XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Application of Machine Learning Techniques in the Pulp and Paper Industry

Breno C. Saturnino Carlos Casais[1], Bruno S. Macêdo[1], Sandro P. da Silva[2], Giovanna G. Miranda[1], Bruno H. Gronner Barbosa[2], Camila M. Saporetti[3], Leonardo Goliatt[4]

[1]*Systems and Automation Engineering Graduate Program, Federal University of Lavras*
Lavras, 37200-000, MG, Brazil
breno.casais1@estudante.ufla.br, brunomsilva200@gmail.com, giovanna.miranda2@estudante.ufla.br
[2]*Dept. of Automatics, Federal University of Lavras*
sandro.silva@ufla.br, brunohb@ufla.br
[3]*Dept. of Computational Modeling, Polytechnic Institute, State University of Rio de Janeiro*
Nova Friburgo, 28625-570, RJ, Brazil
camila.saporetti@iprj.uerj.br
[4]*Dept. of Computational and Applied Mechanics, Federal University of Juiz de Fora*
Juiz de Fora, 36036-900, MG, Brazil
leonardo.goliatt@ufjf.br

Abstract. Bleaching in the pulp and paper industry is essential to produce high-quality paper, but it faces challenges due to the complexity and interdependence of processes, as well as limitations in traditional control methods. Machine learning (ML) techniques appear to be a promising solution to predict pulp brightness and adjust operating conditions in real time. This study evaluates four algorithms (Support Vector Regression (SVR), Random Forest (RF), Linear Regression (LR), and Decision Tree (DT)) to predict brightness at the D1 bleaching stage in a pulp and paper industry. Experiments were performed using all features, with normalized features, with Principal Component Analysis (PCA) as feature selection, and with feature selection based on process knowledge. The RF model obtained the best results, with R^2 of 0.6576, 0.6576, 0.7177, and 0.7291, respectively, in each experiment.

Keywords: Machine Learning, Industry, Paper, Pulp, Bleaching.

1 Introduction

The bleaching process in the paper and pulp industry is a crucial step in producing high-quality paper, where the removal of residual lignin from cellulose fibers is essential to achieve the desired whiteness. This process involves a series of complex chemical reactions and is highly dependent on the precise control of operational variables such as temperature, pH, and the dosage of chemical reagents [1].

Figure 1 illustrates the bleaching process used in the paper and pulp industry, which combines three distinct stages to optimize the final product's whiteness. The process consists of different stages and equipment, represented schematically. The cycle starts with the lignin removal pulp, which undergoes a series of chemical and physical processes, such as the application of chlorine dioxide (ClO_2), sodium hydroxide (NaOH), oxygen (O_2), and hydrogen peroxide (H_2O_2), to remove lignin and other impurities present in the cellulose [2, 3].

However, controlling the variables in the bleaching process used in the paper and pulp industry is challenging due to the dynamic and interdependent nature of the processes, in addition to limitations in traditional measurement and adjustment methods. A critical problem lies in the ability to predict, in advance, the final whiteness of the cellulose and the impact of operational conditions in real-time, allowing for proactive adjustments. The lack of robust and automated predictive tools can lead to fluctuations in the final product quality, increased chemical reagent consumption, and higher waste generation, negatively impacting the process's costs and sustainability.

Given this scenario, the use of Machine Learning (ML) techniques emerges as a promising solution to model the complexity of the bleaching process, integrating large volumes of operational data and providing accurate predictions. These techniques can enable the anticipation of operational adjustments, optimizing resource use, reducing costs, and ensuring the stability of bleached cellulose quality for high-quality paper production.

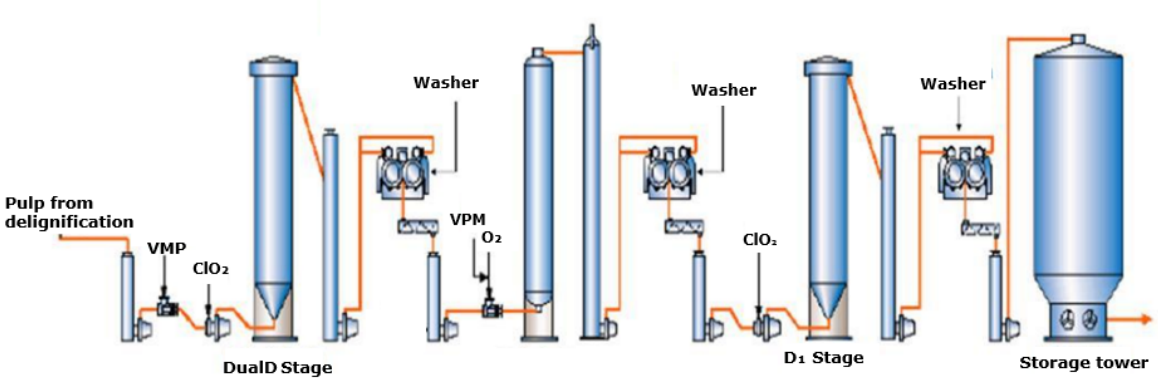


Figure 1. Bleaching Process.

