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

Francisco F.Martins¹ and Aires Camões²

¹ Associate Professor, C-TAC/UM, Territory, Environment and Construction Research Centre, University of Minho, Guimarães, Portugal. Email: ffm@civil.uminho.pt

² Assistant Professor, C-TAC/UM, Territory, Environment and Construction Research Centre, University of Minho, Guimarães, Portugal. Email: aires@civil.uminho.pt.

Number of words: 5864
Number of tables: 5
Number of figures: 6
Abstract

The concrete compressive strength is the most used mechanical property in the design of concrete structures. Therefore, the use of rational models to its prediction, to simulate the effects of its different constituents and its properties can play an important role in the achievement of the safety-economy required. Models to forecast the concrete compressive strength have already been presented before by some researchers. However, the comparison of different rational models and the application of models to predict the importance of the different constituents in the concrete behaviour have not yet been approached. Therefore, developing these models will be necessary namely to take into account the quality, i.e. the activity, of the most used mineral addition in concrete: fly ash. This study compared different Data Mining techniques to predict the compressive strength of fly ash concrete along time. The presented models are able to learn the complex relationships between several variables like the uniaxial compressive strength, the different concrete compounds and its mix design, the different properties of the fly ash used and the relative influence of its.

Keywords: concrete strength, fly ash, data mining, artificial neural networks, support vector machines
1. Introduction

The concrete compressive strength is the most used mechanical property in the design of concrete structures. The prediction of concrete strength before its placement allows engineers to improve planning and quality control. Moreover, a well defined concrete strength prediction can save time, accelerating the overall construction, mainly in large constructions with a lot of concreting stages. However, environmental concerning and the need of higher performance concretes have lead to more complex mixes than the traditional ones. Nowadays the recycled industrial wastes available leads to the quasi-systematic use of mineral addition in concrete, namely fly ash from electrical power plants. Therefore, the factors that affect the concrete compressive strength have increased in number and complexity leading to non linear relationships. These relationships are usually empirical ones 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’s 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: 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 didn’t take into
account the quality of the FA used. In fact, to the best of our knowledge, the physical and the chemical parameters of the FA have never been used before in these predictions.

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

This paper begins to present a brief description of FA properties effect on concrete strength. Then it is presented a definition of the data mining techniques and its application on prediction of concrete behaviour. It is also explained how to evaluate the different algorithms of DM. Finally the results, discussion and conclusions are presented.

2. Influence of FA characteristics 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 Ramezanianpour, 1994, ACI Committee 232, 1996, Joshi and Lohtia, 1997 and Camões, 2002). Therefore, in this section we only intend to address the main aspects directly related to the present work, namely the effect of physical and chemical characteristics of FA on concrete performance, mainly on compressive strength.
Unless physical and 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. Certain FA has also irregular or angular particles. Most of the particles have a diameter of less than 1 µm and 150 µm, and can be thinner or thicker than cement ones. The average diameter of FA particles is usually similar to cement ones, between 7 µm and 12 µm (Joshi and Lohtia, 1997). The Blaine specific surface ranges frequently from 250 m²/kg to 550 m²/kg (Alonso and Wesche, 1991).

Broadly, the physical characteristics of FA have an appreciable variation with respect to its origin. According to Malhotra and Ramezianpour, 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 thinness, as measured by percentage retained on sieve of 45 µm and Blaine specific surface.

FA particles larger than 125 µm are very porous. The occurrence of these particles is associated with large amounts of carbon. This unburned material is responsible for the high specific surface typically found in FA. As a result, high levels of carbon imply greater demand for water in concrete containing FA with high loss on ignition (LOI). The carbon content also affects the strength of concrete actions to freeze-thaw: the higher the carbon content of FA, the lower the resistance of concrete (Alonso and Wesche, 1991).
According to these authors, FA should have a particle size and specific surface similar to or lower than the cement to avoid variations in physical properties of concrete particularly in workability. This recommendation is linked to the presence of carbon in FA, avoiding high amounts of particles larger than 125 µm, endowed with high porosity and with greater concentration of carbon particles.

It is commonly accepted that more fineness leads to greater pozzolanic activity. Like most chemical reactions occur more rapidly with increasing fineness of the particles it is expected that the pozzolanic activity of fly ash must be dependent on the area available for reaction (Jalali, 1991, Neville, 1995).

Furthermore, the spherical shape of FA particles is particularly advantageous from the point of view of the water demand and high specific surface indicates that the material exhibits high reactivity with calcium hydroxide (Neville, 1995).

