

Machine Learning Models to Identify Anomalies in the Production of Flat Glass

Using Machine Learning to Detect Anomalies in Flat Glass Production

Pedro Gabriel Lima¹

 orcid.org/0009-0000-5582-8209

Noam Eyal Resnick¹

 orcid.org/0009-0007-8930-8915

Denis Leite^{1,2}

 orcid.org/0000-0001-8644-2103

Aristóteles Terceiro Neto³

 orcid.org/0009-0002-2439-6081

Alexandre M. A. Maciel¹

 orcid.org/0000-0003-4348-9291

¹Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, Brasil. E-mail: pqsl@ecomp.poli.br

²Mekatronik I.C. Automação Ltda, Recife, Pernambuco, Brazil.

³Vivix Vidros Planos, Recife, Pernambuco, Brazil.

DOI: 10.25286/repa.v9i1.2770

Esta obra apresenta Licença Creative Commons Atribuição-Não Comercial 4.0 Internacional.

Como citar este artigo pela NBR 6023/2018: Pedro Gabriel Lima; Noam Eyal Resnick; Denis Leite; Aristóteles Terceiro Neto; Alexandre M. A. Maciel. Machine Learning Models to Identify Anomalies in the Production of Flat Glass. Revista de Engenharia e Pesquisa Aplicada, v.9, n. 1, p. 19-27, 2024. DOI: 10.25286/repa.v9i1.2740

RESUMO

Este trabalho apresenta uma proposta inovadora para a previsão de defeitos em processos industriais de refino de vidro. Embora seja um processo complexo com vários pontos que podem causar defeitos, a abordagem atual dos especialistas é apenas reativa, ou seja, eles só podem agir após o dano ter sido causado. Este estudo propõe o uso de dados coletados dos processos industriais de uma empresa real como um estudo de caso para criar modelos de previsão, a fim de identificar uma possível falha antes que ela ocorra. O objetivo é usar o SPC Charter como um modelo de entrega e permitir que os especialistas tomem medidas corretivas preventivas, evitando danos e reduzindo os custos de produção.

PALAVRAS-CHAVE: Inovação; Aprendizado de Máquina; Previsão de Defeitos; Vidro Plano; Controle Estatístico de Processos; Regressão Linear Múltipla.;

ABSTRACT

This work presents an innovative proposal for the prediction of defects in industrial glass refining processes. Although it is a complex process with several points that can cause defects, the current approach of specialists is only reactive, that is, they can only act after the damage has been caused. This study proposes the use of data collected from the industrial processes of a real company as a case study to create prediction models, in order to identify a possible failure before it occurs. The objective is to use multiple linear regression as a model and allow specialists to take preventive corrective measures, avoiding damage and reducing production costs.

KEY-WORDS: Innovation; Machine Learning; Defect Forecast; Flat Glass; Statistical Process Control; Multiple Linear Regression.

1 INTRODUCTION

Glass is a versatile, hard and brittle material essential in human life, as it has applications in different types of industries and in everyday life. Despite the glass industry being a relatively little-known sector of the Brazilian economy, the glass market continues to evolve year after year. Recent industry indicators point out that even with a decrease in glass production, there was an increase in sales and import and export numbers [1][2].

The glass industry is divided into four segments, according to the manufactured product: flat, packaging, domestic and special or technical glass [3]. The flat glass manufacturing process is currently quite complex and has several points susceptible to defects in its production line [4][5].

The technological advancement brought about by the fourth industrial revolution (Industry 4.0) allowed the most recent industrial plants to collect and store a large volume of data, thus ensuring greater quality in their manufacturing process and in their manufactured products.

1.1 PROBLEM DESCRIPTION

Currently, it is already possible to notify that an anomaly has occurred through several sensors installed in the production lines, thus generating a large mass of data [4][5]. With these data, some companies use classic techniques that monitor and improve the production process, such as statistical process control (SPC), or control chart [4][5]. However, these techniques are limited to analyze only facts that occurred in the past, not in the future. Therefore, the current challenge is to use past data to be able to predict future information and minimize possible defects.

