popiołu lotnego zależą od jego pochodzenia. Zgodnie z (Malhotrą i Ramezaniapourem, 1994) źródło popiołu lotnego nie jest powiązane z jego uziarnieniem lub powierzchnią właściwą. Autorzy ci uważają, że istnieje bardzo słaba korelacja między uziar-

In this work the predictive capacity of several DM techniques were compared to forecast the concrete compressive strength, taking into account physical-chemical properties of the FA used, in order to quantify its activity and influence on concrete compressive strength. Therefore the main properties of FA, usually determined and currently available, were considered as input variables in the DM models.

## 2. Influence of FA properties on concrete compressive strength

The incorporation of FA in concrete is not new, began in the 1930s in the USA (ACI Committee 232, 1996), and the effects of FA in concrete performance are sufficiently understood and well documented (Wesche, 1991, Malhotra and Ramezaniapour, 1994, ACI Committee 232, 1996, Joshi and Lohtia, 1997 and Camões, 2002). Therefore, in this section only the main aspects directly related to the present work, namely the effect of physical-chemical properties of FA on concrete performance, mainly on compressive strength are shortly presented.

Unless physical-chemical properties of FA can vary considerably depending on their origin and even between different supplies from the same power station, some properties can be generalized.

### 2.1. Physical properties

FA particles have spherical shape, are essentially vitreous (80 %), and have a high fineness. Most of the particles have a diameter in the range of 1  $\mu\text{m}$  to 150  $\mu\text{m}$ , and the average FA fineness is usually similar to cement (Joshi and Lohtia, 1997). The Blaine specific surface ranges frequently from 250  $\text{m}^2/\text{kg}$  to 550  $\text{m}^2/\text{kg}$  (Alonso and Wesche, 1991). According to Malhotra and Ramezaniapour, 1994, the FA source is not related to its fineness or its specific surface. The authors consider that there is a very slight correlation between fineness, as measured by percentage retained on 45  $\mu\text{m}$  sieve and Blaine specific surface area.

FA particles larger than 125  $\mu\text{m}$  are frequently very porous. The occurrence of these particles is associated with large amounts of unburned carbon, which is responsible for the high specific surface area typically found in FA. High level of carbon imply greater water demand in concrete containing FA with high loss on ignition (LOI) (Giergiczny and Giergiczny 2010). The carbon content also affects the concrete resistance to freezing and thawing; the higher the carbon content of FA, the lower the resistance of concrete to frost attack (Alonso and Wesche, 1991). According to these authors, FA should have a particle size and specific surface similar to or lower than cement to avoid variations in physical properties of concrete, particularly its workability. This recommendation is linked to the presence of carbon in FA, avoiding high amounts of particles larger than 125  $\mu\text{m}$ , endowed with high porosity and with greater concentration of carbon particles.

It is commonly accepted that more fineness leads to greater pozzolanic activity. Like all heterophase reactions it occurs more rapidly

nieniem mierzonym udziałem procentowym frakcji pozostającej na sicie 45  $\mu\text{m}$  i powierzchnią właściwą Blaine'a.

Ziarna popiołu lotnego większe niż 125  $\mu\text{m}$  są bardzo porowate. Właściwość ta jest związana z dużym udziałem węgla. Niespalony węgiel jest powodem dużej powierzchni właściwej, którą zazwyczaj charakteryzują się popioły lotne. Wysoka zawartość węgla pociąga za sobą większe zapotrzebowanie na wodę w betonie zawierającym popiół lotny z dużymi stratami prażenia (LOI – loss on ignition). Zawartość węgla wpływa także na mrozoodporność betonu: im wyższa zawartość popiołu lotnego tym mniejsza odporność betonu (Alonso i Wesche, 1991). Według tych autorów popiół lotny powinien mieć wielkość ziaren i powierzchnię właściwą podobną lub mniejszą od cementu, aby uniknąć niejednorodnych właściwości fizycznych betonu, a w szczególności pogorszenia urabialności. To zalecenie jest związane z obecnością węgla w popiele lotnym; zmniejszenie dużej zawartości ziaren większych od 125  $\mu\text{m}$ , bardzo porowatych, złożonych głównie z węgla.

Przyjmuje się, że duży stopień rozdrobnienia jest związany z większą aktywnością pucolanową. Także reakcje heterogeniczne przebiegają szybciej ze wzrostem stopnia rozdrobnienia i przypuszcza się, że aktywność pucolanowa popiołu lotnego powinna zależeć od jego powierzchni właściwej, na której może zajść reakcja (Jalali, 1991, Neville, 1995).

Kulisty kształt ziaren popiołu lotnego jest korzystny szczególnie z punktu widzenia wodożądności, a duża powierzchnia właściwa zwiększa jego reaktywność w stosunku do wodorotlenku wapniowego (Neville, 1995). W związku z tym panuje pogląd że im większa jest powierzchnia właściwa i niższe straty prażenia, tym większa będzie aktywność popiołu lotnego w betonie.

## 2.2. Właściwości chemiczne

Popioły lotne zawierają najwięcej krzemionki,  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$ , i tlenku wapnia. Znacznie mniejsza jest zawartość  $\text{MgO}$ ,  $\text{Na}_2\text{O}$ ,  $\text{K}_2\text{O}$ ,  $\text{SO}_3$ ,  $\text{MnO}$  i  $\text{TiO}_2$ . Popiół lotny zawiera także cząstki nieopalonego węgla (Alonso i Wesche, 1991, Malhotra i Ramezianpour, 1994 i ACI Committee 232, 1996). Zawartości wspomnianych czterech głównych składników popiołu lotnego podlegają znacznym wahaniom, jednak można przewidywać, że ich udział będzie zgodny z normą (ACI Committee 232, 1996) w następujących granicach:  $\text{SiO}_2$  – 35 % do 60 %;  $\text{Al}_2\text{O}_3$  – 10 % do 30 %;  $\text{Fe}_2\text{O}_3$  – 4 % do 20 %;  $\text{CaO}$  – 1 % do 35 %.

Aktywność pucolanowa popiołów lotnych jest ściśle związana z zawartością  $\text{SiO}_2$ , gdyż aktywna krzemionka łączy się z wodorotlenkiem wapnia z utworzeniem uwodnionego krzemianu wapnia. Alonso i Wesche, 1001 wyrażają pogląd, że popioły lotne, w których zawartość  $\text{SiO}_2$  jest mniejsza od 35% mają małą aktywność pucolanową i nie powinny być stosowane do produkcji betonu. Natomiast Halstead, 1986, Mehta, 1985 oraz Joshi i Lohtia, 1997 są zdania, że, z wyjątkiem wapna, zawartości innych składników nie mają znacznego wpływu na właściwości pucolanowe Alonso i Wesche, 1991 stwierdzają, że popiół lotny zawierający więcej wap-

with increasing fineness of the particles then the pozzolanic activity of fly ash must be dependent on the specific surface area (Jalali, 1991, Neville, 1995). Also the spherical shape of FA particles is particularly advantageous for good workability of concrete mixes (Neville, 1995). In this context, it is supposed that the higher specific surfaces and the lower LOI the better activity of FA in concrete will be.

## 2.2. Chemical properties

In the chemical composition of FA four constituents have the major content, namely  $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$  and  $\text{CaO}$ . The FA pozzolanic activity is closely related to the  $\text{SiO}_2$  content, and there are the silica anions  $\text{H}_n\text{SiO}_4^{(4-n)}$  which will react with calcium ions and  $\text{OH}^-$  giving additional quantities of calcium silicate hydrates. According to Alonso and Wesche, 1991, FA with  $\text{SiO}_2$  content lower than 35 % are virtually inactive as pozzolans and should not be used in concrete. However, Halstead, 1986, Mehta, 1985 and Joshi and Lohtia, 1997, have a different opinion and consider that in terms of chemical composition, with the exception of calcium content, the variation of FA constituents does not affect significantly the cementitious or pozzolanic properties. Alonso and Wesche, 1991, also indicate that FA with high content of lime (15 % to 40 %) can have hydraulic properties and their use in concrete should be avoided. Although the content of  $\text{SiO}_2$  appear to be related to the pozzolanic activity, its reduced levels do not imply negative effects on the behavior and properties of fresh or hardened concrete (Camões, 2002). Assuming the minimum values for the total amount of oxides ( $\text{SiO}_2 + \text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3$ ) is criticized by several authors (Swamy, 1993, ACI Committee 262, 1996 and Joshi and Lohtia, 1997). While it may be understandable attempt to ensure the presence of sufficient glass content (according to Mindess, 1994, the higher the glass phase is the better FA will be), one should not confuse the reactivity of the glass, however, the development of mechanical properties is always dependent on the combined effect of several factors, such as the fineness and the strength of cement. Malhotra and Ramezianpour, 1994, mention the good correlation between the amount of  $\text{SiO}_2 + \text{Al}_2\text{O}_3$  and the pozzolanic activity at long term. The same authors also report that high amount of  $\text{Fe}_2\text{O}_3$  can be present in FA as nonreactive hematite and magnetite. These aspects should be the reason why some authors obtained low correlation between the pozzolanic activity index and the content of  $\text{SiO}_2 + \text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3$ . Sulphates,  $\text{SO}_3$ , may affect the optimum FA content as they can influence setting time and mechanical properties of concrete. The excess of  $\text{SO}_3$  in hardened concrete must be avoided as it can cause inner concrete corrosion and contribute also to a worse performance, particularly when concrete is subjected to sulphate attack (Jalali, 1991).

