RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

Machine learning applied to civil engineering: predictive analysis of concrete compressive strength.

Machine learning applied to civil engineering: predictive modeling of concrete compressive strength

Automatic learning applied to civil engineering: predictive analysis of resistance to hormigón compression

Luciana Andrade Villar1
Milton Augusto Pinotti2

ABSTRACT: Significant progress has been observed in the digitalization of civil engineering, driven by the development of computational technologies and the increased availability of technical data from production processes. In this scenario, Artificial Intelligence (AI) techniques, such as Machine Learning (ML) and Deep Learning (DL), are applied to civil engineering to analyze large volumes of data and identify complex patterns in mix design and strength variables. In the area of concrete technology, these techniques have been investigated as promising tools for estimating the mechanical properties of the material, such as compressive strength, from parameters related to mix design and production conditions. This article presents the application and evaluation of Machine Learning models for the predictive analysis of concrete compressive strength ($f$ck). The main objective is to apply supervised regression algorithms to estimate $f$ck from mix design parameters, contributing to technological control and decision-making in structural projects. The methodology adopted included the collection of 300 samples of real mixes from a concrete plant located in Brusque/SC, covering ten attributes related to the composition of the concrete. A preliminary selection of algorithms was carried out, including Linear Regression, Decision Tree, Random Forest, SVR, and K-nearest neighbors (KNN), with SVR being selected as the best-performing model based on data similarity and proximity. Model validation was planned using historical and experimental data, employing k-fold cross-validation. It is concluded that the SVR-based approach offers a promising tool for predicting $f$ck, with potential for practical application in concrete quality control and optimization of mix designs in batching plants.

Keywords: concrete compressive strength; Machine Learning applied to concrete strength; K nearest neighbors; SVR; prediction.

1 INTRODUCTION

Technological innovations have permeated various segments of contemporary society.
promoting significant transformations in production processes and information management
and in decision-making across different fields of knowledge. In the field of civil engineering,
This scenario is no different, given the advancement of computer technologies and...
Data analysis tools have contributed to the improvement of traditional methods.

[1] Student of Civil Engineering at UNIFEBE. Email: luciana.villar@unifebe.edu.br
[2] Supervising professor. Master's degree in Electrical Engineering. Email: pinotti@unifebe.edu.br

w

RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

Project design, execution and control of construction works.

This digital evolution allows for a more thorough analysis of the characteristics of materials.
used in construction. Among the mechanical characteristics, the compressive strength ($f$ck,
Characteristic strength at 28 days (NBR 5738) is a key parameter for performance.
Structural and quality control, in accordance with technical standards.
The compressive strength of concrete ($f$ck) is directly related to the mix design.
and to the production conditions, including the cement-aggregate-water ratio, the relationship
Water/cement (w/c), additives, and the conditions of mixing, placement, compaction, and curing. A
The simultaneous influence of these variables generates complexity, which makes accurate predictions challenging.
with traditional methods (Neville, 2011; Helene and Terzian, 1992).
In this context, technological control of concrete plays a fundamental role in
guaranteeing the quality and structural performance of buildings. This process involves...
verification of the properties of the constituent materials, monitoring of the stages of
production and the carrying out of laboratory tests, including the compression strength test.
one of the main methods used to evaluate the quality of concrete produced.
However, these procedures are carried out after the material has been produced, which limits the...
possibility of anticipating the performance of concrete before its application in construction (Helene;
Terzian, 1992).
In parallel, significant progress is being observed in the digitization of engineering.
civil, driven by the development of computer technologies and the increase in
availability of technical data from production processes. In this scenario,
Artificial Intelligence (AI) techniques, such as Machine Learning (ML) and Deep Learning (DL),
These techniques are applied in civil engineering to analyze large volumes of data and identify patterns.
complexes in dosage and resistance variables (Goodfellow; Bengio; Courville, 2016).
Machine Learning, in particular, stands out for enabling...
development of predictive models capable of learning from historical data,
To identify relationships between variables and make predictions based on observed patterns.
Unlike traditional computer systems, which are based on pre-established rules...
When programmed, machine learning models are able to extract patterns.
directly from the data, allowing the resolution of complex problems through processes.
of machine learning (Goodfellow; Bengio; Courville, 2016).
In the field of concrete technology, these techniques have been investigated as tools.
promising for estimating the mechanical properties of the material, such as resistance to

w

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

compression, based on parameters related to dosage and production conditions.
Given this context, the present work consists of a literature review on the
application of Machine Learning techniques for the analysis and prediction of resistance to
Concrete compression. The research focuses on the theoretical basis of the main...
Regression algorithms used in the literature, with emphasis on SVR (Support Vector) models.
Regression) and Decision Tree Regressor, as well as a description of the procedures.
methodological and evaluation metrics employed in related studies. The tests
experiments with predictive models, including the hyperparameter optimization step,
Validation with real data and comparative analysis of the results will be carried out in a later stage.
subsequent to the investigation, which establishes the natural continuation of this study.
Thus, the general objective of this work is to conduct a literature review on
Predictive models based on Machine Learning techniques aimed at forecasting
compressive strength of concrete, based on mix design parameters and historical data.
production methods. The relevance of this research lies in the possibility of integrating methods
from traditional concrete technology to advanced data analysis tools,
contributing to the improvement of forecasting and quality control processes of
material in civil engineering.