1.2 GOALS

The general objective of this work is to develop machine learning models that can predict possible anomalies in the flat glass production process. For this, the following specific objectives were defined:

- Understanding and analysis of existing variables in a flat glass production line.
- Treatment of data collected from a real database of a production line.
- Development and validation of predictive models.

1.3 JUSTIFICATIVE

The global flat glass market size was valued at USD 273.43 billion in 2021 and is forecast to grow at a compound annual growth rate (CAGR) of 4.3% over the years 2022 to 2030 [1]. The international glass market is very promising, taking into account that the export rate of flat glass in Brazil grew by more than 23% in the last year, even with falling production and productivity rates [2]. In this way, it is essential to minimize the amount of defects that occur in a glass creation process, as this can result in significant financial losses for glass companies.

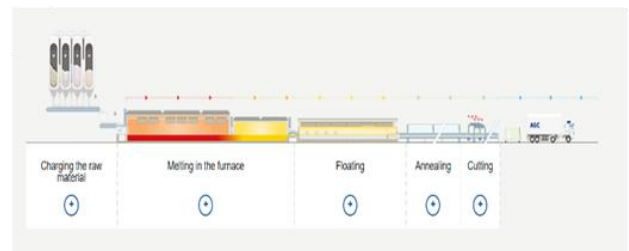


Figure 1 - The Glass Manufacturing Process.

2 THEORETICAL FOUNDATION

The flat glass manufacturing process consists of 5 general steps [6]:

1. **Mixing the raw materials:** The raw materials for the glass are kept in separate containers, the first step is to measure and mix the right amount of each element (sand, limestone, dolomite...)
2. **Furnace melting:** The mixed elements are melted in the furnace at elevated temperatures (1550 °C)
3. **Flotation:** The molten glass comes out of the furnace at high temperatures and rests on an aluminum pool where a glass sheet is formed.
4. **Annealing:** After the flotation process, the glass sheet is placed on a conveyor belt where cooling takes place in a controlled manner to ensure flatness and reduce mechanical defects that can lead to breakage.
5. **Cutting:** The glass sheet is cut into smaller sheets suitable for sale.

Throughout the process there are sensors that collect data and specialists monitor and provide information about the operation of the equipment. There are quality metrics that can be used to define the quality of the glass and the process used for its manufacture [7]:

1. Thickness
2. Flatness
3. Light Transmission
4. Optical Distortion
5. Resistance

2.1 MACHINE LEARNING

Machine Learning (ML) is a subfield of artificial intelligence that focuses on the development of algorithms and statistical models that allow computers to learn from data and perform specific tasks without being explicitly programmed for them [8]. In other words, the machine learns from examples and data, rather than having all the rules coded by a human programmer. There are several types of problems that can be addressed using Machine Learning (ML). Some of the most common types of problems in ML are: Classification, Regression, and Clustering. To solve the proposed problem, grouping and classification algorithms were chosen.

2.1.2 Predictive Algorithms

The multiple linear regression model is the most applied statistical technique for relating a set of two or more variables, the concept of a regression model was introduced to study the relationship between two quantitative variables X and Y.

The concept of a regression model was introduced to study the relationship between two quantitative variables X and Y, The assumed linearity of the relationships makes the models convenient both mathematically and computationally. This simplicity and flexibility have made linear regression the most popular statistical framework across the sciences and standard textbook material.

We first formalize the framework of linear regression. We assume that there are n real-valued observations $y_1, \dots, y_n \in \mathbb{R}$ and corresponding vector-valued observations $x_1, \dots, x_n \in \mathbb{R}^p$ each pair (y_i, x_i) is called a sample. The samples are modeled according to $y_i = x_i^T \beta + u_i$ for all $i \in \{1, \dots, n\}$, where the vector $\beta \in \mathbb{R}^p$ that summarizes the model parameters $\beta_1, \dots, \beta_p \in \mathbb{R}$ is called the regression vector (which is the same across the samples) and $u_1, \dots, u_n \in \mathbb{R}$ the noise (which can be different from one sample to another) [9].