In this context, it is supposed that the higher specific surfaces and the lower LOI the better the activity of FA on concrete will be.

2.2 Chemical properties

The chemical constituents of the majority of FA particles are compounds and crystals of silica, SiO₂, alumina, Al₂O₃, ferric oxide, Fe₂O₃, and lime, CaO. In substantially lower amounts there are other components such as MgO, Na₂O, K₂O, SO₃, MnO and TiO₂. FA also contains carbon particles not consumed in the combustion (Alonso and Wesche, 1991, Malhotra and Ramezanianpour, 1994 and ACI Committee 232, 1996). The alluded four main components of FA record appreciable change, and may have understood values, according to the ACI Committee 232, 1996, between the following: SiO₂ – 35 % to 60 %; Al₂O₃ – 10 % to 30 %; Fe₂O₃ – 4 % to 20 %; CaO – 1 % to 35 %.
The FA pozzolanic activity is closely related to the SiO₂ content, as it is the amorphous silica that will combined with the free lime and water giving resulting in the formation of additional quantities of calcium silicate hydrates. According to Alonso and Wesche, 1991, FA with SiO₂ content less 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 and binder properties and its inclusion in concrete should be avoided.

The role played by the oxides is not consensual. ACI Committee 232, 1996, refers the failure of studies aimed at establishing relations between the amount of oxides - SiO₂, Al₂O₃, Fe₂O₃ - and the performance of the FA. Although the content of SiO₂ appear to be related to the pozzolanic activity, reduced levels of this product do not imply negative effects on the behavior and characteristics of fresh or hardened concrete (Camões, 2002). The establishment of minimum values for the total amount of oxides (SiO₂ + Al₂O₃ + Fe₂O₃) 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 glassy constituents (according to Mindess, 1994, the higher the glass phase is the better FA will be), one should not confuse the reactivity of the glassy phase with increasing resistance, because the development of mechanical characteristics is always achieved by the combined effect of several other factors, such as the fineness and the properties of cement. Malhotra and Ramezanianpour, 1994, mention the existence of a good correlation between the amount of SiO₂ + Al₂O₃ and
the pozzolanic activity at long term. The same authors also reported that the amount of Fe₂O₃ presented in almost FA occurs in the form of hematite and magnetite non-reactive. These aspects should be the reason why researchers have been obtained low correlations between the pozzolanic activity index (measured by the ratio between the compressive strength) and the amount of SiO₂ + Al₂O₃ + Fe₂O₃.

The sulphates, SO₃, may affect the optimum FA content as it can influence its mechanical properties and setting time. A maximum level must be considered to avoid an excess of SO₃ in hardened concrete that can contribute significantly to a worse performance particularly when the concrete is subjected to sulphates attack (Jalali, 1991).

The carbon in FA is the result of incomplete combustion of coal and organic additives used in the process of ash collection. In general, the carbon is not measured directly but using the determination of LOI content. LOI includes, besides the free carbon, losses of combined water and carbon dioxide from the hydrates and carbonates present in the FA. However LOI is assumed, without committing a great error, as approximately equal to the carbon content.

The contribution of the amount of carbon in FA is decisive in the water demand of pastes, mortars and concretes. 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 large specific surface, being able to absorb not only significant amounts of water as well chemical admixtures dissolved during the mixing of concrete, including superplasticizers, air entraining agents or setting time retardants.

According to Alonso and Wesche, 1991, in general, the lower the carbon content the better 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 Ramezanianpour, 1994, concluded that the carbon does not significantly affect the pozzolanic activity index, determined by the relationship between 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 pos-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 researchers (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 for different mixes. 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 ashes. 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, carried on experimental works and 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 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
drawn 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 (Figure 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 (Figure 2).

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. To do so, a hyperplane in a multidimensional space separates the examples
in sets of the same category. The optimal separating hyperplane between the two classes
by 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
hyperplane is at the middle of the margin. The points lying on the wrong side are
weighted down to reduce there influence (Meyer, 2010). When a linear separator cannot
be found there is a transformation via kernel techniques to a higher dimensional space (Figure 3) (Meyer, 2010). The Radial Basis Function kernel was adopted in this study (Cortez, 2010):

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

(1)

The performance is affected by a penalty parameters, C, the width of the ε-insensitive zone and the kernel parameter, γ (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 (Figure 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, γ 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 γ [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 (Effron, 1993) was used, where the data was divided into 5 partitions of equal size. The testing process used the remaining part of the total dataset and the best parameters (H, γ 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:

$$MAD = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

(2)

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

(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 fresh behaviour of concrete, can be seen.
All the compositions were produced with cement, sand as fine aggregate, coarse aggregate and water. Air entraining agents were not used and some mixtures were made including FA and superplasticizer.

