INSTRUCTIONS FOR PREPARATION AND SUBMISSION OF FULL-PAPERS FOR PUBLICATION IN THE PROCEEDINGS OF XL CILAMCE

Pedro B. A. Moreira

pedrobermoreira@gmail.com
Student, Dept. of Engineering, University Veiga de Almeida
Address, Zip-Code, State/Province, Country

Victor M. da Silva

somebody2@somewhere.com
Assistant Professor, Dept. of Engineering, IBMEC/RJ
Address, Zip-Code, State/Province, Country

Abstract. Compressive strength is the main characteristic of concrete. The correct prediction of this parameter results in cost and time reduction. This work built predictive models for 6 different ages of concrete samples (3, 7, 14, 28, 56, and 100 days). Was used a dataset with 9 variables: compressive strength, age, and 7 ingredients (water, cement, fine aggregate, coarse aggregate, fly ash, blast furnace slag, and superplasticizers). Another 6 variables were added to represent the proportions of the main ingredients in each sample (water/cement, fine aggregate/cement, coarse aggregate/cement, fine aggregate/coarse aggregate, water/coarse aggregate, and water/fine aggregate). The predictive models were developed in R language, using the caret package with the Parallel Random Forest algorithm and repeated cross-validation technique to optimize the parameters. The results were satisfactory and compatible with other studies using the same data set. The most important model, 28 days old, obtained RMSE of 4.717. The 3-day model obtained the best result, RMSE of 3.310. The worst result was the 56-day model, with RMSE of 5.939. The work showed that the compressive strength of concrete can be predicted. The choice of creating a model for each age, instead of using age as a predictor, allowed to get compatible results with the available data at each age. It was a promising alternative since good results were achieved by training with just one algorithm. This work facilitates exploration and new efforts to predict the compressive strength of concrete, it can be replicated using different algorithms or the combination of several.

Keywords: Concrete, Compressive Strength, Machine Learning, Prediction

1   Introduction

Compressive strength is the main characteristic of concrete, measured by tests of international standards that consist of the breaking of specimens. Measurement at 28 days is mandatory and represents the grade of the concrete. Knowing in advance what the result will be obtained for a given age, based on the proportions of its ingredients, is of great interest to concrete manufacturers, construction companies, and civil engineers.

This compressive strength is a nonlinear function of its ingredients and age, making it difficult to establish an analytical formula, although some formulas have already been proposed. Hasan and Kabir [1] proposed a mathematical model to predict from the results of tests of 7 and 14 days, and Kabir et al. [2] from 7 days. However, machine learning techniques can be used to model this characteristic from real sample data, using only the ingredients.

```
# Aqui vale incluir um parágrafo para explicar o que é machine learning,
# e por que é possível obter a resistência a partir apenas dos ingredientes.
```

Many previous studies use the same dataset used by Yeh [3] to predict the compressive strength of concrete. Al-Shamiri AK [4] got good results with the regularized extreme learning machine (RELM) technique, and Hameed and Khalid [5] got even better results with the Artificial Neural Networks and cross-validation technique. This set of samples is so well known that there are many pages on the internet of unpublished studies that use it and have good results, such as Abban [6], Raj [7], Modukuru [8] and Pierobon [9]. At the end of the work, the results found are compared to the works cited here.

Unlike previous studies with this dataset, this work does data preparation differently. The age of the concrete is the most unique feature that contributes to its compressive strength. For this reason, age is treated separately in the machine learning models, creating models for each age group.

2   Materials and Methods

2.1   Materials

```
# Editar, porque não vai ter mais os apêndices
```

The methodology was carried out using RStudio software [10], an integrated virtual environment for code development in *R* [11]. Throughout the process, all code executed was documented in the same order as its execution in Appendix 3, and reference was always made to codes throughout the text. All relevant information related to the operating system and installed packages has been presented in Appendix 1. In addition, an online and open-source repository was created in *Github*, housing all the code used to generate this work, the link was made available in[Appendix 2.

2.2   Reproducibility

```
# Editar, porque não vai ter mais os apêndices
# Definir o que é seed
```

In order to guarantee reproducibility, whenever there was some code that could use probabilistic operations, a *seed* was defined before its execution, ensuring that when run on another machine, with the same version of *R* and the same *seed*, get the same result. The *seeds* can be checked throughout Appendix 3 or directly on *Github*.

2.3   Obtaining the data

The data was downloaded from the University of California Irvine website [12]. In total there are 1030 samples with 9 columns. The samples were renamed and an id column was added to facilitate

data manipulation. The columns were reordered to put the new id column in the first position. The first samples can be viewed in the Table 1.

Table 1. First 6 samples

ID	Cement kg/m^3	B.F.S. kg/m^3	Fly ash kg/m^3	Water kg/m^3	Superp. kg/m^3	C.Aggregate kg/m^3	F.Aggregate kg/m^3	Day	Comp.Str. MPa
1	540.0	0.0	0	162	2.5	1040.0	676.0	28	79.99
2	540.0	0.0	0	162	2.5	1055.0	676.0	28	61.89
3	332.5	142.5	0	228	0.0	932.0	594.0	270	40.27
4	332.5	142.5	0	228	0.0	932.0	594.0	365	41.05
5	198.6	132.4	0	192	0.0	978.4	825.5	360	44.30
6	266.0	114.0	0	228	0.0	932.0	670.0	90	47.03

2.4 Data preparation

The preparation of the data consisted of transforming the sample set in order to maintain only relevant data for the subsequent studies. Data that were considered irrelevant or that had the potential to add undesirable noise to the analysis were removed. In addition, the relevant data has been transformed to better fit the studies in the next steps.

2.4.1 Initial data cleaning

Initially, there were 25 duplicate samples that were removed, resulting in a new total of 1005 samples.
The data show the variables in the columns and samples in the rows. However it was found that some samples are identical in proportions of ingredients, changing only the value of age and compressive strength, for example, samples 653, 654, 678 and 681, shown in the Table 2.

Table 2. Samples with same composition

ID	Cement kg/m^3	B.F.S. kg/m^3	Fly ash kg/m^3	Water kg/m^3	Superp. kg/m^3	C.Aggregate kg/m^3	F.Aggregate kg/m^3	Day	Comp.Str. MPa
653	102	153	0	192	0	887	942	3	4.57
678	102	153	0	192	0	887	942	7	7.68
681	102	153	0	192	0	887	942	28	17.28
654	102	153	0	192	0	887	942	90	25.46

In addition, there are also samples with the same values and proportions of ingredients, but with different compressive strength, probably due to differences in the building process. This is the case, for example, of samples 472, 473 and 474, shown in the table **??**.

_____

2.5 Lorem ipsum Sub

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum

dolore eu fugiat nulla pariatur.

2.6   Lorem ipsum Sub 2

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

2.7   Equations, symbols and units

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

$$q_r = -4pr^2 k \frac{dT}{dr}. \tag{1}$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

Cite as Eq. 1.

2.8   Figures and tables

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

Table 3. Coefficients in constitutive relations

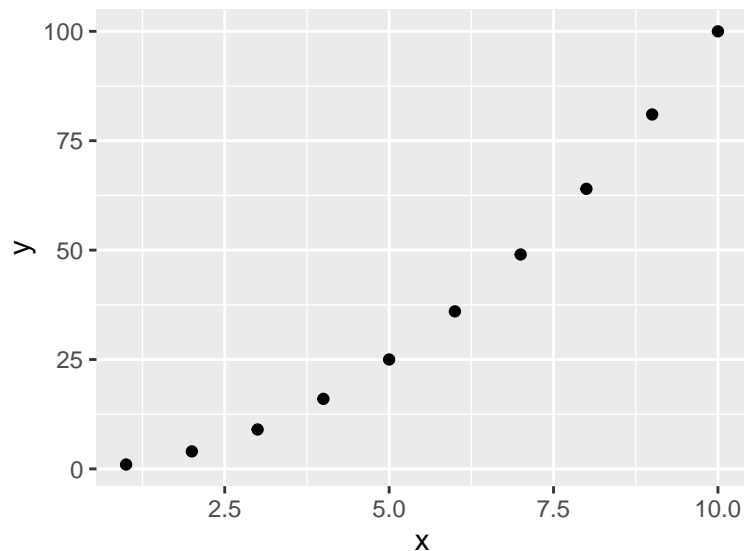| Constitutive relation	Nomenclature	Value
Turbulent tensor	C	0.09
Turbulent tensor	C	0.69
Lateral lift	C	0.08
Virtual mass	C	0.80

Cite as Figure 1 and Table 3.

Figure 1. Pressure variation along the nozzle: experimental data

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

3   tes

Example of reference: "Yeh [3] proposed... Hameed and Khalid [5]. Al-Shamiri AK [4], ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.

References

[1] Hasan, M. & Kabir, A., 2011. Prediction of compressive strength of concrete from early age test result.

[2] Kabir, A., Hasan, M., & Miah, K., 2012. Predicting 28 days compressive strength of concrete from 7 days test result.

[3] Yeh, I.-C., 1998. Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research*, vol. 28, pp. 1797–1808.

[4] Al-Shamiri AK, Yuan T-F, K. J., 2020. Non-tuned machine learning approach for predicting the compressive strength of high-performance concrete. *Materials*, vol. 13, pp. 1023.

[5] Hameed, M. & Khalid, M., 2020. *Prediction of Compressive Strength of High-Performance Concrete: Hybrid Artificial Intelligence Technique*, pp. 323–335.

[6] Abban, D., 2016. Concrete compresive strength.

[7] Raj, P., 2018. Predicting compressive strength of concrete.

[8] Modukuru, P., 2020. Concrete compressive strength prediction using machine learning.

[9] Pierobon, G., 2018. A comprehensive machine learning workflow with multiple modelling using caret and caretensemble in r.

[10] RStudio Team, 2020. *RStudio: Integrated Development Environment for R.* RStudio, Inc., Boston, MA.

[11] R Core Team, 2020. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

[12] , 2008. Concrete compressive strength data set.