2.2 RELATED WORKS

Carvalho [4] in his study evaluated the application of Statistical Process Control (SPC) in the flat glass production process in a family factory in Brazil. The author used control charts to monitor and analyze process data and compared the results with those obtained by traditional methods of quality control. Overall, the study concludes that SPC can be an effective tool to monitor and control quality in the glass production process. Using control charts allows for early detection of process variations and defects, which can help reduce waste, improve productivity, and improve product quality.

Similarly, the research study by Reis [5] also evaluates the application of SPC in the glass manufacturing process. The author concludes that the use of SPC can help reduce defects, improve production efficiency, and improve overall product quality in the glass industry. Reis [5] also emphasizes the importance of training and proper implementation of SPC techniques to ensure their effectiveness.

The work "Development of Machine Learning models to predict glass quality of melting furnace" [12] is an example of application of Machine Learning in the glass industry, where it was successful in predicting glass defects from the recycling process.

3 MATERIALS AND METHODS

In this work, we will follow the standard CRISP-DM process for data mining methodology proposed in [13]. This model has become the de facto standard for data mining, gaining widespread use by the emerging data mining community.

3.1 BUSINESS UNDERSTANDING

The proposed project was carried out in partnership with the flat glass manufacturer VIVIX, one of the most modern flat glass factories in the world and the only large one in the country with 100% national capital.

To support the development of this work, stakeholders from the Vivix industry and the Mekatronik technology supplier were involved. Board 1 shows the role and role of the stakeholders involved in the project.

Board 1 - Table of Stakeholders Involved.

POSITION	FUNCTION
Industrial Transformation Coordinator	Responsible for quality control of the process
Technical Lead from the Hired company	Responsible for data access

Source: Authors.

3.2 DATA UNDERSTANDING

The data provided by VIVIX is collected by the hired company Mekatronik from the MkAnalytics 4.0 tool, through which manufacturing data is integrated for real-time management. The data used represent a real and significant sample of the base used by the company VIVIX in one of its production lines during the period from 2016 to 2022.

3.2.1 Data Dictionary

Board 2 - Board of Data Dictionary.

NAME	TYPE	DESCRIPTION	EXAMPLE
Parâmetro	String	Parameter name	Amostra 4 - Titulação
ParamId	Int	Parameter Identifier	169
Grupo	String	Group name which the parameter belongs to	Testes - Deposição de prata
Form	String	Form used to collect the parameter's value	CEPVIX Transformados - Espelhos
Valor	Float	Parameter's value at a collection point	709
Maximo	Float	Maximum of normal values	750
Minimo	Float	Minimum of normal values	700
InspectionDateHour	Timestamp	Datetime of value collection	2022-08-15 3:00:00
Range	Float	Variation between current and last collection values	1

Source: Authors.

The data dictionary is a collection of definitions about the data values that will be used in the job. From the definition of a data dictionary, it is possible to standardize the variables used and explain what all variable names and values really mean. The dictionary is described in Board 2.

3.3 MODELING

After the data pre-processing step, one of the best practices for analysis is to create visualizations that can help identify patterns and trends in the data. For this, some visualizations were created using specific tools, such as graphs and tables.

Among the views generated, Figure 1 stands out, which presents the missing data matrix per parameter. This matrix is important because it allows identifying which parameters have missing values and the amount of missing data in each of them. With this information, it is possible to define strategies to deal with missing data, such as filling in missing values or excluding records that are missing excessive data.

In short, creating visualizations is a key step in data analysis, as it provides a better understanding of the data and assists in making evidence-based decisions.

Board 3 - Parameter Alias Table.

PARAMETER NAME	ALIAS
Condutividade Água aplicação líquida	P1
Nível de Vidro - PV	P2
pH Água aplicação líquida	P3
ppm O2 banho 1	P4
ppm O2 banho 2	P5
ppm O2 banho 3	P6
ppm O2 White Martins	P7

Source: Authors.