Naproux, 1994, has used 7 different types of FA because it was his intention to study the effect on concrete properties derived from reducing FA particle size. So he used brut FA as received and enhanced ones by grinding or sieving.

Observing Table 1 one can see the chemical composition of the FAs used in the different concrete compositions.

All the compressive strength results were measured on cylindrical specimens with a slenderness ratio of 2.

5. Data mining used data
As already mentioned 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 components of the concrete mixture, namely the chemical and physical characteristics of FA.
In order to simplify the model one intends to minimize the amount of input parameters as possible. Therefore, among the known physical and chemical characteristics of FA described in Table 1, one as chosen for DM input parameters only those known to affect the compressive strength of concrete as mentioned in 2. Furthermore, the content of SO$_3$ was not taking 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 are: FA replacement ratio (% of total cementitious material, TCM, i.e. TCM = C + FA + SF); FA characteristics (SiO$_2$, Al$_2$O$_3$, Fe$_2$O$_3$, 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), high rate water reducing agent (HRWRA - % of solid content related to TCM) and age of samples (Age). The output parameter is 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 see if the importance given by the models to the input variables is according to the experience. The importances for all the 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\textsubscript{2} and Al\textsubscript{2}O\textsubscript{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\textsubscript{2} and SF) and only the SVM model attributes a significant importance to FA content.

Thus, analyzing 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, demonstrating sensitivity to parameters known to affect the compressive strength.

It is also important to note 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 shows 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 use to predict the 3, 7, 28, 56, 90 and 180 days compressive strength values of concrete mixes containing several percentages of cement replaced by FA. 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 is not according to the experience. On the contrary, the SVM model has not this shortcoming, because it demonstrates sensitivity to parameters known to affect the compressive strength. 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 and chemical characteristics in the analysed data concerning compressive strength of concrete seems to be marginal.

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

C-TAC/UM is a research unit (UI4047) of the “Fundação para a Ciência e Tecnologia“ (FCT) currently under partial financial support of the strategic project Pest-OE/ECI/UI4047/2011.

References


**Table captions:**

Table 1: FA chemical composition (%) and Blaine specific surface (m²/kg).

Table 2: Statistical of the input and the output parameters (training set).

Table 3: Statistical of the input and the output parameters (testing set).

Table 4: Mean values of the metrics obtained in the training phase.

Table 5: Importance of the input variables in the evaluation of $f_c$ (%).
Table 1: FA chemical composition (%) and Blaine specific surface (m$^2$/kg).

<table>
<thead>
<tr>
<th></th>
<th>Kim et al.</th>
<th>Naproux</th>
<th>Camões</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO$_2$</td>
<td>55.1</td>
<td>56.7</td>
<td>54.5</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>34.9</td>
<td>27.1</td>
<td>28.2</td>
</tr>
<tr>
<td>Fe$_2$O$_3$</td>
<td>3.7</td>
<td>5.8</td>
<td>5.9</td>
</tr>
<tr>
<td>CaO</td>
<td>3.6</td>
<td>2.9</td>
<td>3.2</td>
</tr>
<tr>
<td>MgO</td>
<td>1.2</td>
<td>0.6</td>
<td>0.9</td>
</tr>
<tr>
<td>K$_2$O</td>
<td>0.7</td>
<td>1.9</td>
<td>2.3</td>
</tr>
<tr>
<td>Na$_2$O</td>
<td>0.4</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>SO$_3$</td>
<td>-</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>LOI</td>
<td>6.8</td>
<td>2.5</td>
<td>3.5</td>
</tr>
<tr>
<td>Blaine</td>
<td>332</td>
<td>680</td>
<td>667</td>
</tr>
</tbody>
</table>