The content of carbon in FA is decisive for the water demand of paste, mortar and concrete. The total water needed to obtain a paste with the same consistency is greater the higher the carbon content. The carbon contained in FA has high porosity and specific surface area, being able to absorb not only significant amounts of water but also the chemical admixtures. According to Alonso and Wesche, 1991, in general, the lower the carbon content the better

na (15% do 40%) ma właściwości hydrauliczne i jego dodatek do betonu jest niepożądany.

Opinie na temat roli omówionych składników nie są zgodne. Norma (ACI Committee 232, 1996) podkreśla brak wyników badań dotyczących wpływu zawartości  $\text{SiO}_2$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{Fe}_2\text{O}_3$  na właściwości popiołu lotnego. Aczkolwiek zawartość  $\text{SiO}_2$  jest związana z aktywnością pucolanową, zmniejszona zawartość tego składnika nie pociąga za sobą negatywnego wpływu na właściwości świeżego lub stwardniałego betonu (Camões, 2002). Ustalanie minimalnych zawartości sumarycznej zawartości tlenków ( $\text{SiO}_2 + \text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3$ ) jest krytykowane przez niektórych autorów (Swamy, 1993, ACI Committee 262, 1996 oraz Joshi i Lohtia, 1997). Opinia o ważnym znaczeniu zawartości szkła może być uzasadniona (według Mindessa, 1994 im większa jego zawartość, tym lepszy popiół lotny), jednak na poprawę właściwości mechanicznych betonu ma także wpływ szereg innych czynników, a przede wszystkim właściwości cementu. Malhotra i Ramezianpour, 1994 mówią o dobrej korelacji pomiędzy sumaryczną zawartością  $\text{SiO}_2 + \text{Al}_2\text{O}_3$  i aktywnością pucolanową popiołu, w długim okresie. Ci sami autorzy podkreślają również, że zawartość  $\text{Fe}_2\text{O}_3$  w prawie każdym popiele lotnym jest związana z niereaktywnym hematytem i magnetytem. Może to być powodem uzyskiwania słabej korelacji aktywności pucolanowej, oznaczanej jako wytrzymałość na ściskanie betonu, a sumaryczną zawartością:  $\text{SiO}_2 + \text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3$  w popiele.

Siarczany wpływają także na korzystną zawartość popiołu lotnego w betonie, gdyż wiążą się z jego właściwościami mechanicznymi oraz czasem wiązania. Maksymalna zawartość popiołu musi uwzględniać problem nadmiaru  $\text{SO}_3$  w betonie, gdyż może on znacznie pogorszyć właściwości betonu, szczególnie gdy może on być narażony na działanie siarczanów (Jalali, 1991).

Węgiel, którego rola w popiele lotnym została już omówiona, nie jest oznaczany bezpośrednio, tylko jako straty prażenia. Straty prażenia, oprócz węgla, obejmują także wodę związaną w hydratych i węglany zawarte w popiele lotnym. Natomiast zakłada się, że straty prażenia, bez większego błędu, są w przybliżeniu równe zawartości węgla.

Zawartość węgla w popiele lotnym jest czynnikiem wpływającym na zapotrzebowanie wody w zaprawach i betonach. Węgiel zawarty w popiele lotnym ma dużą porowatość i powierzchnię właściwą co powoduje absorpcję nie tylko znacznej ilości wody, lecz także domieszek chemicznych dodawanych podczas mieszania betonu, a obejmujących superplastyfikatory, środki napowietrzające lub przyspieszacze twardnienia.

Według Alonso i Wesche, 1991 na ogół im mniejsza zawartość węgla, tym lepszy popiół lotny jako dodatek do zapraw i betonów. Norma (ACI Committee 363, 1992) stanowi, że kiedy popiół lotny stosowany w BWW powinien mieć straty prażenia niższe niż 3%, chociaż wyższe wartości są podawane w normach europejskich. Day, 1995, stwierdza, że zawartość węgla nie powinna przekraczać 8 %, a wskazane są niższe zawartości. Jednakże Malhotra i Ramezianpour, 1994 konkludują, że węgiel nie wpływa

the FA for use in mortar and concrete. ACI Committee 363, 1992, states that when used in high strength concretes the desirable LOI should be less than 3 %, although higher values are considered in standards, enabling its acceptance. Day, 1995, states that the carbon content should not exceed 8 %, much lower levels are preferred. However, Malhotra and Ramezianpour, 1994, concluded that the carbon does not significantly affect the pozzolanic activity index, determined in relation to the compressive strength.

### 3. Data Mining

#### 3.1. Definition and applications to the concrete behaviour

The process of Knowledge Discovery in Databases encompasses three main stages: pre-processing, Data Mining and post-processing. In the Data Mining stage an algorithm is applied for extraction of patterns from the data (Fayyad *et al.*, 1996). Generally, in all Data Mining techniques there are a set of training examples and a set of testing examples. The algorithms learn with the training examples and are tested with the testing examples. During the learning process the several parameters of the algorithms are adjusted to optimize the results. To assess the accuracy of the algorithms some metrics can be used most of them based on errors. After the validation of the algorithms they can be used as models for forecasting the values of the output variables.

Many authors have been succeeded with the application of intelligent tools. Several authors (Kasperkiewicz *et al.*, 1995, Lai and Serra, 1997, Ni and Wang, 2000, and Kim and Kim, 2002) used neural networks to predict 28-day compressive strength of concrete with different composition. Topçu and Saridemir, 2008, developed neural networks and fuzzy logics for predicting the 7, 28 and 90 days compressive strength of concretes containing high-lime and low-lime fly ash. Zarandi *et al.*, 2008, developed fuzzy polynomial neural networks (FPNN) to predict the 28-day compressive strength of concrete. They constructed six different FPNN architectures and used experimental data of 458 different concrete mixes, collected from three distinct sources. They were well succeeded in predicting the compressive strength of concrete mixes. Saridemir, 2009, based on several experimental results gathered from the literature, used neural networks to predict the compressive strength of concretes containing metakaolin and silica fume for different concrete mixes and several curing days. Özcan *et al.*, 2009, basing on experimental works, applied artificial neural network and fuzzy logic models to predict the compressive strength of silica fume concrete. They obtained good results using several concrete mixes at five different ages. Yang *et al.*, 2003, applied neural networks to predict compressive strength, slump value and mix portions of concrete. The networks were trained using standard tables of two companies.

Ruan Xiang, 2009, and Zhitao *et al.*, 2008, applied support vector machine (SVM) to predict concrete carbonation. Chen *et al.*, 2009, have shown that the SVM has a good performance for estimating

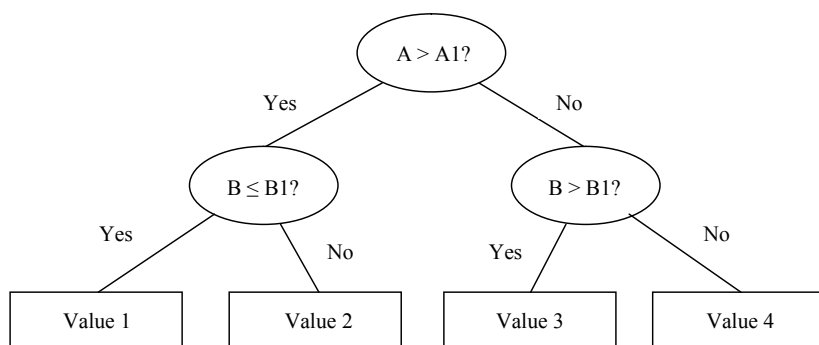


Fig. 1. Example of a regression tree

Rys. 1. Przykład drzewa regresyjnego

w znacznym stopniu na wskaźnik aktywności pucolanowej, określany w oparciu o wytrzymałość na ściskanie.