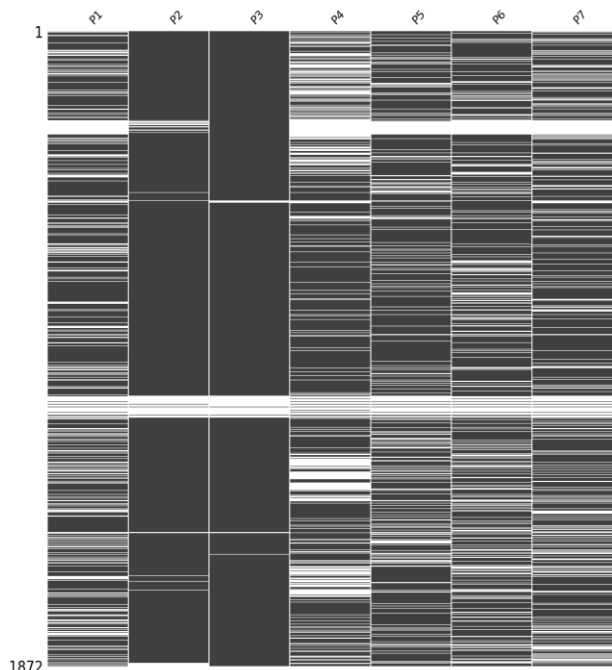
To facilitate visualization, the paper will refer to the parameter name's alias as shown in Board 3, instead of the parameter's full name.

As shown in Figure 2, rows with missing data were removed and columns with few data were also dropped.

Despite the pre-processing performed to normalize the data, we still have a lot of missing data that has been normalized using the mean of the parameter.

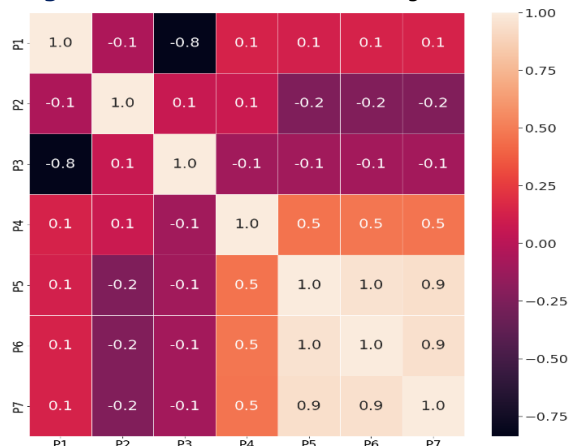
Figure 4 allows the visualization of some parameters that are more related to each other; ppm O2 bath 1, ppm O2 bath 2, ppm O2 bath 3 and Conductivity water liquid application with pH water liquid application.

Figure 2 - Missing Data Matrix.



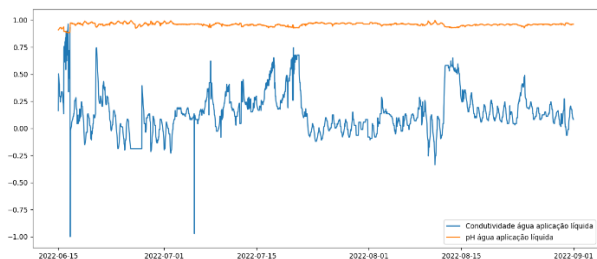
Source: Authors.

Figure 3 - Correlation Matrix Among Parameters.



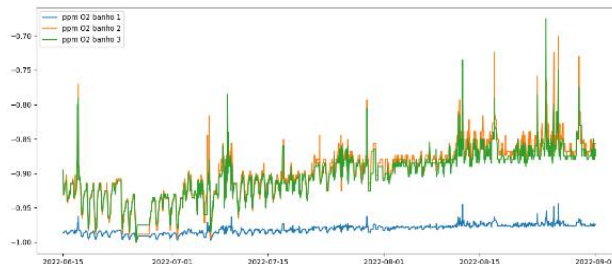
Source: Authors.

Figure 4 - Line graph pH water liquid application x Conductivity water liquid application.



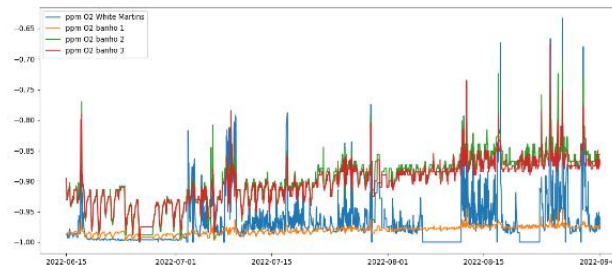
Source: Authors.