Notes: S$\text{iO}_2$, Al$\text{2O}_3$, Fe$\text{2O}_3$, CaO, MgO, K$\text{2O}$, Na$\text{2O}$, SO$_3$, LOI, and Blaine are expressed as percentage and m$^2$/kg, respectively.
Table 2: Statistical of the input and the output parameters (training set).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Min.</th>
<th>Mean</th>
<th>Max.</th>
<th>Standard Deviation</th>
<th>Coefficient of variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td>0</td>
<td>89.12</td>
<td>360</td>
<td>83.43</td>
<td>93.61</td>
</tr>
<tr>
<td>SiO₂</td>
<td>0</td>
<td>51.39</td>
<td>58.10</td>
<td>11.70</td>
<td>22.76</td>
</tr>
<tr>
<td>Al₂O₃</td>
<td>0</td>
<td>30.41</td>
<td>34.90</td>
<td>7.76</td>
<td>25.52</td>
</tr>
<tr>
<td>Fe₂O₃</td>
<td>0</td>
<td>4.404</td>
<td>6.320</td>
<td>1.53</td>
<td>34.78</td>
</tr>
<tr>
<td>CaO</td>
<td>0</td>
<td>3.825</td>
<td>5.425</td>
<td>1.21</td>
<td>31.63</td>
</tr>
<tr>
<td>LOI</td>
<td>0</td>
<td>6.203</td>
<td>7.440</td>
<td>1.95</td>
<td>31.38</td>
</tr>
<tr>
<td>Blaine</td>
<td>0</td>
<td>341.8</td>
<td>680</td>
<td>103.55</td>
<td>30.29</td>
</tr>
<tr>
<td>SF</td>
<td>0</td>
<td>0.5207</td>
<td>44</td>
<td>4.77</td>
<td>916.49</td>
</tr>
<tr>
<td>TCM (kg/m³)</td>
<td>271</td>
<td>460.2</td>
<td>600</td>
<td>91.67</td>
<td>19.92</td>
</tr>
<tr>
<td>W/TCM</td>
<td>0.25</td>
<td>0.3979</td>
<td>0.7343</td>
<td>0.11</td>
<td>28.86</td>
</tr>
<tr>
<td>ssa (kg/m³)</td>
<td>223</td>
<td>622.6</td>
<td>788</td>
<td>143.91</td>
<td>23.11</td>
</tr>
<tr>
<td>ca (kg/m³)</td>
<td>936</td>
<td>1057</td>
<td>1256</td>
<td>91.01</td>
<td>8.61</td>
</tr>
<tr>
<td>HRWRA</td>
<td>0</td>
<td>0.6674</td>
<td>5</td>
<td>0.99</td>
<td>148.91</td>
</tr>
<tr>
<td>Age (days)</td>
<td>3</td>
<td>88.12</td>
<td>365</td>
<td>107.82</td>
<td>122.35</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fₖ (MPa)</td>
<td>7.043</td>
<td>47.604</td>
<td>96.8</td>
<td>19.69</td>
<td>41.37</td>
</tr>
</tbody>
</table>
Table 3: Statistical of the input and the output parameters (testing set).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Min.</th>
<th>Mean</th>
<th>Max.</th>
<th>Standard Deviation</th>
<th>Coefficient of variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inputs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td>0</td>
<td>86.8</td>
<td>360</td>
<td>83.93</td>
<td>96.69</td>
</tr>
<tr>
<td>SiO2</td>
<td>0</td>
<td>51.4</td>
<td>58.1</td>
<td>11.77</td>
<td>22.90</td>
</tr>
<tr>
<td>Al2O3</td>
<td>0</td>
<td>30.43</td>
<td>34.9</td>
<td>7.80</td>
<td>25.64</td>
</tr>
<tr>
<td>Fe2O3</td>
<td>0</td>
<td>4.393</td>
<td>6.320</td>
<td>1.53</td>
<td>34.91</td>
</tr>
<tr>
<td>CaO</td>
<td>0</td>
<td>3.816</td>
<td>5.425</td>
<td>1.21</td>
<td>31.74</td>
</tr>
<tr>
<td>LOI</td>
<td>0</td>
<td>6.196</td>
<td>7.440</td>
<td>1.96</td>
<td>31.57</td>
</tr>
<tr>
<td>Blaine</td>
<td>0</td>
<td>341.6</td>
<td>680</td>
<td>104.11</td>
<td>30.48</td>
</tr>
<tr>
<td>SF</td>
<td>0</td>
<td>0.5238</td>
<td>44</td>
<td>4.80</td>
<td>916.52</td>
</tr>
<tr>
<td>TCM (kg/m³)</td>
<td>271</td>
<td>461.1</td>
<td>600</td>
<td>92.03</td>
<td>19.96</td>
</tr>
<tr>
<td>W/TCM</td>
<td>0.25</td>
<td>0.3973</td>
<td>0.7343</td>
<td>0.12</td>
<td>29.14</td>
</tr>
<tr>
<td>ssa (kg/m³)</td>
<td>223</td>
<td>623.1</td>
<td>788</td>
<td>144.77</td>
<td>23.24</td>
</tr>
<tr>
<td>ca (kg/m³)</td>
<td>936</td>
<td>1056</td>
<td>1256</td>
<td>90.96</td>
<td>8.62</td>
</tr>
<tr>
<td>HRWRA</td>
<td>0</td>
<td>0.6663</td>
<td>5</td>
<td>1.00</td>
<td>150.20</td>
</tr>
<tr>
<td>Age (days)</td>
<td>3</td>
<td>84.82</td>
<td>365</td>
<td>109.26</td>
<td>128.81</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fc (MPa)</td>
<td>7.20</td>
<td>47.54</td>
<td>93.60</td>
<td>20.47</td>
<td>43.07</td>
</tr>
</tbody>
</table>
Table 4: Mean values of the metrics obtained in the training phase.

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>MR</th>
<th>ANN</th>
<th>SVM</th>
<th>k-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAD</td>
<td>7.486</td>
<td>7.026</td>
<td>4.380</td>
<td>5.610</td>
<td>7.722</td>
</tr>
<tr>
<td>R²</td>
<td>0.776</td>
<td>0.817</td>
<td>0.867</td>
<td>0.858</td>
<td>0.750</td>
</tr>
</tbody>
</table>

Table 5: Importance of the input variables in the evaluation of $f_c$ (%).

<table>
<thead>
<tr>
<th></th>
<th>RT</th>
<th>MR</th>
<th>ANN</th>
<th>SVM</th>
<th>k-NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>FA</td>
<td>0</td>
<td>0.33</td>
<td>1.51</td>
<td>19.49</td>
<td>2.68</td>
</tr>
<tr>
<td>SiO2</td>
<td>0</td>
<td>22.21</td>
<td>5.19</td>
<td>1.06</td>
<td>0</td>
</tr>
<tr>
<td>Al2O3</td>
<td>0</td>
<td>31.24</td>
<td>10.64</td>
<td>1.82</td>
<td>16.31</td>
</tr>
<tr>
<td>Fe2O3</td>
<td>0</td>
<td>0.92</td>
<td>5.07</td>
<td>1.93</td>
<td>4.55</td>
</tr>
<tr>
<td>CaO</td>
<td>0</td>
<td>9.41</td>
<td>31.33</td>
<td>1.01</td>
<td>1.17</td>
</tr>
<tr>
<td>LOI</td>
<td>0</td>
<td>11.56</td>
<td>4.82</td>
<td>0.79</td>
<td>17.78</td>
</tr>
<tr>
<td>Blaine</td>
<td>0</td>
<td>0.7</td>
<td>5.98</td>
<td>1.09</td>
<td>13.98</td>
</tr>
<tr>
<td>SF</td>
<td>0</td>
<td>0.09</td>
<td>2.65</td>
<td>0.98</td>
<td>0</td>
</tr>
<tr>
<td>TCM</td>
<td>19.92</td>
<td>9.93</td>
<td>1.22</td>
<td>17.16</td>
<td>6.07</td>
</tr>
<tr>
<td>W/TCM</td>
<td>0</td>
<td>0</td>
<td>3.58</td>
<td>16.98</td>
<td>9.73</td>
</tr>
<tr>
<td>SSA</td>
<td>0</td>
<td>6.15</td>
<td>1.71</td>
<td>0.37</td>
<td>1.49</td>
</tr>
<tr>
<td>CA</td>
<td>9.73</td>
<td>1.68</td>
<td>4.59</td>
<td>1.35</td>
<td>6.12</td>
</tr>
<tr>
<td>HRWRA</td>
<td>37.16</td>
<td>2.29</td>
<td>14.97</td>
<td>1.05</td>
<td>2.63</td>
</tr>
<tr>
<td>Age</td>
<td>33.19</td>
<td>3.49</td>
<td>6.74</td>
<td>34.92</td>
<td>17.49</td>
</tr>
</tbody>
</table>
Figure captions:

Figure 1: Example of a regression tree.
Figure 2: Example of a multilayer perceptron.
Figure 3: Example of a SVM transformation.
Figure 4: Example of k-nearest neighbours.
Figure 5: Performance of the SVM model using the training data set.
Figure 6: Performance of the SVM model using the testing data set.
Figure 1: Example of a regression tree.

Figure 2: Example of a multilayer perceptron.
Figure 3: Example of a SVM transformation.

Figure 4: Example of k-nearest neighbours.
Figure 5: Performance of the SVM model using the training data set.

Figure 6: Performance of the SVM model using the testing data set.