This study aims to advance the field of the paper and pulp industry by developing and validating ML approaches. The specific objectives are:

- Develop and evaluate ML models to predict the whiteness of cellulose in the bleaching process, allowing for the control and optimization of operational variables.
- Identify and select the most relevant attributes for building predictive models, using dimensionality reduction techniques.
- Analyze the influence of each variable on the predictive model, highlighting the operational conditions that most impact whiteness.

The rest of the article is organized as follows: Section 2 reviews some related works on the topic of this study. Section 3 presents the methodology used in this research, such as the database, the method for selecting the most relevant attributes, cross-validation, regression methods, and the metrics for evaluating the models' performance. Section 4 discusses the results obtained. Finally, the conclusions are presented in Section 5.

2 Related Works

The paper industry is a topic that has been studied by several researchers. These studies involve analyses and ML tools proposed to understand the problem and find solutions to predict outcomes.

Gomes [4] validated ML models, such as Linear Regression and Decision Trees, to estimate laboratory measurements from on-line instrument data. The predictive model showed an R^2 of 62.3%, indicating the need for more data to make reliable predictions.

De Paula [5] used ML to predict pulp whiteness, testing algorithms like Decision Trees and Random Forest. The results showed that the algorithms were effective for creating virtual whiteness sensors, improving process efficiency.

The study by Carvalho [6] demonstrated the feasibility of predictive models for whiteness in stages of the cellulose bleaching process, highlighting XGBoost and LightGBM with an R^2 of 0.80, outperforming Linear Regression. Integrating XGBoost with the PI System allowed for real-time predictions, optimizing the process and improving product quality.

Favaro [7] applied models like XGBoost and Neural Networks to predict whiteness in the bleaching process, achieving superior performance, with a MAE of 0.872 and an R^2 of 0.454. XGBoost stood out for its accuracy. Henrique [8] used Neural Networks to predict whiteness and the KAPPA number in four stages of the bleaching process, achieving consistent predictions and aiding in the optimization of reagent control.

Despite the advances, few studies use automated methodologies to select the most relevant attributes from databases, typically relying on expert knowledge. There is, therefore, a gap for creating methodologies that integrate automatic attribute selection and the comparison of different ML models.

3 Methodology

The dataset for this study includes variables collected between October 2023 and October 2024, covering aspects of the bleaching process in a paper and pulp industry in Brazil. With 51,480 samples and 74 variables, the data is divided into four categories: Chemicals and Reagents, Operational Parameters, Production Parameters,

and Quality Variables. The main variables include data on the flows and concentrations of NaOH, hydrogen peroxide (H_2O_2), chlorine dioxide (CLO_2), and caustic soda, as well as parameters such as pH, viscosity, and pulp temperature, which are essential for controlling the bleaching process.

Table 1 summarizes the relevant variables for each category, including data such as paper whiteness at different production stages, which is used as the main quality indicator.

Table 1. Bleaching Process Variables

Chemicals and Reagents	Operational Parameters	Production Parameters	Quality Variable
NaOH dilution screw press	Stand Pipe Level	Daily Production	Whiteness EOP Stage
Caustic Soda	Consistency for Reactor	Production Campaign	Whiteness D1 Stage
H_2O_2 Flow	pH entry EOP	Estimated Production EOP	Kappa EOP Stage
FV50 Peroxide Process	Pulp Temperature after Mixer		
Chlorine Dioxide Flow	Pulp Temperature in Reactor		
NaOH for D1	Washing Flow		
Caustic Soda D2 Tower	Liquor for Stage		

The data underwent a filtering process, considering only productions greater than 1000 t/d. Outliers were eliminated using the Interquartile Range (IQR) method, and missing values were filled with the mean of the respective columns. Irrelevant columns, such as time and campaign, were discarded, resulting in a clean dataset for the subsequent analyses.

3.1 Regression Methods

In this section, the regression methods that were used in this work will be presented, which were Decision Tree (DT), Linear Regression (LR), Random Forest (RF) and Support Vector Regression (SVR).