Figure 5 - Line graph ppm O2 bath 1 x ppm O2 bath 2 x ppm O2 bath3.



Source: Authors.

Figure 6 - Line graph ppm O2 bath 1 x ppm O2 bath 2 x ppm O2 bath3 x ppm O2 White Martins.



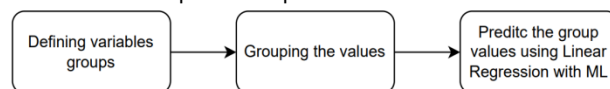
Source: Authors.

These groups were shown graphically in figures 4, 5 and 6 containing the values of the time series of the figures in the same graph.

3.3.1 Predictive Data Analysis

The purpose of predictive analysis is to create a machine model capable of classifying a value from a given sensor. For the forecasting process we use multiple linear regression in the clustered data generated by the clustering algorithms. The complete process is described in the flowchart of Figure 7.

Figure 7 - Flowchart describing the predictive classification experiment process.



Source: Authors.

DBSCAN and K-Means algorithms were used for grouping. Clustering was performed using data from the time series of each variable independently and the time series of some variables together. These variables together were chosen based on their correlation seen in figure 3.

The groups are:

- PPM O2 bath 1, PPM O2 bath 2, PPM O2 bath 3

- PPM O2 bath 1, PPM O2 bath 2, PPM O2 bath 3, PPM O2 White Martins

- Conductivity water liquid application, pH water liquid application

a) Clustering using K-Means

To perform a grouping using K-Means it is necessary to choose the value of K (number of groups). For this, we use the “elbow” method to choose the optimal amount per treated data set.

For datasets dealing with a single parameter we have a one-dimensional time series, the only variable being the sensor value at each time point. Thus, it is possible to order the points sequentially and facilitate the grouping of K-Means. As for the sets dealing with groups of sensors, it is necessary to reduce the dimensionality of the data to 2 so that the distance function has meaning. For this, the PCA (Principal Component Analysis) algorithm was used, which provides the parameters with the greatest significance at a given point.

b) Clustering using DBSCAN

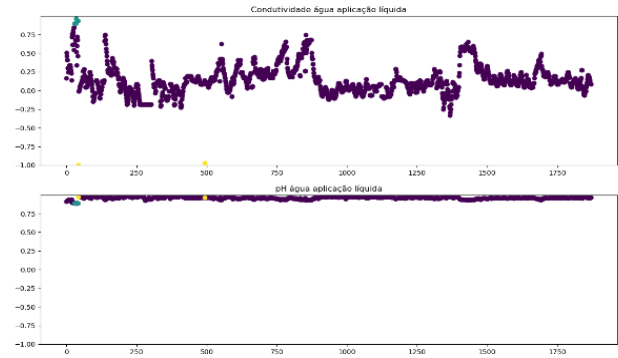
For the DBSCAN algorithm, it is possible change two parameters: the minimum distance between points to define a new group (eps) and the minimum number of points to be considered a group (n). The algorithm was executed 9 times (eps equal to 0.01, 0.05 and 0.1 and n equal to 1, 2 and 3) and after analyzing the resulting graphs, the values of eps equal to 0.05 and n equal to 1 were chosen. Also in K-Means with one-dimensional sets, the ordering of values before execution was used to improve performance, but in this case it was not necessary to reduce dimensionality for larger sets since DBSCAN supports grouping of N-dimensional data.

c) Classification Using K-NN

A classification algorithm needs labels in the training phase, an attribute not present in the original data. For the purpose of data classification, it was used the sets resulting from the clustering phase as labels for each data point. With the labeled data the first step of the classification process was the division of data for training and testing. It was decided 70% for training and 30% for testing with the points being chosen randomly. The K-NN algorithm was run for each previously grouped group and generated the confusion matrices based on the test set for analysis.