2. THEORETICAL FOUNDATION

This study covers the principles of concrete technology and the computational foundations of...
Machine Learning, establishing the connection between dosage variables and capacity.
Predictive capabilities of supervised algorithms.

2.1 Concrete Technology

Concrete is a composite material made up of a binder paste (Portland cement and
water) and aggregates (fine and coarse), whose performance results from chemical and physical interaction.
in the interfacial transition zone between the paste and the particles (Mehta; Monteiro, 2014). This
The interaction determines crucial properties, such as workability in the fresh state and...
strength in the hardened state. Its wide application in civil engineering stems from attributes
such as high mechanical strength and durability under diverse environmental conditions,
Its versatility in molding and relatively economical cost solidify its position as a raw material.

RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

essential in buildings and infrastructure works (Neville, 2011).

The complexity of the interactions between the components of concrete and environmental factors. This introduces an inherent variability to the material. To monitor this behavior, they use- if standardized test protocols: while the slump test evaluates the For workability and consistency in the fresh state, axial compression is employed. to quantify the characteristic resistance (fck) at 28 days. In this scenario, the pronounced non-linearity of the interactions between the inputs, evidenced by the moderate correlation between the Cement consumption and fck (r=0.52) justify the adoption of Machine Learning techniques for a more robust and assertive prediction (Hoefelmann, 2021).

2.1.1 COMPOSITION

The composition of concrete is the primary factor that defines its performance. Cement Portland, when reacting with water in a hydration process, forms the gel CSH (silicate of hydrated calcium), which acts as the binding matrix, agglutinating the aggregates.

The quality and quantity of aggregates (fine and coarse) are equally crucial. influencing the workability, density, and final strength of the concrete, being the Particle size distribution is a fundamental aspect to be controlled (Carpinteiro, 2005). The relationship water/cement ratio (ÿ) is one of the most critical parameters, as it directly governs porosity. of the cement paste and, consequently, the strength and durability of the material.

In a regional context, Resner (2021) analyzed traces of concrete in plants in Santa Catarina, identifying dosage patterns that align with the NBR 12655 guidelines. (ABNT, 2015) for site-mixed concrete. Table 1 presents a summary of the mix proportions. typical observations, highlighting the range of variation of the components and their influence on the compressive strength.

Table 1 – Typical traits

Component	Average trace (kg/m³)	SC trace (kg/m³)	Influence fck
Cement	350	300-450	High (r=0.52)
Water	175	150-200	Negative (ratio ÿ 0.45-0.55)
Aggregate Kid	700	650-750	Average (particle size)
Aggregate Big	1050	1000-1100	Medium (shape and strength)
Age Healing	28 days	7-28 days	High (r=0.48)

Source: Author, based on an adaptation from Resner (2021).

w

RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

2.1.2 Influencing Factors and Variability

The variables that influence the compressive strength (fck) of concrete are diverse, being
The water/cement ratio is the determining factor for the quality of the mixture. According to the
principles established by Neville (2011), a smaller proportion of water in relation to
The volume of cement reduces the porosity of the hardened paste, increasing its mechanical strength.
provided that complete hydration of the cement particles is ensured.
To balance this relationship without compromising workability, additives are used.
Superplasticizers, which allow for reduced water consumption while maintaining the necessary flow.
to thickening. Monitoring this consistency in the fresh state is carried out by means
from the slump test, which ensures compliance with
The mixture is prepared before molding. The development of strength is then monitored during the process.
curing period, under controlled temperature and humidity conditions, at an age of 28 days.
as the normative framework for validating fck results, following the criteria of
current technological controls.

2.2 COMPRESSIVE STRENGTH OF CONCRETE

The characteristic compressive strength ($f$ck, MPa) at 28 days defines the sizing.
structural (NBR 6118:2023).
This property is validated as demonstrated by cylinder tests, which calculate $f$ck.
such as the 5th percentile (95% reliability). In this context, for ready-mixed concrete in SC,
A load of 25-30 MPa balances cost, durability, and productivity.

2.2.1 Dosage and Main Influences

The water/cement ratio (w/c) is the determining factor for compressive strength (fck).
Using reduced indices (such as the range of 0.45 to 0.55) decreases the porosity of the paste.
hardened cement, resulting in increased mechanical strength according to the principles of
Abrams' Law (NEVILLE, 2011). This dosage ensures the balance between hydration.
complete separation of cement particles and the workability necessary for compaction, which
It mitigates the formation of initial cracks and meets the durability requirements stipulated in NBR.
6118, item 7.4.2.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

2.2.2 Specific Factors of FCK Resistance and Control of VARIABILITY

Several factors modulate the ultimate compressive strength (fck). In addition to the water-

Regarding cement, key factors include the type of cement, the quality of the aggregates, and the efficiency of the curing process.