Decision Tree (DT)

Decision Trees (DTs) are tree-based ML algorithms capable of generating expert systems to solve classification and regression problems. A DT is constructed from a set of tests performed on the input data. The tree's internal nodes represent the tests. In the case of regression problems, these tests are quantitative and are performed with the output value defined for the set of attributes compared to the value that gives the output a split. Tests are typically performed by checking whether the value of a feature is greater than or less than the split value. [9].

Linear Regression (LR)

Linear Regression models the relationship between dependent and independent variables by adjusting the coefficients to minimize the squared errors. The model equation is:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \epsilon$$

where y is the dependent variable, $x_1, x_2, \ldots, x_n$ are the independent variables, β_0 is the intercept, and $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients.

Random Forest (RF)

The Random Forest (RF) algorithm [10] is a ML technique based on an ensemble model. It constructs k decision trees from the training dataset in k iterations. In each iteration, a subset of samples is randomly selected. To construct each decision tree from this subset, RF also randomly selects a set of candidate features.

Support Vector Regression (SVR)

Support Vector Regression (SVR) is an application of the Support Vector Machine (SVM) model aimed at regression problems [11]. This model can be either linear or non-linear, depending on the kernel function used.

The model is:

$$\hat{y_i} = \varepsilon \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b$$

where $\hat{y_i}$ is the value estimated or predicted by the SVR model for a given point $\mathbf{x}$, α_i and α_i^* are the dual coefficients obtained during the resolution of the SVR dual optimization problem, which represent the Lagrange multipliers associated with the problem constraints.

3.2 Grid-Search

The Grid-Search procedure is an exhaustive technique based on a defined subset of the hyperparameter space. Hyperparameters are specified using the minimum value (lower bound), maximum value (upper bound), and number of steps [12]. There are three different scales that can be used: linear scale, quadratic scale, and logarithmic scale. The performance of each combination is evaluated using a number of performance metrics. When using Grid-Search with cross-validation, the process is repeated for each combination of parameters tuned for a regressor to improve performance.

3.3 Evaluation Metrics

The evaluation metrics used include the Coefficient of Determination (R^2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The R^2 represents the proportion of variability explained by the model. The MSE calculates the average of the squared errors, while the RMSE is the square root of the MSE, expressed in the original units of the data. The MAE, in turn, measures the average of the absolute errors, reflecting the average magnitude of the error.

4 Results and Discussion

The computational experiments were conducted using the Python programming language and implementations based on the libraries Scikit-Learn [13], Matplotlib [14], Pandas [14]. All experiments were executed on a computer with the following specifications: Intel(R) Core (TM) i5-1135G7, 8 GB of RAM, and Windows 10 operating system. To evaluate the methodology, a k-fold cross-validation (KF) with $k = 5$ was used.

In this study, four regression models were evaluated for predicting the whiteness at the D1 bleaching stage in a paper and pulp industry. The models tested were: DT, LR, RF, and SVR. For each model, a systematic search for the best hyperparameters was applied. Initially, a set of possible values for each hyperparameter was defined, as shown in Table 2. Then, the Grid-Search method was used to exhaustively test all combinations of these values, evaluating each model using k-fold cross-validation.

The evaluation metrics were calculated for each hyperparameter combination, and the best set of parameters for each model was selected based on the average performance obtained during cross-validation. Four distinct strategies were used in the process: the first strategy used all available features, while the second involved data normalization. The third strategy applied dimensionality reduction through PCA. Finally, the fourth strategy involved feature selection based on process knowledge.

In the first experiment Table 3, all available features were used for the four regression algorithms. The RF model achieved the best performance, with the lowest MSE (0.0728), RMSE (0.2699), MAE (0.1883), and R^2 of 0.6576, explaining 65.76% of the data variance. This is due to its ability to capture nonlinear relationships. The DT model performed moderately, with R^2 of 0.3047. The LR and SVR models had lower performance, with LR achieving R^2 of 0.1832 and SVR showing the worst performance, with MSE of 0.2068 and R^2 of 0.0277, due to its high sensitivity to parameters and data complexity.

Table 2. Hyperparameter ranges for each regressor.

Regressor	Hyperparameters	Values (grid)
SVR	kernel	{linear, rbf, poly}
	C	{0.1, 1, 10, 100}
	gamma	{scale, auto, 0.001, 0.01, 1}
	epsilon	{0.1, 0.2, 0.5}
	degree	{2, 3, 4}
RF	n_estimators	{50, 100, 200}
	max_depth	{None, 10, 20, 30}
	min_samples_split	{2, 5, 10}
	min_samples_leaf	{1, 2, 4}
	max_features	{auto, sqrt, log2}
LR	fit_intercept	{True, False}
	positive	{True, False}
DT	criterion	{squared_error,friedman_mse}
	max_depth	{None, 5, 10, 20}
	min_samples_split	{2, 5, 10}
	min_samples_leaf	{1, 2, 4}

Table 3. Performance metrics for the models.

Model	MSE	RMSE	MAE	R^2
SVR	0.2068	0.4547	0.3285	0.0277
RF	0.0728	0.2699	0.1883	0.6576
LR	0.1737	0.4168	0.2999	0.1832
DT	0.1479	0.3845	0.2091	0.3047

Next, the independent variables were normalized using the Standard Scaling technique, which adjusts the data to have a zero mean and a standard deviation of 1. Table 4 shows the results obtained after normalization for the four evaluated models. The SVR model showed improvements, with R^2 increasing from 0.0277 to 0.3334. The RF model showed no significant changes. For the LR and DT models, the results remained almost the same.

Table 4. Performance metrics for the models with normalization.

Model	MSE	RMSE	MAE	R^2
SVR	0.1418	0.4265	0.2638	0.3334
RF	0.0728	0.2699	0.1883	0.6576
LR	0.1737	0.4168	0.2999	0.1832
DT	0.1478	0.3844	0.2090	0.3052

Next, the feature selection strategy was applied using PCA, resulting in dimensionality reduction to 10 variables. Table 5 shows the results obtained with feature selection. The RF model showed superior performance

after feature selection, with the R^2 increasing from 0.6576 to 0.7177, along with reductions in MSE, RMSE, and MAE values. The performance gain can be attributed to the elimination of variables that do not contribute to the variability of the target variable. The SVR model also showed improved performance, with R^2 rising from 0.0277 to 0.49597, along with reduced error values. The LR and DT models did not show significant improvements, with R^2 of 0.1403 and 0.1964, respectively.

Table 5. Performance metrics for the models with PCA-based feature selection.

Model	MSE	RMSE	MAE	R^2
SVR	0.1575	0.2368	0.2792	0.49597
RF	0.0919	0.3032	0.2080	0.7177
LR	0.1828	0.4276	0.3094	0.1403
DT	0.1709	0.4134	0.2795	0.1964

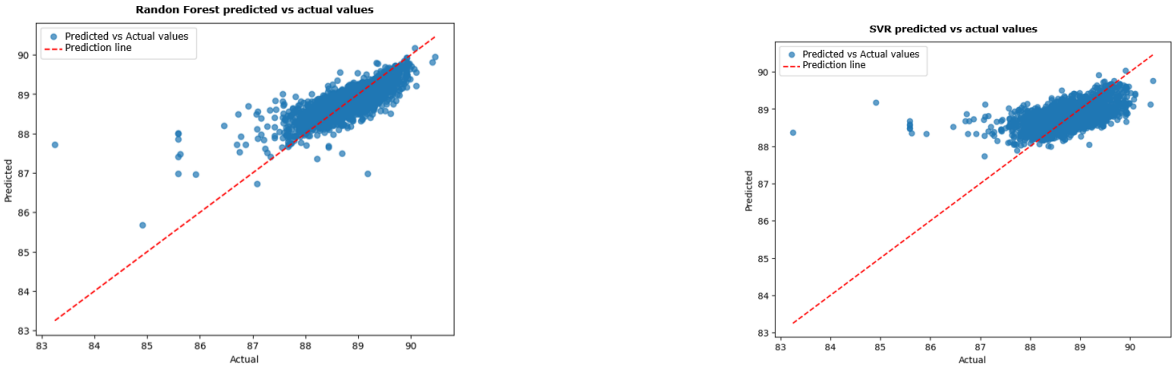
Finally, in the manual feature selection step based on process knowledge, variables were selected considering critical parameters such as pH, reagent flows, and intermediate stage measurements. Table 6 shows the results with this approach. The RF model showed the best performance, with $R^2 = 0.7291$, superior to the R^2 obtained in previous steps. There were also reductions in MSE, RMSE, and MAE, indicating that the selected variables capture important process relationships well. The SVR model obtained $R^2 = 0.4282$, with improvements compared to previous steps, but still inferior to RF. The LR model achieved $R^2 = 0.1723$, showing slightly better performance, while the DT model had $R^2 = 0.3323$, also showing slight improvement.

Table 6. Performance metrics for the models (manual feature selection).

Model	MSE	RMSE	MAE	R^2
SVR	0.1471	0.3836	0.2694	0.4282
RF	0.0728	0.2697	0.1749	0.7291
LR	0.1845	0.4296	0.3118	0.1723
DT	0.1484	0.3852	0.1909	0.3323

4.1 Best Hyperparameter Combinations

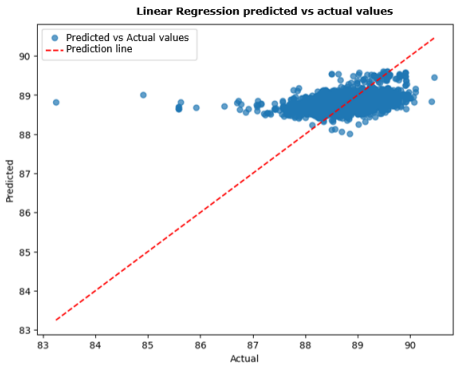
Figure 2 and 3 present the graphs resulting from the best hyperparameter combinations for each regression model considered in this study. It can be observed that for each model, the performance curve evolves according to the adjustments made.



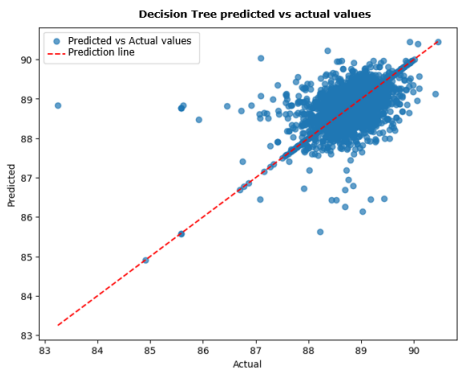
(a) Scatter plot between actual and predicted values for the RF model.

(b) Scatter plot between actual and predicted values for the SVR model.

Figure 2. Scatter plots for the different models evaluated.



(a) Scatter plot between actual and predicted values for the LR model.



(b) Scatter plot between actual and predicted values for the DT model.

Figure 3. Scatter plots for the different models evaluated.

5 Conclusions

This study used the SVR, RF, LR, and DT algorithms to predict whiteness at the D1 bleaching stage in a paper and pulp industry. Experiments were conducted using all features, with normalized features, PCA for feature selection, and feature selection based on process knowledge. The RF model achieved the best results, with R^2 of 0.6576, 0.6576, 0.7177, and 0.7291, respectively, in each experiment. The results showed the RF model's ability to capture nonlinear relationships and interactions between explanatory variables, which are important characteristics in complex systems like cellulose bleaching. Another important point is that the SVR model improved its performance after applying normalization and reducing the number of features.

As future experiments, it is planned to apply other feature selection techniques and also apply some metaheuristics, such as Grey Wolf Optimization (GWO), Differential Evolution (DE), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO), to conduct new tests and assess their performance in the selection of model hyperparameters.

References

[1] T. Ferdous, M. Quaiyyum, and M. S. Jahan. Chlorine dioxide bleaching of nineteen non-wood plant pulps. *Nordic Pulp & Paper Research Journal*, vol. 35, n. 4, pp. 569–576, 2020.

[2] M. Langendorf. Modelagem e simulação de um reator de dióxido de cloro, 2023.

[3] K. Rullifank, M. Roefinal, M. Kostanti, L. Sartika, and others. Pulp and paper industry: An overview on pulping technologies, factors, and challenges. In *IOP Conference Series: Materials Science and Engineering*, volume 845, pp. 012005. IOP Publishing, 2020.

[4] L. V. Gomes. Aplicação de metodologias estatísticas e machine learning para validação de analisadores contínuos para indústria de celulose, 2021.

[5] K. G. F. De Paula. Sensor virtual de alvura em polpa branqueada de celulose baseado em inteligência artificial, 2022.

[6] R. D. Carvalho. Modelo preditivo de alvura final em estágio de branqueamento de polpa celulósica., 2024.

[7] I. C. Favaro. Preditor de alvura utilizando machine learning para controle do processo de branqueamento, 2024.

[8] P. M. Henrique. Predição de variáveis no processo de branqueamento da celulose por meio de redes neurais artificiais., 2024.

[9] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen. *Classification and regression trees*. CRC press, 1984.

[10] L. Breiman. Random forests. *Machine learning*, vol. 45, pp. 5–32, 2001.

[11] A. Smola. Learning with kernels [ph. d. thesis]. *Technische Universitat Berlin*, 1998.

[12] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, vol. 13, n. Feb, pp. 281–305. Disponível em `https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf`, 2012.

[13] F. Pedregosa. Scikit-learn: Machine learning in python fabian. *Journal of machine learning research*, vol. 12, pp. 2825, 2011.

[14] D. Y. Chen. *Análise de dados com Python e Pandas*. Novatec Editora, 2018.