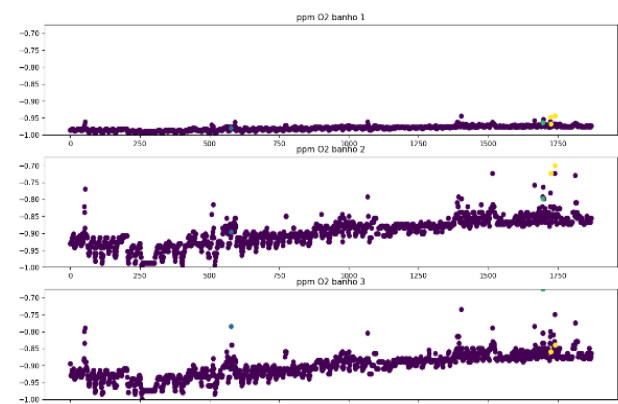
3.3 GRAPHICAL RESULTS

Figure 8 - DBSCAN Clustering Conductivity x pH.



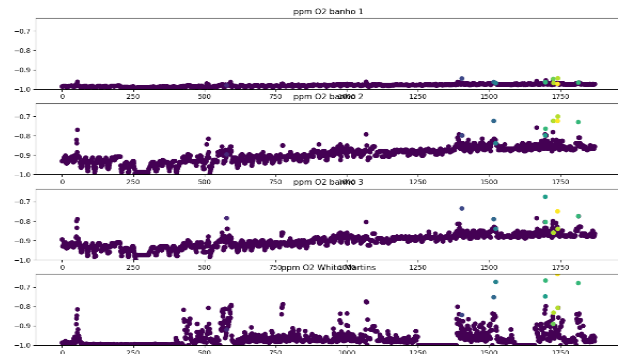
Source: Authors.

Figure 9 - DBSCAN PPM O2 Bath Grouping.



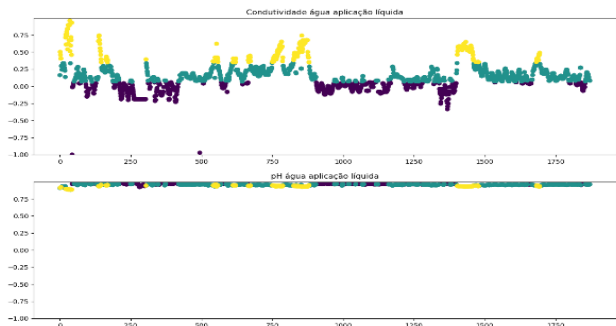
Source: Authors.

Figure 10 - DBSCAN PPM O2 Baths + White Martins grouping.



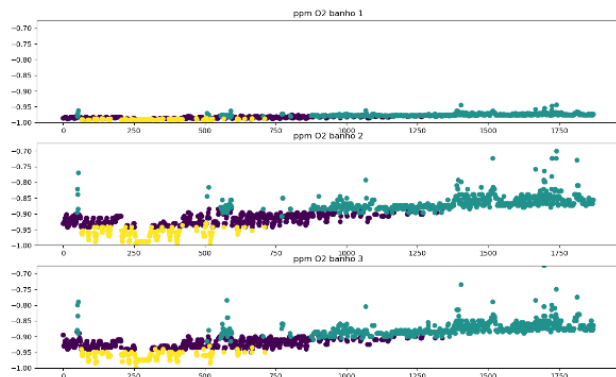
Source: Authors.

Figure 11 - K-Means Clustering Conductivity x pH.



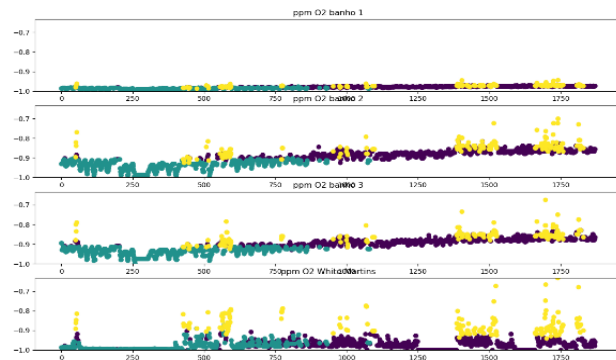
Source: Authors.

Figure 12 - K-Means PPM O2 Bath Grouping.