### 3. Zgłębianie danych

#### 3.1. Definicja i zastosowania w badaniach właściwości betonu

Proces uzyskiwania wiadomości z baz danych składa się z trzech głównych etapów: wstępna analiza danych, zgłębianie wiedzy i obróbka znalezionych informacji. W etapie zgłębiania danych stosuje się algorytm w celu znalezienia regularnych wzorców i systematycznych współzależności pomiędzy zmiennymi (Fayyad *et al.*, 1996). W zasadzie we wszystkich technikach zgłębiania danych są przykłady ciągów uczących i przykłady ciągów testujących. Algorytmy te uczą się za pomocą ciągów uczących i są sprawdzane za pomocą ciągów testujących. Podczas uczenia się kilka parametrów algorytmu jest modyfikowanych w celu optymalizacji wyniku. Do oszacowania dokładności algorytmów można stosować różne metody, które w większości oparte są na ocenie wielkości błędu. Po zatwierdzeniu algorytmów używa się ich jako modeli do prognozowania wartości zmiennych wyjściowych.

Wielu autorów z powodzeniem stosuje narzędzia sztucznej inteligencji. Kilku badaczy (Kasperkiewicz *et al.*, 1995, Lai i Serra, 1997, Ni i Wang, 2000, oraz Kim i Kim, 2002) stosowało sieci neuronowe do prognozowania wytrzymałości na ściskanie betonu po 28 dniach, w funkcji jego składu. Topçu i Saridemir, 2008, zastosowali sieci neuronowe i logikę rozmytą do prognozowania wytrzymałości na ściskanie betonów po 7, 28 i 90 dniach, z popiołem lotnym o małej i dużej zawartości wapna. Zarandi i in., 2008, zastosowali rozmyte wielomianowe sieci neuronowe RWSN do prognozowania wytrzymałości betonu na ściskanie po 28 dniach. Zbudowali sześć odmiennych architektur RWSN, wykorzystując dane doświadczalne obejmujące 458 betonów o różnym składzie, pochodzących z trzech różnych źródeł. Okazało się, że prognozy wartości wytrzymałości na ściskanie tych betonów były bardzo trafne. Saridemir, 2009, zebrał dane z kilku eksperymentów opisanych w literaturze i zastosował sieci neuronowe do prognozowania wytrzymałości na ściskanie betonów zawierających meta-kaolin i pył krzemionkowy po kilku okresach twardnienia. Özcan

the exposed temperature of fire-damaged concrete structures. Gupta, 2007, applied the SVM to predict the compressive strength of high performance concrete.

#### 3.2. DM algorithms

The DM algorithms used in this study were Regression Trees (RT), Multiple Regressions (MR), Artificial Neural Networks (ANN), Support Vector Machines (SVM) and k-Nearest Neighbours (k-NN).

The Decision Trees (Quinlan, 1986) have an inverted tree structure composed of nodes and descendent branches. The result of a test performed at each node indicates the branch to continue the process. This process is repeated until the final decision can be taken and a class is attributed to the register. The regression trees are a particular case of the decision trees where the classes are replaced by values (Fig. 1).

The multiple regressions are similar to simple regressions but with several independent variables, instead of one independent variable.

The artificial neural networks are based on the architecture of the human brain. They are based on process units named neurons that have a large number of interconnections that allow the communication between them. Each neuron has a set of input and output connections that have an associate weight. The level of activation of a neuron is determined by an activation function (Haykin, 1999 and Cortez, 2010). The neuron receives signal from the input connections and calculates a new activation value which is send through the output connections. The value is a result of the calculation of the value of the neuron activation, using the activation function that has as input argument the value of the weighted sum of the input values. In this study it was used the multilayer perceptron configuration (Haykin, 1999) composed by an input layer, a hidden layer and an output layer (Fig. 2).

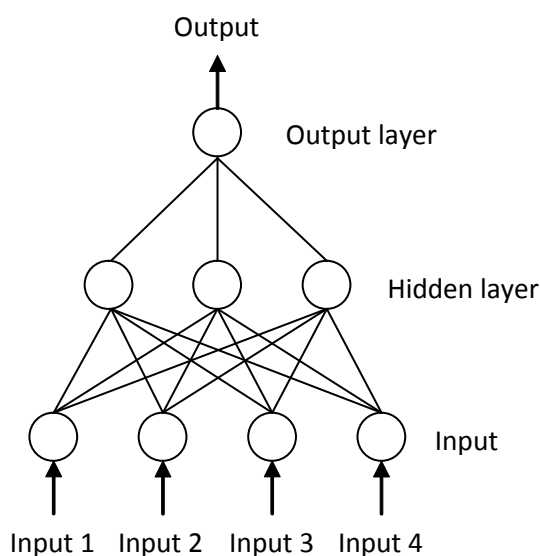


Fig. 2. Example of a multilayer perceptron

Rys. 2. Przykład wielowarstwowego perceptronu

i in., 2009 przeprowadzili doświadczenia i zastosowali sztuczne sieci neuronowe oraz modele logiki rozmytej do prognozowania wytrzymałości na ściskanie betonu zawierającego pył krzemionkowy. Otrzymali poprawne wyniki wykorzystując dane pomiarowe uzyskane po pięciu różnych okresach twardnienia betonów o różnym składzie. Yang i in., 2003, zastosowali sieci neuronowe do prognozowania wytrzymałości na ściskanie betonów o różnym składzie. Sztuczne sieci neuronowe uczono wykorzystując tablice normowe z dwóch firm.

Ruan Xiang, 2009, i Zhitao *et al.*, 2008, zastosowali maszynę wektorów nośnych (MWN) do prognozowania stopnia karbonatyzacji betonu. Chen i in., 2009, wykazali, że maszyna wektorów nośnych ma dobrą efektywność w prognozowaniu temperatury, na której działanie był wystawiony Beton w trakcie pożaru. Gupta, 2007, zastosował MWN do prognozowania wytrzymałości na ściskanie BWW.

### 3.2. Algorytmy zgłębiania danych

W niniejszej pracy zastosowano następujące algorytmy zgłębiania danych: drzewa regresji (DR), regresję wieloraką (RW), sztuczne sieci neuronowe (SSN), maszynę wektorów nośnych (MWN) i algorytm k najbliższych sąsiadów (algorytm k – nn).

Drzewa decyzyjne (Quinlan, 1986) mają budowę odwróconego drzewa złożonego z węzłów i opadających gałęzi. Wynik testu przeprowadzonego w każdym węźle wskazuje gałąź kontynuacji procesu. Proces jest powtarzany, aż do podjęcia końcowej decyzji i liść jest przypisany do atrybutu. Drzewa regresyjne są szczególnym przypadkiem drzew decyzyjnych, w których liście są zamienione na wartości, co jest pokazane na rysunku 1.

Regresja wieloraka jest podobna do regresji jednokierunkowej, lecz z kilkoma niezależnymi zmiennymi zamiast jednej zmiennej niezależnej.

Sztuczne sieci neuronowe są wzorowane na architekturze ludzkiego mózgu. Są budowane na jednostkach procesowych nazywanych neuronami, które mają dużą liczbę połączeń pozwalających na komunikację między nimi. Każdy neuron ma zbiór połączeń wejścia i wyjścia z przypisanymi doń wagami. Poziom aktywacji neuronu jest określany przez funkcję aktywacji (Haykin, 1999 and Cortez, 2010). Neuron otrzymuje sygnały z połączeń wejściowych i oblicza nową wartość aktywacji, która jest przesyłana przez połączenia wyjścia. Wartość jest wynikiem obliczenia aktywacji neuronu przy zastosowaniu funkcji aktywacji, która ma jako argument wejściowy wartość ważonej sumy wartości wejściowych. W niniejszej pracy zastosowano architekturę wielowarstwowego perceptronu (Haykin, 1999) złożonego z warstwy wejściowej, warstwy ukrytej i warstwy wyjściowej, co pokazano na rysunku 2.

Maszyny wektorów nośnych były przedmiotem badań Cortesa and Vapnika, 1995 dla klasyfikacji binarnej. Podstawowym pomysłem było podzielenie zbioru

The Support Vector Machines were developed by Cortes and Vapnik, 1995, for binary classification. The basic idea was to separate the dataset in two classes or categories. For this aim, a hyperplane in a multidimensional space separates the examples in sets of the same category. The optimal separating hyperplane between the two classes is maximizing the margin between the closest points of the two classes (Meyer, 2010). The points lying in the boundaries are called support vectors and the optimal separating hiperplane is at the middle of the margin. The points lying on the wrong side are weighted down to reduce they influence (Meyer, 2010). When a linear separator cannot be found there is a transformation via kernel techniques to a higher dimensional space (Fig. 3) (Meyer, 2010). The Radial Basis Function kernel was adopted in this study (Cortez, 2010):

$$K(x, x') = \exp(-\gamma \|x - x'\|^2), \quad \gamma > 0 \quad [1]$$

The performance is affected by a penalty parameters, C, the width of the  $\epsilon$ -insensitive zone and the kernel parameter,  $\gamma$  (Cortez, 2010).

The k-Nearest Neighbour (Hechenbichler and Schliep, 2004) is a simple supervised learning algorithm that can be used in classification and regression problems. In classification problems an instance query is classified according to its neighbours classes (Fig. 4). The class in majority among the nearest neighbours is attributed to the query instance. In regression problems the property value for the instance query is obtained as the average of the weighted values of the k-nearest neighbours. This implies the calculation of the distance between the target and its neighbours in the multidimensional space. Generally, the weights are attributed according to the distance. The closest neighbours are more weighted than the more distant neighbours.

The training process of this study used 2/3 of the total dataset and included the optimization of the parameters involved in the different techniques (H,  $\gamma$  and k). It was used a grid search for the number of hidden nodes H [0, 2, 4, 6, ..., 20], the parameter of the kernel  $\gamma$  [ $2^{-15}$ ,  $2^{-13}$ , ...,  $2^3$ ] and the number of nearest neighbours k [2,3,4,...,12]. To access the predictive performance of the models, the 5-fold cross-validation (Efron, 1993) was used, where the data was divided into 5 partitions of equal size.

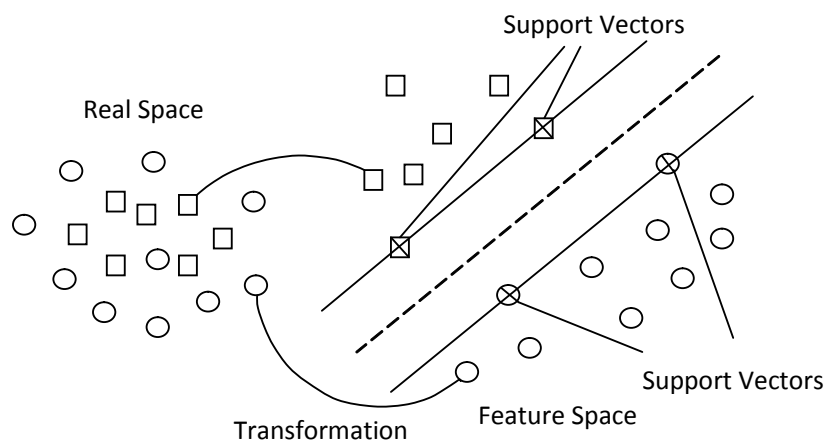


Fig. 3. Example of a SVM transformation

Rys. 3. Przykład transformowania w modelu maszyny wektorów nośnych (MWN)

Table 1 / Tablica 1

FA CHEMICAL COMPOSITION (%) AND BLAINE SPECIFIC SURFACE, m<sup>2</sup>/kgPOPIÓŁ LOTNY: SKŁAD CHEMICZNY (%) I POWIERZCHNIA WŁAŚCIWA BLAINE'A, m<sup>2</sup>/kg

	Kim et al.	Naproux							Camões
SiO <sub>2</sub>	55.1	56.7	54.5	55.1	55.8	56.6	57.3	58.1	50.3
Al <sub>2</sub> O <sub>3</sub>	34.9	27.1	28.2	28.1	27.9	27.3	27.0	26.5	26.8
Fe <sub>2</sub> O <sub>3</sub>	3.7	5.8	5.9	5.5	5.1	5.8	5.8	6.1	6.3
CaO	3.6	2.9	3.2	3.0	2.8	2.9	2.9	3.0	5.4
MgO	1.2	0.6	0.9	0.8	0.8	0.7	0.6	0.6	1.6
K <sub>2</sub> O	0.7	1.9	2.3	2.1	2.0	1.8	1.6	1.5	1.9
Na <sub>2</sub> O	0.4	0.1	0.3	0.2	0.2	0.1	0.1	0.0	0.6
SO <sub>3</sub>	-	0.6	0.3	0.4	0.5	0.6	0.7	0.8	0.6
LOI, Strata prażenia, %	6.8	2.5	3.5	3.2	2.9	2.5	2.1	1.8	7.4
Blaine, Powierzchnia właściwa Blaine'a, m <sup>2</sup> /kg	332	680	667	569	450	310	232	158	388

ru danych na dwie klasy lub kategorie. Aby to zrobić wprowadza się hiperpłaszczyznę w przestrzeni wielowymiarowej, która dzieli dane na zbiory tej samej kategorii. Optymalna hiperpłaszczyzna podziału między dwoma klasami powstaje przez maksymalizację marginesu między najbliższymi punktami obu klas (Meyer, 2010). Punkty danych leżące w obszarach prób są nazywane wektorami nośnymi i optymalny klasyfikator hiperpłaszczyznowy leży pośrodku hiperpłaszczyzn marginesowych. Dla zbiorów danych, które nie są liniowo separabilne wprowadza się zmienne zwisające dla pomniejszenia wpływu tych punktów na rozdzielenie zbiorów tej samej kategorii (Meyer, 2010). Jeśli nie można znaleźć klasyfikatora liniowego wówczas stosuje się transformację do wyższej wymiarowej przestrzeni przez zastosowanie funkcji jądrowych co jest pokazane na rysunku 3 (Meyer, 2010). W niniejszej pracy zastosowano radialną funkcję bazową w postaci (Cortez, 2010):

$$K(x, x') = \exp\left(-\gamma \|x - x'\|^2\right), \quad \gamma > 0 \quad [1]$$

Na skuteczność transformacji wpływ mają: pojemność C, szerokość między hiperpłaszczyznami marginesowymi (parametr  $\epsilon$ ) i parametr  $\gamma$  funkcji jądrowej (Cortez, 2010).

Klasyfikator k najbliższych sąsiadów (Hechenbichler i Schliep, 2004) jest prostym algorytmem uczenia nadzorowanego, który może być użyty w zagadnieniach klasyfikacji i regresji. W zagadnieniach klasyfikacji przykładowy obiekt jest klasyfikowany stosownie do klas jego sąsiadów co jest pokazane na rysunku 4. Klasa większości pośród najbliższych sąsiadów jest klasą, do której jest przypisywany przykładowy obiekt. W zagadnieniach regresji wartość właściwości przykładowego obiektu jest obliczana jako średnia ważonych wartości k najbliższych sąsiadów. Pociąga to za sobą obliczenie odległości między obiektem i jego sąsiadami w wielowymiarowej przestrzeni. Na ogół wagi są przypisywane stosownie do odległości. Najbliżsi sąsiedzi mają większe wagi niż sąsiedzi dalsi.

W procesie uczenia w niniejszych badaniach użyto 2/3 dostępnych przykładów i osiągnięto optymalizację parametrów używa-

The testing process used the remaining part of the total dataset and the best parameters (H,  $\gamma$  and k) of the training phase.

To evaluate the performance of the regression models the coefficient of determination ( $R^2$ ), the mean absolute deviation (MAD) and the root mean squared error (RMSE) were used. The last two metrics are given by:

$$MAD = \frac{1}{N} \times \sum_{i=1}^N |y_i - \hat{y}_i| \quad [2]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad [3]$$

where N denotes the number of examples,  $y_i$  the real value and  $\hat{y}_i$  the value estimated by the model.

The complexity of the models imposes not only the analysis of the metrics but also careful interpretation of the results. To help this interpretation it is also important to evaluate the relative importance of the input parameters in the model. Therefore, it is necessary to carry on a sensibility analysis (Kewley *et al.*, 2000). This analysis is applied after the training phase and allows analysing the response of the model when the input parameters are changed from its minimum to its maximum value. During the process each parameter is changed while the others remain with their middle values. When the parameter is relevant it is obtained a high variance at the model output. A higher variance means a greater importance.

#### 4. Materials and concrete compositions

The data used in this study was collected from Kim *et al.*, 1992, Naproux, 1994, and Camões, 2002, where the used mix-designs as other details, namely concerning the production and behaviour of fresh concrete, can be followed. All the compositions were produced with cement, sand as fine aggregate, coarse aggregate and water. Air entraining agents were not used and some mixtures

Table 2 / Tablica 2

STATISTICAL DESCRIPTION OF THE INPUT AND THE OUTPUT PARAMETERS (TRAINING SET)

OPIS STATYSTYCZNY PARAMETRÓW WEJŚCIA I WYJŚCIA (ZBIÓR UCZĄCY)

Parameters Parametry	Min.	Mean Średnia	Max.	Standard deviation Odchylenie standardowe	Coefficient of variation Współczynnik zmienności, %
Inputs/Wejścia					
FA/Popiół lotny	0	89.12	360	83.43	93.61
S <sub>i</sub> O <sub>2</sub>	0	51.39	58.10	11.70	22.76
Al <sub>2</sub> O <sub>3</sub>	0	30.41	34.90	7.76	25.52
Fe <sub>2</sub> O <sub>3</sub>	0	4.404	6.320	1.53	34.78
CaO	0	3.825	5.425	1.21	31.63
LOI, Strata prażenia, %	0	6.203	7.440	1.95	31.38
Blaine, Powierzchnia właściwa Blaine'a, m <sup>2</sup> /kg	0	341.8	680	103.55	30.29
SF/Pk	0	0.5207	44	4.77	916.49
TCM, kg/m <sup>3</sup> Suma spoiw SW	271	460.2	600	91.67	19.92
W/TCM W/SW	0.25	0.3979	0.7343	0.11	28.86
ssa, kd, kg/m <sup>3</sup>	223	622.6	788	143.91	23.11
ca, kg, kg/m <sup>3</sup>	936	1057	1256	91.01	8.61
HRWRA	0	0.6674	5	0.99	148.91
Age, days Wiek próbki, dni	3	88.12	365	107.82	122.35
Output / Wyjście					
f <sub>c</sub> , MPa Wytrzymałość na ściskanie, MPa	7.043	47.604	96.8	19.69	41.37

nych w różnych metodach (H,  $\gamma$  oraz k). Zastosowano wyczerpujące poszukiwanie dla liczby ukrytych węzłów H [0, 2, 4, 6, ..., 20], parametru funkcji jądrowej  $\gamma$  [2<sup>-15</sup>, 2<sup>-13</sup>, ..., 2<sup>3</sup>] oraz liczby najbliższych sąsiadów k [2, 3, 4, ..., 12]. Do otrzymania dobrej efektywności prognozowania wyżej omówionych modeli zastosowano sprawdzanie krzyżowe, przy czym początkowy zbiór przykładów podzielono na 5 równych części.

W procesie testowania dokładności klasyfikacji użyto pozostałej części zbioru dostępnych przykładów oraz najlepszych parametrów (H,  $\gamma$  oraz k) z fazy uczenia.

Do oszacowania efektywności modeli regresyjnych zastosowano współczynnik determinacji (R<sup>2</sup>), średniego odchylenia bezwzględnego (ŚOB) i pierwiastka średniego błędu kwadratowego (ŚBK). Dwie ostatnie statystyczne miary błędu dane są wzorami:

$$MAD = \frac{1}{N} \times \sum_{i=1}^N |y_i - \hat{y}_i| \quad [2]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad [3]$$

gdzie: N oznacza liczbę obiektów,  $y_i$  wartość rzeczywistą, a  $\hat{y}_i$  wartość prognozowaną przez model.

were made including FA and superplasticizer. Naproux, 1994, has used 7 different types of FA because the main goal of his study was the effect on concrete properties by reducing FA particle size. He used FA as received and enhanced their properties by grinding or sieving.

In Table 1 the FA chemical compositions used in the different concretes are presented. Compressive strength were measured on cylindrical specimens with the height to diameter ratio of 2.

## 5. Data mining used

As aforementioned the compressive strength results were gathered from Kim *et al.*, 1992, and Naproux, 1994, and compiled by Waller *et al.*, 1997. Results from Camões, 2002, were also used. Kim *et al.*, 1992, have performed 26 different concrete compositions and have tested 24 at 6 ages and 2 at 3 ages. Naproux, 1994, have performed 13 concrete compositions and have tested all of them at 3 different ages. Camões, 2002, have performed 11 compositions and have tested all at 6 distinct ages.

The database analysed here was composed of 255 records being 150 extracted from Kim *et al.*, 1992, 39 from Naproux, 1994, and 64 from Camões, 2002. Instead of using only one DM model, one intends to use several models to compare their performances and to include the quantification of the importance of the different



Table 3 / Tablica 3

STATISTICAL DESCRIPTION OF THE INPUT AND THE OUTPUT PARAMETERS (TESTING SET)

OPIS STATYSTYCZNY PARAMETRÓW WEJŚCIA I WYJŚCIA (ZBIÓR TESTUJĄCY)

Parameters Parametry	Min.	Mean Średnia	Max.	Standard Deviation Odchylenie standardowe	Coefficient of variation, Współczynnik zmienności, %
Inputs/Wejścia					
FA/Popiół lotny	0	86.8	360	83.93	96.69
SiO <sub>2</sub>	0	51.4	58.1	11.77	22.90
Al <sub>2</sub> O <sub>3</sub>	0	30.43	34.9	7.80	25.64
Fe <sub>2</sub> O <sub>3</sub>	0	4.393	6.320	1.53	34.91
CaO	0	3.816	5.425	1.21	31.74
LOI, Strata prażenia, %	0	6.196	7.440	1.96	31.57
Blaine, Powierzchnia właściwa Blaine'a, m <sup>2</sup> /kg	0	341.6	680	104.11	30.48
SF/Pk	0	0.5238	44	4.80	916.52
TCM, Suma spoiw SW, kg/m <sup>3</sup>	271	461.1	600	92.03	19.96
W/TCM [W/SW]	0.25	0.3973	0.7343	0.12	29.14
Ssa, kd, kg/m <sup>3</sup>	223	623.1	788	144.77	23.24
ca, kg, kg/m <sup>3</sup>	936	1056	1256	90.96	8.62
HRWRA	0	0.6663	5	1.00	150.20
Age, days Wiek próbki, dni	3	84.82	365	109.26	128.81
Output/Wyjście					
f <sub>c</sub> , Wytrzymałość na ściskanie, MPa	7.20	47.54	93.60	20.47	43.07

Złożoność modeli narzuca nie tylko analizę wyników obliczeń według powyższych wzorów, ale także ostrożną interpretację otrzymanych wyników. Aby ułatwić interpretację wyników ważną jest ocena względnej cenności parametrów wejściowych modelu. Dlatego konieczna staje się analiza wrażliwości (Kewley *et al.*, 2000). Analiza ta jest wykonywana po fazie uczenia i pozwala na analizę odpowiedzi modelu, gdy parametry wejściowe są zmieniane od minimalnej do maksymalnej wartości. W analizie wrażliwości zmieniana jest wartość każdego parametru przy niezmiennych średnich wartościach pozostałych parametrów. Jeśli parametr jest istotny dla badanego modelu, to otrzymuje się dużą wartość wariancji wyjścia modelu. Duża wartość wariancji oznacza większą wrażliwość modelu na wartość badanego parametru.

#### 4. Materiały i składy betonu

Dane wykorzystane w niniejszej pracy są wynikami badań betonów autorstwa Kima i in., 1992, Naprouxa, 1994, i Camõesa, 2002, którzy w badaniach uwzględnili również warunki produkcyjne i zachowanie się świeżego betonu.

Wszystkie mieszanki zawierały cement, piasek jako kruszywo drobne, kruszywo grube i wodę. Domieszki napowietrzające nie były stosowane, a kilka mieszanin sporządzono z popiołem lotnym i superplastyfikatorem.

components of the concrete mixture, namely the chemical and physical characteristics of FA.

In order to simplified the model the amount of input parameters should be minimizing as possible. Therefore, among the known physical-chemical properties of FA presented in Table 1, for DM input parameters were chosen only those known to affect the compressive strength of concrete, as mentioned in point 2. Fur-

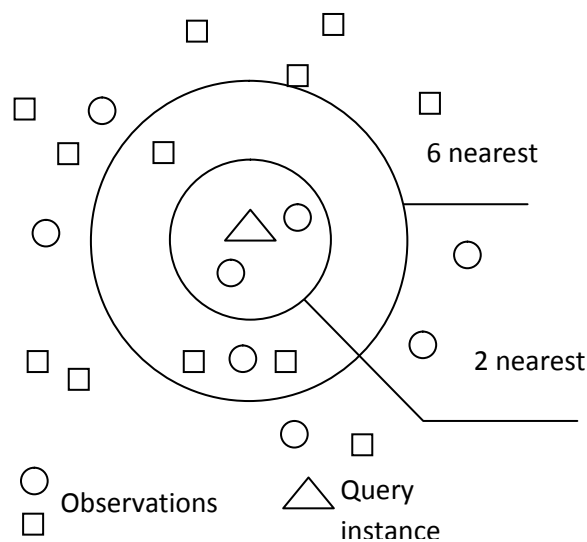


Fig. 4. Example of k-nearest neighbours

Rys. 4. Przykład klasyfikacji według modelu k – najbliższych sąsiadów

Naproux, 1994, stosował 7 rodzajów popiołu lotnego, ponieważ zamierzał badać wpływ zmniejszenia rozmiaru ziarna popiołu lotnego na właściwości betonu. Zastosował suchy popiół lotny i poprawił jego właściwości przez mielenie i przesiewanie.

W tabelicy 1 podano skład chemiczny popiołu lotnego użytego do sporządzenia różnych mieszanin betonowych.

Wszystkie wyniki wytrzymałości na ściskanie uzyskano na próbkach walcowych o stosunku wysokości do średnicy równym 2.

## 5. Wyniki wykorzystywane w zgłębianiu danych

Jak już wspomniano, wartości wytrzymałości na ściskanie zebrano z prac Kima i in., 1992, i Naprouxa, 1994, oraz wartości zestawione przez Wallera i in., 1997. Wykorzystano również wyniki Camõesa, 2002. Kim i in., 1992, sporządzili 26 betonów o różnym składzie i zbadali 24 z nich po 6 okresach dojrzewania, a dwie po 3 okresach twardnienia. Naproux, 1994, sporządził betony o 13 składach i wszystkie badał po 3 różnych okresach twardnienia. Camões, 2002, sporządził 11 składów i wszystkie badał po 6 różnych okresach.

Baza danych analizowana w niniejszej pracy zawierała 255 rekordów z czego 150 pochodziło z pracy Kima *et al.*, 1992, 39 z pracy Naprouxa, 1994, i 64 z pracy Camõesa, 2002. Do zgłębiania danych użyto kilku modeli w celu porównania ich efektywności i dokonania kwantyfikacji ważności wpływu różnych składników mieszaniny betonowej na jej właściwości, a więc właściwości chemicznych i fizycznych popiołu lotnego.

W celu uproszczenia modelu dążono do możliwie jak największej minimalizacji liczby parametrów wejścia. Dlatego spośród znanych fizycznych i chemicznych właściwości popiołu lotnego zawartych w tabelicy 1, do zgłębiania danych wybrano parametry wejścia tylko takie, których wpływ na wytrzymałość betonu na ściskanie jest znany i podano je w tabelicy 2. Pominęto zawartość  $SO_3$ , bowiem w pracy Kima i in., 1992 brak jest danych związanych z tym czynnikiem.

Dane zostały podzielone na dwa zbiory. Zbiór uczący składał się ze 169 rekordów, a zbiór testujący z 84. Parametrami wejścia są: zastąpienie popiołem lotnym ( jako % sumy wszystkich spoiw) % WS, czyli  $WS = C + Pk + PI =$  cement + pył krzemionkowy + popiół lotny); właściwości popiołu lotnego ( $SiO_2$ ,  $Al_2O_3$ ,  $Fe_2O_3$ , CaO, % straty prażenia i powierzchnia właściwa wg Blaine'a w  $m^2/kg$ ); iloraz zastąpienia przez pył krzemionkowy (Pk - % sumy spoiw), suma materiałów wiążących WS, stosunek woda/WS (W/WS), kruszywo drobne (kd), kruszywo grube (kg), duży dodatek domieszki zmniejszającej zawartość wody, (DZW - % w odniesieniu do zawartości materiałów wiążących WS) i wiek próbek (wiek). Parametrem wyjścia jest wytrzymałość betonu na ściskanie ( $f_c$ ). Tabele 2 i 3 zawierają niektóre dane statystyczne parametrów używanych w zbiorach uczących i testujących. Współczynnik zmienności parametrów dla ciągów uczących i testujących jest bardzo podobny. Oznacza to, że zmienność jest podobna w obu zbiorach da-

thermore, the content of  $SO_3$  was not taken into account because there is a lack of information on this parameter in Kim *et al.*, 1992.

The data were separated in two sets. The training set composed of 169 records and the testing set composed of 84 records. The input parameters were: FA replacement ratio (% of total cementitious material, TCM, i.e.  $TCM = C + FA + SF$ ); FA characteristics ( $SiO_2$ ,  $Al_2O_3$ ,  $Fe_2O_3$ , CaO, LOI, in % and Blaine, in  $m^2/kg$ ); silica fume replacement ratio (SF - % of total cementitious material), total cementitious material (TCM), water/TCM ratio (W/TCM); fine aggregate (ssa), coarse aggregate (ca), water reducing agent (WRA - % of solid content related to TCM) and age of samples (Age). The output parameter was the compressive strength of concrete ( $f_c$ ). Tables 2 and 3 present some statistical data of the parameters used in the training and testing databases, respectively. The coefficient of variation of the parameters for training and testing set is quite similar. This means that the variability is similar in both datasets. The SF has the higher variability while the coarse aggregate (ca) has the lower variability.

## 6. Results and discussion

With the built database the predictive models were trained to forecast the concrete compressive strength. Tables 4 show the errors and the coefficient of determination obtained in the training phase. It can be seen that the ANN and the SVM models have the best predictive capacity with lower errors and higher coefficient of determination.

However, it is also important to verify if the importance given by the models to the input variables is according to the experience. The importances for all models are presented in Table 5.

As it was earlier explained the models are based on different algorithms. Therefore, the importance attributed to the input parameters differs from model to model.

It seems that the RT model should not be considered because of the null importance attributed to the majority of the input variables. The MR model also seems to be not applicable since it concentrates almost half of the total importance on only two parameters ( $SiO_2$  and  $Al_2O_3$ ) and does not consider a reasonable importance to the age (only 3 % of importance).

Considering the three remaining models, ANN model gave a non expectable and an apparently exaggerated importance to CaO content (31 %), only the SVM and k-NN attributes reasonable importance to age and only the SVM model attributes a significant importance to W/TCM ratio (17 %). Furthermore, the k-NN model attribute null influence to two parameters ( $SiO_2$  and SF) and only the SVM model attributes a significant importance to FA content.

Thus, analysing the results and comparing them with the experience, i.e. with what is expected, it seems that the SVM was the best model for the tested data, showing sensitivity to parameters known to affect the compressive strength of concrete.

Table 4 / Tablica 4

MEAN VALUES OF THE METRICS OBTAINED IN THE TRAINING PHASE  
ŚREDNIE WARTOŚCI STATYSTYCZNYCH MIAR BŁĘDÓW UZYSKANE W FAZIE UCZENIA

	RT	MR	ANN	SVM	k-NN
MAD	7.486	7.026	4.380	5.610	7.722
RMSE	9.354	8.418	7.414	7.435	9.861
R <sup>2</sup>	0.776	0.817	0.867	0.858	0.750

nych. Zmienność Pk jest duża, podczas gdy kruszywo grube (kg) ma niższą zmienność.

## 6. Wyniki i ich omówienie

Wykorzystując zbudowaną bazę danych modele prognostyczne uczono przewidywania wytrzymałości betonu na ściskanie. Tablica 4 pokazuje błędy i współczynnik determinacji uzyskane podczas uczenia. Można zauważyć, że sztuczna sieć neuronowa i maszyna wektorów nośnych mają najlepszą zdolność prognozowania z niższymi wartościami błędów i wyższą wartością współczynnika determinacji. Ale należy również zauważyć, że cenność danej wejściowej dla modelu jest zgodna z doświadczeniem. Cenność każdej zmiennej wejściowej w poszczególnych modelach jest pokazana w tablicy 5.

Jak już wspomniano w modelach stosuje się różne algorytmy obliczeniowe. Dlatego cenność danej wejściowej dla modelu zmienia się w zależności od modelu.

Wydaje się, że model DR (drzewa regresji) nie powinien być brany pod uwagę, ze względu na zerową cenność większości danych wejściowych. Model regresji wielorakiej (RW) także wydaje się nieprzydatny, gdyż prawie połowa cenności zmiennych wej-

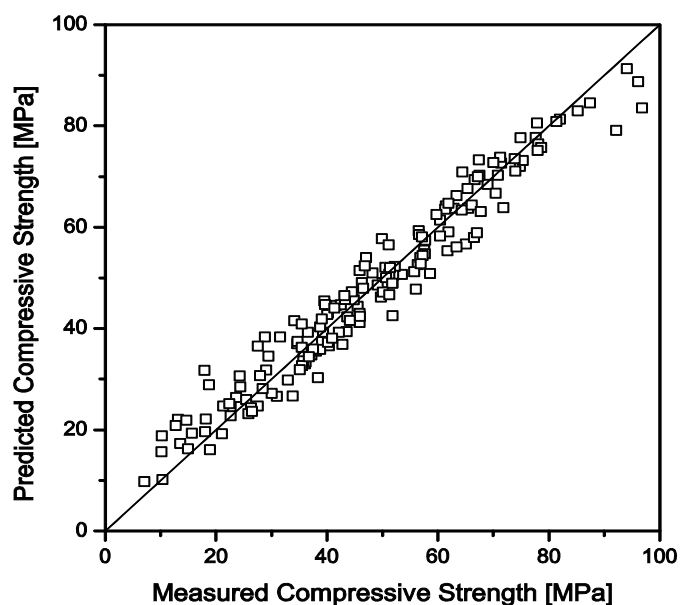


Fig. 5. Performance of the SVM model using the training data set

Rys. 5. Efektywność modelu maszyny wektorów nośnych w fazie uczenia

Table 5 / Tablica 5

IMPORTANCE OF THE INPUT VARIABLES IN THE EVALUATION OF  $f_c$  %  
CENNOŚĆ ZMIENNEJ WEJŚCIOWEJ W PROGNOZOWANIU WYTRZYMAŁOŚCI BETONU NA ŚCISKANIE  $f_c$  %, PRZY UŻYCIU RÓŻNYCH MODELI

	RT	MR	ANN	SVM	k-NN
FA [PI]	0	0.33	1.51	19.49	2.68
SiO <sub>2</sub>	0	22.21	5.19	1.06	0
Al <sub>2</sub> O <sub>3</sub>	0	31.24	10.64	1.82	16.31
Fe <sub>2</sub> O <sub>3</sub>	0	0.92	5.07	1.93	4.55
CaO	0	9.41	31.33	1.01	1.17
LOI	0	11.56	4.82	0.79	17.78
Blaine	0	0.7	5.98	1.09	13.98
SF [Pk]	0	0.09	2.65	0.98	0
TCM [SW]	19.92	9.93	1.22	17.16	6.07
W/TCM [W/SW]	0	0	3.58	16.98	9.73
SSA	0	6.15	1.71	0.37	1.49
CA	9.73	1.68	4.59	1.35	6.12
HRWRA [DDRW]	37.16	2.29	14.97	1.05	2.63
Age	33.19	3.49	6.74	34.92	17.49

It is also important to mention that, according to the SVM model, the influence of FA characteristic parameters in the compressive strength of concrete proved to be marginal. Even the LOI or the Blaine reveals little influence (0.8 % and 1.1 % respectively).

Figures 5 and 6 show the comparison between the measured and predicted concrete compressive strengths for the SVM model using the training and the testing set, respectively. These figures confirm the good predictive capacity of the SVM model.

## 7. Conclusions

The data mining techniques have the capacity of learning with examples. In this study several data mining techniques were used to predict the 3, 7, 28, 56, 90 and 180 days compressive strength of concrete mixes containing different content of FA, replacing cement. The training phase indicates the ANN and the SVM as the best predictive capacity models. However, the importance attributed by the ANN model to the input parameters does not agree with the experience. On the contrary, the SVM model has not this shortcoming, because it demonstrates sensitivity to parameters known to affect the compressive strength of concrete. This sensitivity associated to predictions very close to the experimental results makes the SVM model the best one.

According to the SVM model, the influence of FA physical-chemical properties in the analysed data concerning compressive strength of concrete seems to be marginal.

Based on the analysed data the compressive strength of concrete containing fly ash can be accurately predicted using SVM without spending much time.

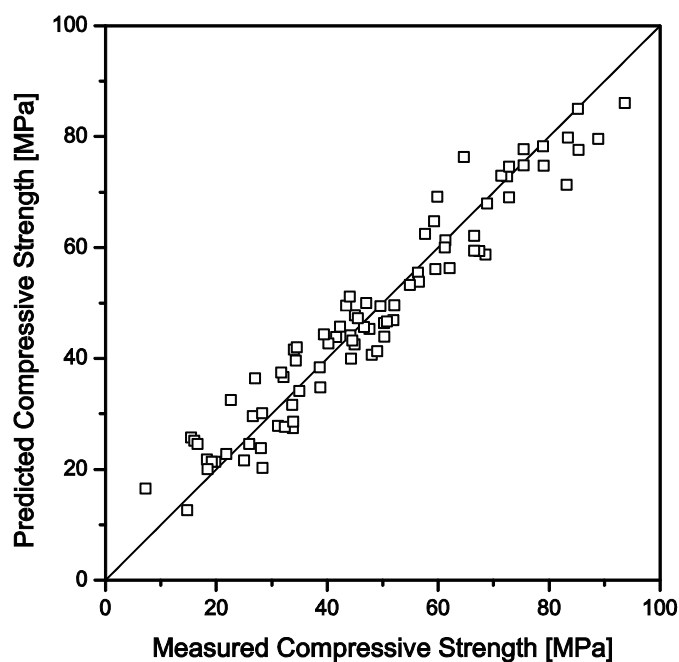


Fig. 6. Performance of the SVM model using the testing data set

Rys. 6. Efektywność modelu maszyny wektorów nośnych w fazie testowania

ściowych przypada na dwie zmienne ( $\text{SiO}_2$  i  $\text{Al}_2\text{O}_3$ ) i nie uwzględnia ważności wpływu wieku próbki na wytrzymałość na ściskanie (tylko 3% cenności tej zmiennej).

Rozważając trzy pozostałe modele można zauważyć, że model sztucznych sieci neuronowych daje nieprzewidywalną i oczywiście wyolbrzymioną cenność zawartości CaO (31%), tylko model maszyny wektorów nośnych i k najbliższych sąsiadów przypisują racjonalną cenność wiekowi próbki i tylko model maszyny wektorów nośnych przypisuje istotną cenność stosunkowi W/Ws (17%). Dalej model k najbliższych sąsiadów przypisuje zerowy wpływ dwóm parametrom ( $\text{SiO}_2$  i Pk) oraz tylko model maszyny wektorów nośnych przypisuje istotną cenność zawartości popiołu lotnego.

W oparciu o analizę wyników i porównując je z doświadczeniem, czyli z tym co jest oczekiwane, wydaje się, że model maszyny wektorów nośnych jest najlepszym modelem dla zbadanych danych, wykazującym wrażliwość na parametry znane jako posiadające istotny wpływ na wytrzymałość betonu na ściskanie.

Należy również zauważyć, że zgodnie z modelem maszyny wektorów nośnych wpływ właściwości popiołów lotnych na wytrzymałość betonu na ściskanie jest niewielki. Nawet strata prażenia lub powierzchnia właściwa Blaine'a wykazują mały wpływ (odpowiednio 0.8 % i 1.1 %).

Na rysunkach 5 i 6 pokazano porównanie między zmierzonymi i prognozowanymi wartościami wytrzymałości betonu na ściskanie w przypadku modelu maszyny wektorów nośnych, odpowiednio dla ciągu uczącego i testującego. Rysunki te potwierdzają przydatność modelu maszyny wektorów nośnych w prognozowaniu.

## Acknowledgements

This study has been carried out under the framework of the strategic plan (2011-2013) of Territory, Environment and Construction Centre (C-TAC/UM), PEst-OE/ECI/UI4047/2011, approved by the Portuguese Foundation for Science and Technology (FCT).

## Literatura / References

1. ACI Committee 232, Use of fly ash in concrete. ACI committee report 232.2R-96. ACI, April, p. 34 (1996).
2. ACI Committee 363, State-of-the-art report on high strength concrete. ACI committee report 363R-92. ACI, September, p. 55 (1992).
3. J. L. Alonso, K. Wesche, Characterization of fly ash. Fly ash in concrete: properties and performance, Report of technical committee 67-FAB – use of fly ash in building (K. Wesche ed.) RILEM, E&FN SPON, pp. 3-23 (1991).
4. A. Camões, (2002). High performance concrete incorporating fly ash, PhD Thesis, University of Minho, p. 456 (2002).
5. B.-T. Chen, T.-P. Chang, J.-Y. Shih, J.-J. Wang, Estimation of exposed temperature for fire-damaged concrete using support vector machine. *Computational Materials Science*, **44**, 3, pp. 913-920 (2009).
6. C. Cortes, V. Vapnik, Support vector networks. *Machine Learning*. Kluwer Academic Publishers, **20**, 3, pp. 273-297 (1995).
7. P. Cortez, Data mining with neural networks and support vector machines using the R/miner Tool. *Proceedings of 10th Industrial Conference on Data Mining, Advances in Data Mining, Applications and theoretical aspects*. (P. Perner (ed.)). *Lecture Notes in Computer Science*, Springer, Berlin, Germany, pp. 572-583 (2010).
8. Day, W. Ken, Concrete mix design, quality control and specification. E&FN SPON, p. 350, London 1995.
9. B. Efron, R. Tibshirani, In *An Introduction to the Bootstrap*. Chapman & Hall 1993.
10. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery: an overview. In *Advances in Knowledge Discovery and Data Mining* (Fayyad et al. eds.). AAAI Press / The MIT Press, Cambridge MA, pp. 471-493 (1996).
11. E. Giergiczny, Z. Giergiczny, *Cement Wapno Beton*, **77**, 157 (2010).
12. S. M. Gupta, Support vector machines based modelling of concrete strength. *Engineering and Technology*, Word Academy of Science, **36**, pp. 305-311 (2007).
13. W. Halstead, Use of fly ash in concrete. *Transportation research board publication n° 127*, Washington, p. 66, 1986.
14. S. Haykin, *Neural networks - a comprehensive foundation*. 2nd ed. New Jersey: Prentice-Hall 1999.
15. K. Hechenbichler, K. Schliep, Weighted k-nearest-neighbor techniques and ordinal classification. *Discussion Paper 399*, SFB 386, Ludwig-Maximilians University Munich 2004. Available from: <[http://epub.uni-muenchen.de/archive/00001769/01/paper\\_399.pdf](http://epub.uni-muenchen.de/archive/00001769/01/paper_399.pdf)>. [Accessed 7 August 2010]
16. S. Jalali, A study of factors affecting the mechanical behaviour of lime-fly ash mixtures. *Coventry Polytechnic, PhD Theses*, February, p. 411, 1991.
17. R. C. Joshi, R. P. Lohtia (1997). Fly ash in concrete: production, properties and uses, *Advances in Concrete Technology* (V. M. Malhotra ed.), Volume 2, Gordon and Breach Science Publishers, p. 269, Ottawa, Ontario, Canada 1997.
18. J. Kasperkiewicz, J. Racz, A. Dubrawski, HPC strength prediction using artificial neural network. *Journal of Computing in Civil Engineering*, **9**, 4, pp. 279-284 (1995).
19. R. Kewley, M. Embrechts, C. Brenemam, (2000). Data strip mining for the virtual design of pharmaceuticals with neural networks. *IEEE Transactions on Neural Networks*, **11**, 3, pp. 668-679 (2000).
20. J. I. Kim, D. K. Kim, Application of neural networks for estimation of concrete strength. *KSCE Journal of Civil Engineering*, **6**, 4, pp. 429-438 (2002).

## 7. Wnioski

Techniki zgłębiania danych mają zdolność uczenia się przy użyciu przykładów. W niniejszej pracy wybrane techniki zgłębiania danych zastosowano do prognozowania wartości wytrzymałości betonu na ściskanie w funkcji zmian składników spoiwowych, w których kilka procent cementu zastąpiono popiołem lotnym. Badania przeprowadzono na próbkach betonu po 3, 7, 28, 56, 90 i 180 dniach dojrzewania. Faza uczenia wskazuje, że sztuczne sieci neuronowe oraz maszyna wektorów nośnych są modelami o najlepszej zdolności uczenia się. Jednakowoż cenność parametrów wejściowych przypisana przez model sztucznych sieci neuronowych nie zgadza się z praktyką. Natomiast maszyna wektorów nośnych jest pozbawiona tej wady, gdyż wykazuje wrażliwość na parametry znane jako znacznie wpływające na wytrzymałość betonu na ściskanie. Wrażliwość tego modelu na zmienne wejściowe jest bardzo bliska do wyników praktycznych i tym samym model maszyny wektorów nośnych jest najbardziej przydatny do prognozowania wartości wytrzymałości betonu na ściskanie.

Model maszyny wektorów nośnych w obszarze zanalizowanych danych wykazuje mało znaczący wpływ właściwości fizycznych i chemicznych popiołu lotnego na wytrzymałość betonu na ściskanie.

Na gruncie zanalizowanych danych można stwierdzić, że model maszyny wektorów nośnych trafnie prognozuje wartości wytrzymałości na ściskanie betonu zawierającego popiół lotny, bez dużego nakładu czasu na wykonanie obliczeń.

21. J. K. Kim, Y. D. Park, K. Y. Sung, S. G. Lee, The production of high-strength fly ash in Korea. 4th CANMET/ACI International Conference on Fly Ash, Silica Fume, Slag and Natural Pozzolans in Concrete (V.M. Malhotra ed.), Supplementary papers, Istanbul. May 1992.
22. S. Lai, M. Serra, Concrete strength prediction by means of neural network. *Construction and Building Materials*, **11**, 2, 93-98 (1997).
23. V. M. Malhotra, A. A. Ramezani-pour, Fly ash in concrete. 2nd edition, CANMET – Canada Centre for Mineral and Energy Technology, pp. 307, 1994.
24. P. K. Mehta, Influence of fly ash characteristics of the strength of Portland fly ash mixtures. *Cement and concrete research*, Elsevier Science Ltd, Volume 15, Number 4, pp. 669-674, 1985.
25. D. Meyer, Support Vector Machines. The interface to libsvm in package e1071, 2010. Available from: <http://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf>. [Accessed 7 August 2010].
26. S. Mindess, (1994). *Materials selection, proportioning and quality control. High performance concretes and applications* (S.P. Shah and S.H. Ahmad ed.), Edward Arnold, pp. 1-25, London 1994.
27. P. Naproux, Les microcendres (cendres volantes traitées) et leur emploi dans les bétons hydrauliques. PhD Thesis. Institut National de Sciences Appliquées de Toulouse, France 1994.
28. A. M. Neville, (1995). *Properties of concrete*, Fourth edition, Longman Group Limited, p. 844, England 1995.
29. H.-G. Ni, J.-Z. Wang, Prediction of compressive strength of concrete by neural networks. *Cement and Concrete Research* **30**, 1245-1250 (2000).
30. F. Özcan, C. D. Atiş, O. Karahan, E. Uncuoğlu, H. Tanyildizi, Comparison of artificial neural network and fuzzy logic models for predicting of long-term compressive strength of silica fume concrete. *Advances in Engineering Software*, **40**, pp. 856-863 (2009).
31. J. Quinlan, *Induction of Decision Trees*. Machine Learning, Kluwer Academic Publishers **1**, pp. 81-106, 1986.
32. R Development Core Team. (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3. Available from: <http://www.R-project.org>. [Accessed 7 August 2010].
33. Ruan Xiang, Prediction of concrete carbonation depth based on support vector regression. *Third International Symposium on Intelligent Information Technology Application*, iita, vol. 3, pp. 172-175, 2009.
34. M. Saridemir, Prediction of compressive strength of concretes containing metakaolin and silica fume by artificial neural networks. *Advances in Engineering Software*, **40**, pp. 350-355 (2009).
35. R. Swamy, Fly ash and slag: standards and specifications – help or hindrance? *Materials and structures / Matériaux et constructions*, **26**, pp. 600-613 (1993).
36. I. B. Topçu, M. Saridemir, Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic. *Computation Materials Science*, **41**, pp. 305-311 (2008).
37. V. Waller, P. Naproux, F. de Larrard, Contribution des fumées de silice et des cendres volantes silico-alumineuses à la résistance en compression du béton. Quantification. *Bulletin des Laboratoires des Ponts et Chaussées*. 208. Mars-Avril – réf. 4098, pp. 53-65, 1997.
38. K. Wesche, Fly ash in concrete: properties and performance. Report of technical committee 67-FAB – use of fly ash in building (K. Wesche ed.) RILEM, E&FN SPON, pp. 3-23, 1991.
39. D.-S. Yang, S.-K. Park, J.-H. Lee, A prediction on mix proportion factor and strength of concrete using neural network. *KSCE Journal of Civil Engineering*, **7** (5), pp. 525-536 (2003).
40. M. H. F. Zarandi, I. B. Türksen, J. Sobhani, A. A. Ramezani-pour, Fuzzy polynomial neural networks for approximation of the compressive strength of concrete. *Applied Soft Computing*, **8**, 488-498, (2008).
41. L. Zhitao, H. Hongming, Z. Shengli, Research on support vector machine's prediction of concrete carbonization. *International Seminar on Business and Information Management*, isbim, vol. 1, pp. 319-322, 2008.