(Neville, 2011). In addition, environmental variables, such as temperature and relative humidity,

They play a critical role in the hydration kinetics of cement. According to Mehta and

Monteiro (2014), continuous monitoring of these conditions is essential and should follow the

guidelines of NBR 14931, especially in precast elements, in which control

Strict temperature and consistency control in the fresh state ensures structural integrity.

This dependence on multiple factors is reflected in the productive variability of concrete, which

It commonly exhibits dispersion, with coefficients of variation between 10% and 15% (NBR 12655).

These fluctuations result from unavoidable variations in the mixing, transport, and curing stages.

A significant limitation of the conventional process is that destructive testing, carried out

According to NBR 5739, results at 28 days are provided late, which limits interventions.

immediate. Therefore, understanding this variability reinforces the importance of methods of

More agile and rigorous controls, based on current technical standards.

2.3 MACHINE LEARNING AND PREDICTIVE MODELS

Machine Learning (ML) is a field of artificial intelligence that enables systems to...

Learn and improve your performance from data, without being explicitly

programmed for each specific task (Goodfellow et al., 2016). This ability to

Identifying complex patterns and making predictions makes ML a valuable tool in

civil engineering, especially in the modeling of materials such as concrete.

Predicting the compressive strength ($f$ck) of concrete is a remarkable challenge, given the nature

non-linearity of the interactions between its components, such as the water/cement ratio (ÿ), the type of

cement, the proportion of aggregates and the curing age.

Traditional methods often simplify these relationships, while ML offers...

a robust approach to capturing this complexity and optimizing quality control.

2.3.1 ML ALGORITHMS

A machine learning algorithm is a set of instructions for solving problems through...

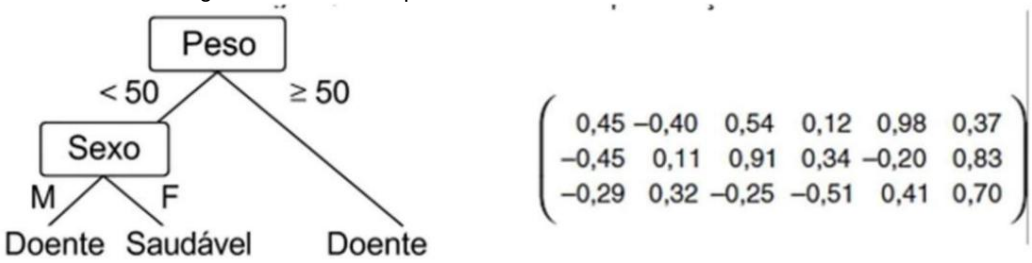
w

RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

Learning from data, without explicit programming. In the context of predicting ƒck from Specifically, attribute hypotheses (ÿ, cure) are constructed to map non-linear relationships, in that each algorithm uses a specific representational bias, such as decision trees (hierarchy) or neural networks (weights), limiting the hypotheses for generalization (Figure 1).

Figure 1 - Different representational biases



(a) Decision tree Source: Carvalho et al. (2021, p. 49).

(b) Neural networks

El sesgo de búsqueda explora modelos de forma eficaz. Essentiali para predições robust in concrete data.

2.3.2 REGRESSION AND CLASSIFICATION

A predictive machine learning (ML) algorithm is a function that, based on a From a set of labeled examples, an estimator is constructed. The attribute takes values within a previously known domain and, when that domain is composed of nominal values, The problem is classified as a classification (or concept learning) problem, and the estimator The generated value is called a classifier. On the other hand, if the domain is an infinite set and Ordered set of values, this is a regression problem and the estimator is known as regressor. (Dietterich, 1998).

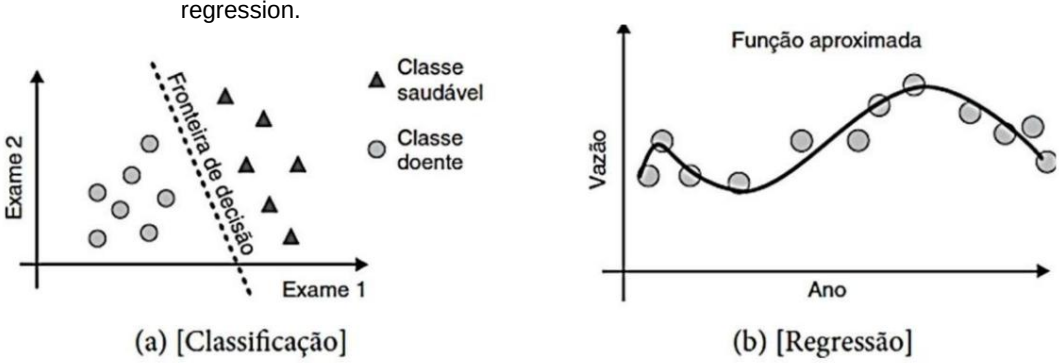
Both the classifier and the regressor are functions that receive a non-example. labeled and produce an output: in the case of the classifier, it assigns the example to one of the classes. possible; in the case of the regressor, it estimates a real value corresponding to the example presented. Figure 2 illustrates the definition of classification and regression. (Carvalho et al., 2021).

w

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

Figure 2 – Graph representing the classification and the regression.



Source: Carvalho et al. (2021), p 49.

2.3.3 Regression Algorithms

The CART (Classification and Regression Trees) decision tree is a regression algorithm.
which builds a hierarchical decision structure based on recursive divisions of the data,
using criteria such as Gini or Entropy to minimize impurities, making it interpretable and
robust to outliers (Breiman et al., 1984).
The Random Forest (RF) extends CART through an ensemble of
Multiple trees (bagging + feature randomness), reducing variance and overfitting, with
aggregated predictions by the average (Breiman, 2001).
The robustness of this method supports its technical viability in civil engineering, as studies
Recent studies validate the application of CART and Random Forest in predicting ƒck, according to NBR standards.
6118 (2023) and NBR 12655 (2022), which emphasize the control of variables such as the ratio ÿ and the
cure.

Silva et al. (2023) analyzed Brazilian concretes using Random Forest, obtaining $R^2 = 0.95$ and
highlighting ÿ and aggregates as main predictors. Omotayo, Arum and Ikumapayi (2024),
In turn, they compared ML methods, with RF outperforming CART (RMSE = 3.2 MPa).
vs. 4.8 MPa). Xu et al. (2021), on the other hand, proposed an ensemble model for concrete.
ready, achieving $R^2 = 0.92$ in variable traces.
K-Nearest Neighbors (KNN) is an instance-based learning algorithm.
which makes predictions based on the similarity between samples in the attribute space.
Unlike tree-based models, KNN does not construct a function of
Explicit mapping during training, but stores the entire dataset for
to make inferences at the time of prediction.

w

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

2.4 Metrics for Evaluating Regression Models

Evaluating the performance of Machine Learning models, especially in Regression tasks, such as predicting the characteristic compressive strength (ƒck) of Specifically, these are crucial for determining its accuracy and generalizability. Statistical metrics allow us to quantify the discrepancy between the values predicted by the model. and the actual observed values, providing an objective basis for comparison and selection of algorithms (Harrison, 2020; Carvalho, 2021). Therefore, the appropriate choice of metrics is fundamental to understanding the characteristics of the model errors and their adequacy to problem in question.

2.4.1 Coefficient of Determination (R^2)

The Coefficient of Determination, or R^2, is a widely used metric for evaluating the proportion of the variance of the dependent variable (ƒÿÿ) explained by the independent variables of the model. Its value ranges from 0 to 1, where values closer to 1 indicate that the model explains a larger portion of the data variability, that is, it presents a better fit. It is calculated according to Equation 1:

$$= \quad - \quad \frac{}{\sum_{i=1}^{n}(y_i - \hat{y})} \quad ÿ \quad ÿ\,(\,) \qquad (1)$$

Where

= Coefficient of determination

= predictive function yi = Known target attribute. Represents the actual value of ƒ ÿÿ for the i-th sample.

xi = value predicted by the model for the i-th sample.

Average of the actual values of ƒ ÿÿ.

n = total number of samples.

For Random Forest models applied to predicting ƒck, an R^2 between 0.92 is expected. and 0.98, indicating high explanatory power.

2.4.2 MEAN SQUARE ERROR (MSE) AND ROOT MEAN SQUARE ERROR (RMSE)

The Mean Squared Error (MSE) measures the average of the squares of the errors, that is, the difference

w

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

between predicted values and actual values. It is a metric that penalizes larger errors in form.
The most significant factor is the squaring of the equation, making it sensitive to outliers. The RMSE is the same metric as MSE, but normalized to the same unit as the target variable, making The analysis and interpretation of the error become more intuitive. The MSE is defined according to equation 2 and the RMSE according to equation 3:

$$( ) = \_ \quad (\ddot{y} ( )) \qquad (2)$$

Where:

Average error

$$( ) = \ddot{y} \qquad (3)$$

Where:

= Root mean square error

2.4.3 MEAN ABSOLUTE ERROR (MAE)

The Mean Absolute Error (MAE) calculates the average of the absolute values of the errors. Unlike Unlike MSE/RMSE, MAE does not square errors, making it less sensitive to outliers.
and more robust in error distributions with extreme values. The MAE is expressed by
Equation 4:

$$( ) = \_ \quad [\ddot{y} ( )] \qquad (4)$$

Where:

Mean Absolute Error

w

RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

2.4.4 Waste Analysis and Assessment Flow

In addition to quantitative metrics, graphical analysis of residuals (differences between values) (observed and predicted) is fundamental (Hoefelmann, 2021). This evaluation allows Identify the characteristics of a good model: residuals distributed symmetrically around from zero, with no discernible patterns, which indicates that the errors are random and that there is no bias. systematic in predictions (Harrison, 2020).

Plots of residuals versus predicted values or versus independent variables can reveal Problems such as heteroscedasticity or omitted variables. Therefore, the evaluation flow Machine learning models typically involve dividing the dataset into Training and test subsets (train_test_split) to ensure the model is evaluated based on data not seen during training.

For a more robust assessment and to mitigate dependence on the specific division of the data, Techniques such as k-fold cross-validation are employed. In it, the The dataset is divided into 'k' parts, and the model is trained and tested 'k' times, using A different part is tested in each iteration; the metrics are then calculated as the average. of the 'k' results.

These evaluation metrics (R^2, RMSE, MAE) were applied in real-world contexts, aligned to NBR 12655 for ƒcÿ tests. The studies by Hoefelmann (2021) and Chou et al. (2011) They confirm the robustness of RF in capturing non-linearities in curing and aggregates, with Random residuals indicating good generalization.

3 METHODOLOGICAL PROCEDURES

This chapter presents the methodological approach established for carrying out the work. understanding the phases from data collection to model verification predictive, given the purpose of creating a Machine Learning system capable of estimating the compressive strength (fck) of concrete based on its composition.

The study begins with a literature review to provide a technical foundation for the research. followed by the use of a database from a local concrete plant.

Once the dataset was defined, the analysis phase began, consisting of extracting the... variables, their preprocessing, and subsequent partitioning into training and test samples.

During the modeling phase, an initial screening was performed among several regression algorithms.

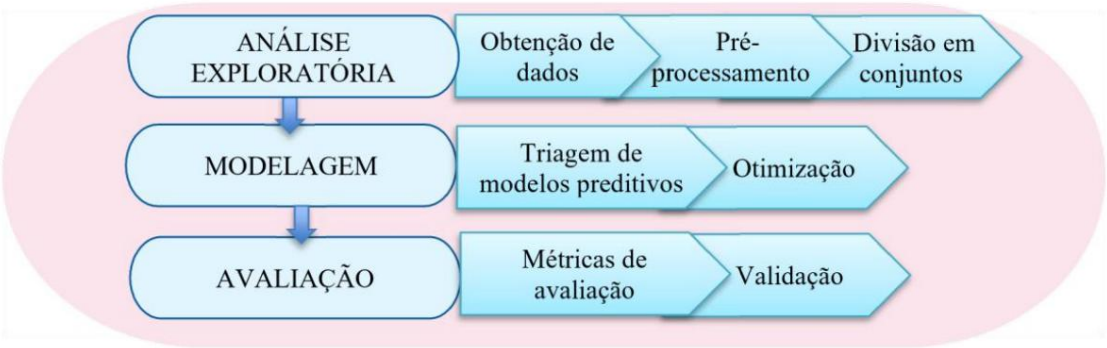
w

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

for comparison, progressing towards optimizing the one with the best performance.
Finally, the evaluation stage describes the performance indicators and the validation process.
of the developed tool. Figure 3 presents the methodological flowchart that summarizes the
work process.

Figura 3 – Fluxograma metodológico



Fonte: Os autores (2026).

The data processing and algorithm development were carried out in
The Python language, using the Anaconda distribution and the Jupyter application.

3.1 DATASET

In order to ensure the integrity, relevance, and compliance of the information with the scope.
For this research, a dataset was chosen that covers dosage parameters and
The results of compression strength tests. This dataset is composed of real records collected.
directly at a concrete batching plant located in the municipality of Brusque/SC. The
Data was extracted from technical production reports and laboratory tests, including
a robust sampling that reflects practical manufacturing conditions and technological control.
of the region.

Table 2 – Trace parameters

Cement component		Sand Artificial	Natural sand	Brita 1	Brita 0	Water,	Superplasticizer Additive
Line 1	277	522	352	655 276	185 2.2	488 320	195 0.0
Line 2	360	449 454 Source:		1.08			1.08

The authors (2026).

Based on the composition of the traits presented in Table 2, the following were carried out:

w

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

Molding and curing procedures for cylindrical test specimens for each mix design.
experimental. To ensure the representativeness of the sample and the reliability of the results,
Serial moldings were performed for each dosage, strictly following the criteria of
compaction and wet curing. Compressive strength readings were obtained by means
from the breaking of the test specimens in a properly calibrated hydraulic press, with records
performed at 28 days of age. The resulting dataset, consolidated from the
Technical reports from the regional plant served as the basis for the training and validation of
predictive models.

3.1 PRE-PROCESSING

3.1.1 INTEGRATION

The dataset used in this research was compiled from...
Technical records from the regional plant, integrating dosage variables and results of
resistance in a unified database. This database is composed of 10 attribute columns,
encompassing the components of the mixture (cement, artificial sand, natural sand, crushed stone 1, crushed stone 0,
Water, additive and superplasticizer) and the mechanical performance results (R7 and R28).
In total, 300 records of real test specimens were processed. After the integration and...
After structuring, the data was exported to the Jupyter environment in .csv format, allowing
The start of preprocessing and predictive analytics procedures.

3.1.2 ATTRIBUTE EVALUATION

The attribute evaluation phase was conducted with the aim of selecting the variables of
greater predictive power and ensuring the statistical integrity of the model. The following were removed
identification records of the test specimens (ID CP), as well as the columns that presented
constant values or values that did not present relevant variability for learning
algorithm.
Unlike the identifiers, the 7-day resistance result (R7) was maintained in
dataset. The persistence of this attribute allows the model to capture the kinetics of
initial hardening of concrete, serving as a valuable technical indicator for prediction
from resistance to 28 days (R28). This integration allows the algorithm to identify patterns.
of strength gain that depend on the chemical interaction between the cement and the additives to

RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

over time.

The exclusion of these variables was performed using the .drop function from the Pandas library, which This allowed us to optimize the dataset. With this procedure, the structure of the dataset was... refined, retaining only the essential attributes for modeling, which reduces the computational complexity and mitigates the risk of overfitting. Figure 4 shows the statistical summary of the dataset

Figure 4 – Statistical data of the attributes.

	Cimento	Areia Artificial	Areia Natural	Brita 1	Brita 0	Água	Aditivo	Superplastificante	R7 (Mpa)	R28 (Mpa)
count	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000	300.000000
mean	304.666667	497.666667	386.000000	599.333333	290.666667	188.333333	2.140000	0.360000	27.845667	33.617667
std	39.191950	34.470028	48.163601	78.856091	20.776455	4.721922	0.113326	0.509968	3.249906	2.726052
min	277.000000	449.000000	352.000000	488.000000	276.000000	185.000000	1.980000	0.000000	20.100000	25.800000
25%	277.000000	449.000000	352.000000	488.000000	276.000000	185.000000	1.980000	0.000000	25.675000	31.600000
50%	277.000000	522.000000	352.000000	655.000000	276.000000	185.000000	2.220000	0.000000	27.600000	34.000000
75%	360.000000	522.000000	454.000000	655.000000	320.000000	195.000000	2.220000	1.080000	30.000000	35.500000
max	360.000000	522.000000	454.000000	655.000000	320.000000	195.000000	2.220000	1.080000	35.000000	41.200000

Source: The authors (2026).

3.1.3 STANDARDIZATION

Data standardization is a pre-processing step aimed at normalizing the scales. of the variables, ensuring that all have a mean of zero and a standard deviation of one. This This procedure is fundamental to the performance of machine learning algorithms. because it prevents attributes with higher orders of magnitude from exerting influence. disproportionate during model training.

In the context of this research, the application of standardization is justified by the numerical disparity. Among the input variables are cement consumption (in kg/m³) and additive content. To perform this task, the StandardScaler function, available in the module, was used. preprocessing of the Scikit-learn library.

3.2 DIVISION OF DATA SETS

The initial stage in building a predictive architecture consists of training the algorithm. with the consolidated database. For this purpose, the scikit-learn library was used. (sklearn), specifically through the train_test_split function. This feature allows for splitting The information was randomly distributed into two distinct groups: one focused on learning and the other

w

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

dedicated to validating the results. In this research, it was defined that 80% of the records would be allocated for training, and the remaining 20% for the testing phase.

To ensure an unbiased assessment of the system's performance, the data was divided. in training and test sets. This distribution ensures that the model's effectiveness is Validated on data not seen during the learning phase. The organization of the variables. (attributes) was structured as follows:

- X_train: includes the predictor attributes (dosages) used for the learning from the model;
- y_train: stores the target variable corresponding to the compressive strength. (fck) for training;
- X_test: includes the input variables that will be used to test the accuracy of model;
- y_test: contains the actual resistance values associated with X_test, allowing statistical verification of the forecasts.

This data partitioning methodology is essential to verify the potential of Generalization of the model in the face of new dosages, mitigating the risk of overfitting the data. originals.

3.3 Preliminary Evaluation of Regression Algorithms

With the aim of identifying the algorithm with the greatest generalization capacity for the To estimate the concrete's resistance, an initial comparative screening was carried out between the regression models are available in the scikit-learn library. To ensure the selection of The following algorithms were evaluated to determine the most efficient and robust model:

- Linear Regression (sklearn.linear_model): classic statistical model for analysis of linear relationships;
- Decision Tree Regressor — CART (sklearn.tree): algorithm that subdivides the data into hierarchical structures;
- Random Forest Regressor (sklearn.ensemble): technique of an ensemble that combines multiple decision trees;
- Support Vector Regressor — SVR (sklearn.svm): an algorithm that searches Find the best hyperplane for regression.

RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

- K nearest neighbors — KNN (sklearn.neighbors): a model based on similarity and proximity of the data;

The training of each model was performed using the fit(X_train, y_train) function, which submits the dosage parameters and their respective resistance results to processing of algorithms. This procedure allows the algorithms to identify patterns and interdependencies present in the dataset. After the training phase, the following was used: The predict(X_test) function was used so that the models could process the variables from the test set. generating the vector y_pred. This set stores the estimates produced by each algorithm. regression (listed above), enabling subsequent statistical analysis in relation to Actual values measured in the laboratory.

3.4 EVALUATION OF THE MODELS

The evaluation of the effectiveness of the models detailed in the previous section was carried out based on Four statistical regression indicators: R^2, MAE, MSE, and RMSE. These metrics allow Compare the actual results obtained on the test set (y_test) with the generated estimates. through algorithms (y_pred), enabling the quantification of the accuracy of the predictions. The implementation of these indicators was carried out through specific functions of The scikit-learn library. The R^2 coefficient allowed us to verify the percentage of variance explained. by the model, while MAE and RMSE provided the magnitude of the mean error and the sensitivity to larger deviations, respectively.

3.5 Algorithm Optimization

The performance of the predictive model was improved through the configuration of hyperparameters, adjusting the internal parameters of the algorithms to increase precision and robustness of the estimates. For this purpose, the random search technique was used, with Cross-validation was performed using the RandomizedSearchCV function from the scikit-learn library. This This approach enabled the efficient exploration of various combinations of pre-parameters. defined.

The hyperparameters and their respective variation ranges were defined according to the Specifics of the K-Nearest Neighbors (KNN) model. The search was configured to process a given number of random combinations and, at the end of the procedure,

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

The configuration that showed the best average performance in the validation was selected for compose the final version of the predictive model.

3.6 MODEL VALIDATION

To validate the algorithm developed in this research, two procedures were adopted. distinct. The first consisted of cross-validation of historical data from the Brusque plant, integrated into the standard Machine Learning training process. The second procedure This involved validation using experimental data collected at the plant itself, with the purpose of to verify the model's ability to generalize to unprecedented dosages that were not presented during the learning phase.

3.6.1 Validation with Experimental Data

To validate the developed algorithm, a verification step was performed using data. Experiments were collected directly at the batching plant. The objective of this procedure was to... to test the model's ability to generalize to real-world dosages that did not include the initial training set. The traits were selected to replicate the conditions operational aspects of the database, maintaining the same input specifications and relationships. The water/cement ratios are presented in Table 2.

The data collection followed the plant's laboratory standards, in which the results of Compressive strength was measured using calibrated hydraulic presses, as per the NBR 5738 and NBR 5739 standards. Unlike the machining and measurement method of roughness proposed by Canal (2022), this validation was based on technological control. of the hardened concrete at 28 days. The new values were duly tabulated and imported into the Jupyter environment, where, using the predict(X_valid) function, it was generated the set of predictions y_pred_valid.

Finally, the actual results of the experimental dosages were compared with the Estimates generated by the algorithm. Performance was evaluated using the R^2 metric. MAE, MSE, and RMSE, in addition to residue analysis, attest to the tool's reliability. for practical applications in monitoring the strength of regional concrete.

w

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

4. ANALYSIS OF RESULTS

This section presents the main findings resulting from the application of the models of regression to experimental data, with emphasis on the dependent variable resistance to Concrete compression ($f$ck). First, the performance metrics are described. obtained in the preliminary evaluation stage of the models, followed by a comparison between the different algorithms were tested, with the SVR and the regressor Decision Tree standing out, which They showed similar performance and the best results in this initial phase. Subsequently, the impact of hyperparameters on the model's behavior is examined. The interpretation of error metrics and residual distribution is performed, which makes it possible a reliable evaluation of the results. Finally, the tests performed with are presented. data from the experiment, with the aim of evaluating the model's ability to generalize the Data not observed during training.

4.1 RESULTS OF THE PRELIMINARY EVALUATION OF THE ALGORITHMS

Table 3 presents the performance indicators of the algorithms evaluated in the stage. Preliminary findings. The SVR (Support Vector Regression) algorithm stood out in this phase. presenting the highest coefficient of determination (R^2).

Table 3 – Results of the preliminary assessment

Algorithm	R2 MOTHER	MSE	RMSE	
SVR	0.76	0.48	0.63	0.76 0.75
Decision Tree	0.48	0.65	0.81	0.70 0.61
KNNeighbors	0.79	0.89	0.63	0.69 0.96
Random Forest	0.98	0.63	0.69	0.96 0.98
Linear Regression				

Source: The authors (2026).

4.2 OPTIMIZATION OF THE SVR ALGORITHM

The SVR (Support Vector Regression) and Decision Tree Regressor algorithms were subjected to a hyperparameter tuning process using the function RandomizedSearchCV, with the purpose of identifying the best configuration for the dataset. of the data in question.
The search was parameterized to test 50 random combinations of hyperparameters within

w

RCMOS – Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

pre-established intervals for each algorithm. After execution is complete, the
The combination that showed the best performance in the validation set was selected.
as the final configuration of each model. The hyperparameters subjected to this search,
their respective ranges of variation and the identified optimal values are shown in the Tables.
4 and 5.

Table 4 – Hyperparameter Selection - SVR Algorithm

Kernel Intervals [linear,		Best parameter
sigmoid] [0.1, 1, 10,	rbf, poly, 100]	To be defined
W		To be defined
Epsilon [0.01, 0.1, 0.2, 0.5]		To be defined
Gamma [scale, auto, 0.01, 0.1] TBD Degree [2, 3, 4]		
		TBD Tol
[1e-3, 1e-4, 1e-5]		To be defined

Source: The authors (2026).

Table 5 – Hyperparameter Selection – Decision Tree Regressor Best parameter

Algorithm	Intervals	[squared_error,
Criterion	friedman_mse, absolute_error, poisson] To be defined max_depth [None, 10, 20,	
30]		TBD
min_samples_split [2, 5, 10]		To be defined
min_samples_leaf [1, 2, 4]		To be defined
max_features [sqrt, log2, None] ccp_alpha		To be defined
[0.0, 0.01, 0.1]		To be defined

Source: The authors (2026).

4.3 Feature Importance

Attribute importance analysis, which quantifies the contribution of each variable in
The model's predictions showed that R7 and R28 (concrete strengths at 7 and 28 days) are...
the most determining factors for prediction. These results are consistent, since the
The initial and final strengths of concrete directly reflect the quality of the mix and its
capacity to develop resistance. The other parameters of the trait also showed
significant contributions. However, it was identified that trait 2 generated outliers.
(outliers) that impacted the model's performance.

w

RCMOS - Multidisciplinary Scientific Journal The Knowledge.
ISSN: 2675-9128. São Paulo-SP.

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

4.4 FUTURE WORK: EXPERIMENTAL TESTS WITH SVR AND DECISION TREE

This work consisted of a literature review on the application of algorithms.
Machine Learning for predicting the compressive strength of concrete ($f$ck), with emphasis
in the SVR (Support Vector Regression) and Decision Tree Regressor models. The rationale
The theoretical framework presented, along with a description of the methodological procedures and the...
evaluation metrics provided the necessary conceptual basis for the continuation of
investigation.
As a future step, experimental tests are planned with the SVR algorithms and
Decision Tree Regressor, using the dataset described in section 3.1. Both
The models will undergo a hyperparameter optimization process using the function.
RandomizedSearchCV, within pre-established intervals, with the objective of
Identify the configuration with the best predictive performance.
After the optimization phase, the models will be validated with experimental data and evaluated.
Based on the metrics R^2, MAE, MSE, and RMSE. Comparative analysis of the results.
This will allow us to identify which of the two algorithms offers the best balance between precision and...
generalization capacity for predicting $f$ck, contributing to the improvement of
Technological control of concrete in civil engineering.

FINAL CONSIDERATIONS

This paper presents a literature review on the application of techniques of
Machine Learning for predicting the compressive strength of concrete ($f$ck), with emphasis
in the SVR (Support Vector Regression) and Decision Tree Regressor algorithms. A
The theoretical foundation addressed the main concepts related to concrete technology.
to the factors that influence its resistance and to the fundamentals of predictive models, as well
such as the evaluation metrics used in the specialized literature.
Based on the review conducted, it was possible to ascertain that the use of regression algorithms
This has proven to be a promising approach for estimating $f$ck based on parameters of
dosage and historical production data. The literature consulted indicates that both SVR and Decision Tree
have the potential to capture non-linear relationships between variables.
input and concrete strength, which can aid in technological control and in

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

Optimization of features.

The analysis of the methodological procedures described in related studies allowed identify the fundamental steps for the development of predictive models, including the data preprocessing, splitting training and test sets, evaluation Preliminary algorithms and hyperparameter optimization. The metrics R^2, MAE, MSE and RMSE measurements proved suitable for the comparative evaluation of model performance.

It should be noted that this work constitutes an initial stage of investigation, with the main objectives being... contributions to the systematization of the theoretical framework and the definition of the methodology to be used applied. Experimental tests with the SVR and Decision Tree Regressor algorithms, including Hyperparameter optimization via RandomizedSearchCV and validation with real data.

These will be carried out at a later stage, which represents the natural continuation of this research.

Future results are expected to contribute to the improvement of technological control. concrete in civil engineering, integrating traditional methods with advanced tools. data analysis and enabling faster and more efficient estimates of resistance to material even before the tests were carried out at 28 days.

REFERENCES

Brazilian Association of Technical Standards. NBR 5738: Concrete — Procedure for molding and curing test specimens. Rio de Janeiro, 2015.

Brazilian Association of Technical Standards. NBR 5739: Concrete — Compression test of cylindrical specimens. Rio de Janeiro, 2018.

Brazilian Association of Technical Standards. NBR 6118: Design of concrete structures — Procedure. Rio de Janeiro, 2023.

Brazilian Association of Technical Standards. NBR 12655: Portland cement concrete — Preparation, control, receipt and acceptance — Procedure. Rio de Janeiro, 2022.

BISHOP, CM Pattern recognition and machine learning. New York: Springer, 2006.

BREIMAN, L. Random forests. Machine Learning, vol. 45, no. 1, p. 5–32, 2001.

BREIMAN, L. et al. Classification and regression trees. New York: Chapman & Hall, 1984.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. Cambridge: MIT Press, 2016.

HELENE, P.; TERZIAN, P. Concrete mix design and control manual. São Paulo: Pini, 1992.

w

Year VII, v.1 2026 | Submission: 05/23/2026 | Accepted: 05/26/2026 | Publication: 05/29/2026

MEHTA, PK; MONTEIRO, PJM Concrete: microstructure, properties and materials. 2nd ed. São Paulo: IBRACON, 2014.

NEVILLE, AM Properties of concrete. 5. ed. London: Pearson, 2011.

OMOTAYO, T.S.; ARUM, C.; IKUMAFAYI, CM Comparative analysis of machine learning models for predicting concrete compressive strength. Asian Journal of Civil Engineering, vol. 25, p. 1–12, 2024.

SILVA, R. et al. Prediction of concrete compressive strength using Random Forest. Ibracon Journal of Structures and Materials, v. 16, n. 2, p. 1-15, 2023.
XU, J. et al. Ensemble learning models for predicting concrete compressive strength— Construction and Building Materials, vol. 303, p. 124-132, 2021.

w