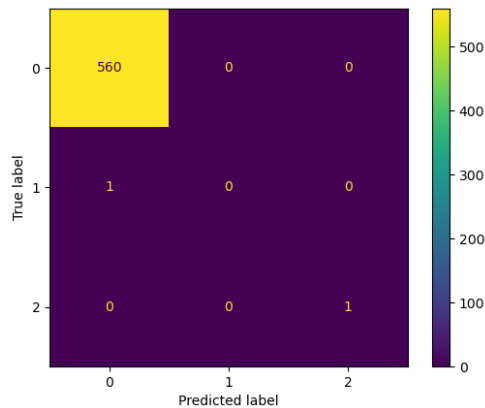
Source: Authors.

Figure 13 - K-Means PPM O2 Baths + White Martins grouping.



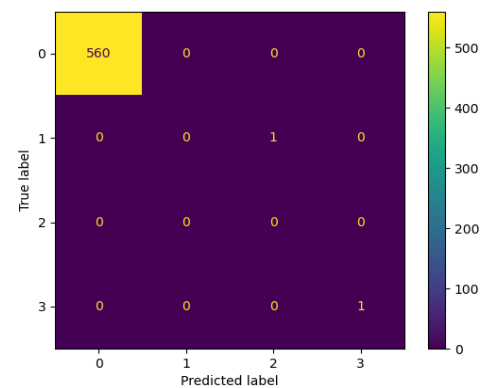
Source: Authors.

Figure 14 - Confusion Matrix - Classification Conductivity x pH DBSCAN.



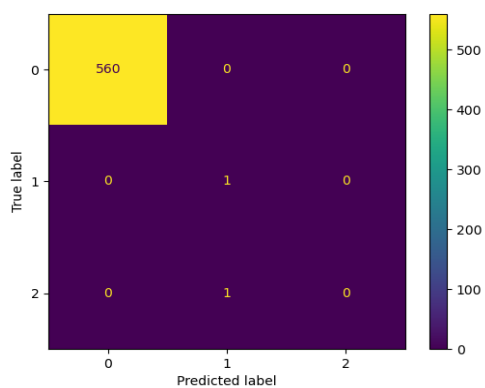
Source: Authors.

Figure 15 - Confusion Matrix - Classification PPM O2 Baths DBSCAN.



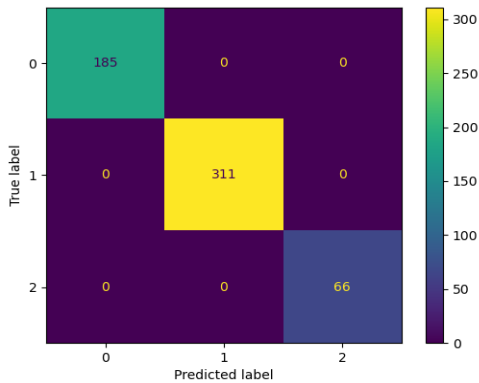
Source: Authors.

Figure 16 - Confusion Matrix - Classification PPM O2 Baths + White Martins DBSCAN.



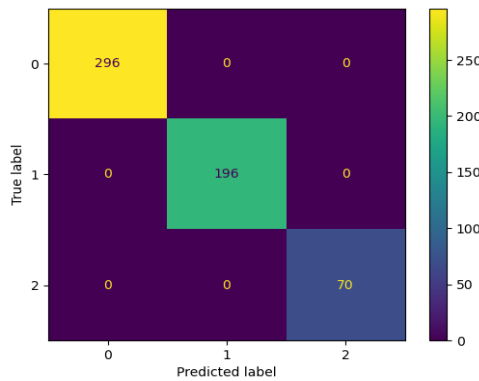
Source: Authors.

Figure 17 - Confusion Matrix – Classification Conductivity x pH K-Means.



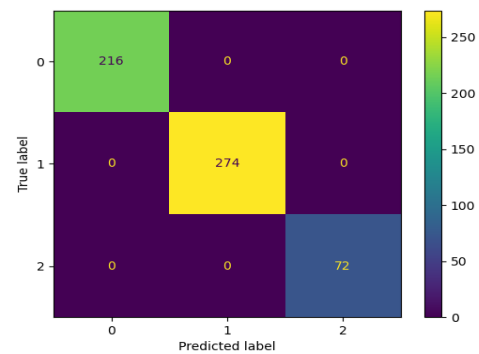
Source: Authors.

Figure 18 - Confusion Matrix - Classification PPM O2 Baths K-means.



Source: Authors.

Figure 19 - Confusion Matrix - Classification PPM O2 Baths + White Martins K-Means.



Source: Authors.

Due to the low mean variation, the DBSCAN algorithm grouped most of the data points into a single group leaving only the outliers in separate groups. K-Means, on the other hand, forced the division of the points into 3 sets (which was the optimal case calculated using the elbow method) dividing the points into low, medium and high. A difference appears in the PPM O2 Baths + White Martins group where DBSCAN finds 4 groups and K-Means finds 3.

In the classification, we can clearly see that a grouping using K-Means generates a much more accurate classification in K-NN, since the accuracy rate in this model was 100%, with DBSCAN as the labeling base, K-NN had a low error. This error can also be caused by the fact that most points belong to a single group in DBSCAN, unlike K-Means where data points are better distributed among groups.

REFERENCES

- [1] Global flat Glass Market, Grand View Research. Available in <<https://www.grandviewresearch.com/industry-analysis/global-flat-glass-market>> Accessed on: 22, mar. 2023..
- [2] Panorama Abravidro 2022, Abravidro. Available in <https://abravidro.org.br/wp-content/uploads/2022/05/Panorama_Abravidro_2022.pdf>. Accessed on 22 mar. 2023.
- [3] Freire, Laura Lúcia Ramos. A Indústria de vidros planos. Caderno Setorial ETENE. Fortaleza: Banco do Nordeste do Brasil, ano 1, n.1, nov.2016. (Série Caderno Setorial ETENE, n.1).
- [4] Carvalho, Alan Queiroz. Metodologia para Controle Estatístico de Processo em uma indústria de transformação de vidro. 28-Set-2018. Available in: <<https://repositorio.ufgd.edu.br/jspui/handle/prefix/2127>>. Accessed on: 22, mar. 2023.
- [5] Reis, Mariana de Oliveira. Aplicação do Controlo Estatístico de Processos na Indústria Vidreira. Fev-2017. Available in <<https://run.unl.pt/handle/10362/21793>>. Accessed on 22, mar 2023.

5. CONCLUSIONS

This work was completed with the presentation of its results to the stakeholders and, the implementation of the models was the responsibility of the company.

- [6] Sand Glass, AGC. Disponível em: <<https://www.agc-glass.eu/en/products/sand-glass>> Accessed on: 22 mar. 2023.
- [7] Costa, C.S.M.P. Controlo da Qualidade de Fachadas em Vidro. Available in <<https://fenix.tecnico.ulisboa.pt/downloadFile/1407770020546310/ControloQualidadeFachadasVidro.pdf>>. Accessed on: 22, mar. 2023.
- [8] Mitchell, T. M. (1997). Machine Learning. McGraw-Hill Science/Engineering/Math. Available in: <<https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf>>.
- [9] PALMA, L. F. AGRUPAMENTO DE DADOS: K-MÉDIAS. 2008 - UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA. Available in: <https://www2.ufrb.edu.br/bcet/components/com_chronoforms5/chronoforms/uploads/tcc/20190604200511_2018.2_TCC_Luann_Farias_Palma_Agrupamento_de_dados_-_K_medias.pdf> Accessed on: 27 mar. 2023.
- [10] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. 2017. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. ACM Trans. Database Syst. 42, 3, Article 19 (September 2017), 21 pages. Available in: <<https://doi.org/10.1145/3068335>>
- [11] Ferrero, C. A. Algoritmo kNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia. USP – São Carlos 2009.
- [12] Rodrigues, F. J. L. S. N. (2022). Development of Machine Learning Models to Predict Glass Quality of Melting Furnace. Available in: <<https://repositorio-aberto.up.pt/handle/10216/142742>>.
- [13] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide..