algorithms

MDPI

Article

Analysis of the State and Fault Detection of a Plastic Injection Machine—A Machine Learning-Based Approach

João Costa [1], Rui Silva [2], Gonçalo Martins [2], Jorge Barreiros [1,3] and Mateus Mendes [1,3,4,*]

[1] Coimbra Institute of Engineering, Polytechnic University of Coimbra, Rua Pedro Nunes-Quinta da Nora, 3030-199 Coimbra, Portugal; a2022143368@isec.pt (J.C.); jmsousa@isec.pt (J.B.)
[2] Sinmetro LDA, Rua dos Costas, Lote 19, Loja 74, R/C, 2415-567 Leiria, Portugal; rui.silva@sinmetro.pt (R.S.); gmartins@sinmetro.pt (G.M.)
[3] RCM^{2+} Research Centre for Asset Management and Systems Engineering, Rua Pedro Nunes, 3030-199 Coimbra, Portugal
[4] Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, 3030-290 Coimbra, Portugal
* Correspondence: mmendes@isec.pt

Abstract

Predictive maintenance is essential for minimizing unplanned downtime and optimizing industrial processes. In the case of plastic injection molding machines, failures that lead to downtime, slowing production, or manufacturing defects can cause large financial losses or even endanger people and property. As industrialization advances, proactive equipment management enhances cost efficiency, reliability, and operational continuity. This study aims to detect machine anomalies as early as possible, using sensors, statistical analysis and classification models. A case study was carried out, including machine characterization and data collection. Clustering methods identified operational patterns and anomalies, classifying the machine's behavior into distinct states, validated by company experts. Dimensionality reduction with PCA contributed to highlighting salient features and reducing noise. State classification was carried out using the resulting cluster data. Classification using XGBoost achieved the best performance among the machine learning models tested, reaching an accuracy of 83%. This approach can contribute to maximizing plastic injection machines' availability and reducing losses due to malfunctions and downtime.

Keywords: predictive maintenance; plastic injection machine; fault detection; DBSCAN; PCA; XGBoost

check for updates

Academic Editor: Sándor Szénási

Received: 16 July 2025
Revised: 11 August 2025
Accepted: 13 August 2025
Published: 18 August 2025

Citation: Costa, J.; Silva, R.; Martins, G.; Barreiros, J.; Mendes, M. Analysis of the State and Fault Detection of a Plastic Injection Machine—A Machine Learning-Based Approach. *Algorithms* **2025**, *18*, 521. https://doi.org/10.3390/a18080521

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Unplanned downtime in industrial machinery poses a significant challenge to manufacturing efficiency. Unexpected machine failures can lead to severe financial losses, production delays, and increased maintenance costs. In plastic injection molding, where precision and continuity are crucial, predicting and preventing such failures is essential to maintain productivity and reduce operational risks. Failures in these machines can stem from various sources, including mechanical wear, electrical faults and human errors, making early detection a complex but necessary task.

Plastic Injection Molding (PIM) machines are heavy industrial equipment that require specialized maintenance interventions. Ideally, they operate continuously for many hours, days, or even weeks in order to maximize production and minimize setup time. Nonetheless, they can suffer numerous problems that require fast, qualified maintenance

interventions. A number of possible problems and solutions are discussed below. The problems include fixed plate deformation, obstructions in the injection system, pressure and temperature variations, mold cooling failures, and other common challenges in this type of equipment.

Recent advancements in data science and machine learning have enabled the development of predictive maintenance strategies aimed at reducing unplanned stoppages. Anomaly detection has shown promising results in identifying patterns associated with machine failures, allowing early interventions [1]. However, despite the growing body of research, many industrial applications still rely on reactive maintenance, leading to inefficiencies and high costs. This study aims to contribute to bridging this gap for this particular type of equipment, by applying data-driven clustering and classification methods to detect failures in a plastic injection machine at a very early stage.

The main objectives of this research are as follows: (1) to analyze machine behavior through literature and dataset analysis; (2) to group and identify states through clustering techniques and (3) to classify the data to be used later in fault detection in a real-world industrial context. The study follows a structured methodology, starting with data collection and machine characterization, followed by clustering analysis and classification. The present paper is an expanded and revised version of [2].

The remainder of this paper is organized as follows: Section 2 describes the operating principles and main components of the plastic injection machine. Section 3 reviews the current state of the art. Section 4 outlines the data and methodology adopted in this study. Section 5 discusses the results, and Section 7 presents the conclusions and future research directions.

2. Plastic Injection Machines

Plastic Injection Machines are large and complex equipment, heating the plastic and passing molten plastic through more or less complex molds for production. Figure 1 shows pictures of the PIM subject of study. Table 1 shows the components that make up the machine, its location and a brief description of what it does.



(**a**) (**b**)

Figure 1. Plastic Injection Machine.

Table 1. Name, location, function and appearance of the most important parts of a PIM.

Name	Location	Function	Appearance
Barrel	In the middle of the machine, surrounds the screw	Captures and mixes the plastic, maintaining uniformity and pressure during injection	Long and cylindrical; pointed at one end; covered by heating bands
Gates	Between channels and mold cavity	Control the flow of plastic going into the cavity	Small openings
Heaters	Around the barrel	Heats the barrel to melt the plastic	Metal bands around the barrel
Hopper	At the top of the machine	The place where plastic is introduced in its solid form. Sometimes contains a dryer to remove moisture	Funnel (Conical Shape)
Mold	Connected to the Fixed and Movable Platens	Gives the final shape to the molten plastic, forming the desired part	Metal block, typically two parts, containing a cavity, cooling channels, and vents
Mold Cavity	Inside the mold	Creates the final shape of the mold and contains cooling cavities	A space in the mold that forms the desired shape
Movable Platen	Connected to one half of the mold	Presses one half of the mold against the other during part manufacturing and releases it once the part is finished and cooled	Flat, rectangular, and metallic
Nozzle	At the end of the barrel, near the mold	Directs the plastic into the mold cavity and prevents it from cooling before entering the mold	Tapered outlet
Pellets	Inside the barrel and nozzle	Plastic material inserted into the machine for molding. Common plastics include ABS, PP, or Nylon, sometimes with additives	Small plastic granules
Reciprocating Screw	Inside the barrel	Mixes and compresses the plastic coming from the hopper	A metal spiral
Runners	Inside the mold	Directs the molten plastic, maintains a uniform flow, and reduces plastic waste	Long and narrow channels
Sprue	Central channel that connects the nozzle to the entry point of the molten plastic	Directs the molten plastic from the nozzle to the runners	Cone-shaped channel

2.1. Steps in Injection Molding

Injection molding is a critical manufacturing process in the plastics industry, capable of producing high-volume, high-precision components for a wide range of applications. Here is a detailed description of the plastic injection molding process.

1. **Feeding and Preparation of Raw Material** —The process begins with loading plastic raw materials, usually in the form of small pellets or granules, into the machine's hopper. These thermoplastic materials, such as polypropylene (PP), polyethylene (PE), polystyrene (PS), polycarbonate (PC), or acrylonitrile butadiene styrene (ABS), can be mixed with additives such as colorants, UV stabilizers, or reinforcing agents to alter the appearance of the part. The hopper feeds the material into a heated cylinder by gravity.
2. **Plasticizing the Material**—Inside the cylinder, a reciprocating screw rotates and moves the plastic forward. As the material advances, it is gradually heated by both external electric heaters surrounding the barrel and the frictional heat generated by the shearing action of the screw. This combined heat melts the pellets into a homogeneous, viscous molten state called a pillow. At the front of the barrel, a check valve prevents the molten plastic from flowing backwards, ensuring that the full volume of material is injected forward when needed.
3. **Injection into the Mold**—Once a sufficient amount of molten plastic has accumulated in front of the screw (a process known as "shot size preparation"), the screw stops rotating and moves forward, acting as a plunger. It forces the molten plastic through the nozzle and into the mold cavity under high pressure. The mold, which is tightly closed under tons of pressure, contains the negative shape of the final part. The high injection pressure ensures that the molten plastic fills every detail of the cavity, including thin walls, small features and complex geometries.

4. **Cooling and Solidification**—Once the mold is filled, the cooling phase begins. The plastic begins to solidify when it comes into contact with the cooler walls of the mold. Most molds have an integrated cooling system, usually channels that circulate water or oil, to control and accelerate the cooling process. Cooling time is crucial and depends on the material, part thickness and mold design. Adequate cooling ensures dimensional stability and prevents problems such as warpage or sink marks.
5. **Mold Opening and Part Ejection**—After sufficient cooling time, the mold opens and ejector pins push the solidified part out of the cavity. In multi-cavity molds, several parts can be ejected simultaneously. Sometimes, robotic arms or conveyors assist in removing and organizing molded parts, especially in automated production lines. Once the part is ejected, the mold closes again and the next cycle begins. A complete injection molding cycle can take anywhere from a few seconds to a few minutes.
6. **Post-Molding Operations**—Although injection molding produces parts that are nearly finished in shape, some post-processing may be required. In this case, the piece is automatically inspected for defects, then transferred to an oven where it is hardened and strengthened before being stored.

2.2. PIM Problems and Failures

PIMs are complex devices that require careful optimization and maintenance to operate smoothly and safely. According to industry operators who perform daily maintenance on injection molding machines, a number of issues are common [3].

One of the most common problems in PIM machines is the obstruction of the injection devices due to plastic residue accumulation or contaminants. This issue may arise from improper cleaning procedures, low-quality materials, or incorrect processing conditions. Additionally, incorrect material selection and improper processing parameters can lead to material degradation, overheating, and nozzle blockage. Regular maintenance and cleaning of injection devices are essential to prevent obstructions. Operators should use appropriate cleaning agents to remove accumulated residues. Using compatible materials and optimizing processing parameters also significantly reduces obstruction risks and enhances machine efficiency.

The Mold Cooling System is another sensitive part. Inefficient mold cooling can result from poor cooling channel design and inadequate heat dissipation capacity, leading to uneven temperature distribution, extended cycle times, higher defect risks, and reduced product quality. Optimizing the layout and configuration of cooling channels ensures uniform temperature distribution throughout the mold. Proper positioning, consistent channel diameters, and appropriate spacing improve cooling efficiency, leading to better productivity and higher-quality injected parts.

Pressure and Temperature Variations in Injection can also be a source of problems. Pressure variations in the injection process are influenced by changes in material viscosity and polymer flow behavior. Temperature fluctuations and humidity content can alter material flow properties, causing injection pressure instability. Controlling material properties through proper storage, handling, and humidity monitoring is critical. Injection parameters must be adjusted based on material type, product shape, and mold design. Proper pressure calibration minimizes variations and ensures consistent production quality.

Part Adhesion and Removal Issues are another typical problem. Part adhesion to the mold can result from inadequate use of mold release agents. Insufficient extraction force or poorly designed ejector pins may lead to deformation or incomplete removal of molded parts. Applying mold release agents correctly and optimizing surface finish reduces friction and facilitates part removal. Adjusting the extraction force and modifying the ejector pin design ensure efficient part ejection, minimizing defects and material waste.

The Hydraulic System can also cause frequent failures. Hydraulic failures, including oil leaks due to worn seals, damaged hoses, or loose connections, can significantly impact machine performance. Insufficient hydraulic pressure caused by pump failures, valve blockages, or fluid contamination also disrupts operation. Regular hydraulic system inspections prevent oil leaks and ensure all connections are secure. Monitoring pressure gauges and performing preventive maintenance on pumps, valves, and filters, help maintain optimal hydraulic performance and avoid machine failures.

The Electrical and Control System must also be monitored to prevent potential failures. Electrical issues, such as power fluctuations or wiring failures, can cause unexpected machine shutdowns and production delays. Malfunctions in the control system, including software or hardware issues, can impact machine performance and process regulation. Installing surge protectors, voltage regulators, and uninterruptible power supplies (UPS) stabilizes the power supply and prevents sudden failures. Routine inspections of electrical wiring, terminals, and connectors help prevent malfunctions. Diagnosing and resolving control system issues in advance ensures stable operation and high-quality production.

By addressing these issues through proper maintenance and optimization strategies, the efficiency, durability, and quality of plastic injection molding machines can be significantly improved.

3. Literature Review

A comprehensive literature review was conducted, searching scientific databases such as Scopus, IEEE Xplore, and ScienceDirect. Keywords used included "plastic injection molding failures," "Predictive Maintenance in plastic injection machine," "mold cooling efficiency," and "hydraulic system failures in injection molding machines." The papers were selected based on their relevance to fault identification, predictive maintenance, and operational optimization in plastic injection molding (PIM) machines.

3.1. Fault Detection

Fault detection is an important area of research, and different methods have been applied in the past. Abid et al. [4] present a survey of different fault detection methods. Datta et al. [5] also present a review focused on pipeline methods. Nor et al. [6] focus specifically on data-driven fault detection for chemical processes. Park et al. [7] present a review aiming to cover fault detection in industrial processes in general. Yu et al. [8] focus on deep learning based methods. This shows the importance of the area of fault detection in industrial settings. However, in the specific case of PIM, there is still a need to develop context-specific methods and algorithms.

Zhang and Alexander [9] explore the use of pressure signals measured directly in the mold cavity as a source of information for identifying faults in the injection process. The study's main hypothesis is that different types of faults, such as incomplete filling, the presence of bubbles, or burn marks, manifest themselves in a characteristic way in the cavity's pressure profile throughout the molding cycle. To process these signals, the authors apply Principal Component Analysis (PCA) as a dimensionality reduction technique, facilitating the extraction of relevant patterns. They then use wavelet transforms to decompose the signals and extract representative dynamic features. These features are then fed into an artificial neural network, which is responsible for classifying the different types of faults based on previously labeled examples. The study demonstrates that cavity pressure signals provide a robust basis for early fault diagnosis, with the potential to significantly improve the quality of the parts produced and the efficiency of the process. In addition, Zhang and Alexander emphasize the importance of approaches based on real process data for

continuous and adaptive monitoring, anticipating current trends in predictive maintenance and intelligent manufacturing.

Kozjek et al. [10] explore the application of data mining techniques for fault diagnosis in PIM processes. The study examines how algorithms such as J48, Random Forest, and k-Nearest Neighbors can identify operational failure patterns, aiming to improve production efficiency and maintenance planning. This study demonstrates how Data Mining (DM) can uncover defective operational patterns and improve quality and productivity in PIM. DM provides an alternative to traditional methods like statistical process control and experimental design by discovering significant patterns and relationships in industrial data. The dataset consists of six months of operational records from five European PIM machines: Process parameter logs input and output values per cycle, Alarm logs, and Tool change records. Approximately 2.2 million cycles were recorded, with over ten different tools used per machine. Alarms were classified into First-degree alarms (immediate machine stoppage) or Second-degree alarms (non-critical issues). The final dataset contained 62 numerical attributes and a binary classification label. To ensure an unbiased classification baseline of 50%, the dataset included equal instances of normal and faulty cycles. A Python-based system was developed for data processing, including reading, encoding, filtering, transformation, and querying. The results indicate that the J48, Random Forest, JRip, Naïve Bayes, and k-NN algorithms effectively identify patterns related to defective operating conditions. All tested algorithms outperformed the standard accuracy benchmark of 50.0%. J48, Random Forest, JRip, and Naïve Bayes exhibited higher classification accuracy than k-NN, as they leverage target attribute information during model induction. Random Forest offers the advantage of easy parameter tuning and relatively high predictive performance. However, its interpretability is limited, whereas J48 and JRip provide interpretable rule-based models. Key parameters affecting defective operating conditions in Injection Molding Units (UMSs) for a selected tool and machine include temperature, opening time, and cycle time.

Ke and Huang [11] propose an approach for classifying the quality of injection-molded components through machine learning, focusing on Multilayer Perceptron (MLP) models. The study relies on the extraction of quality indices derived from process data, primarily internal pressure and other variables measured during the production cycle. The research evaluates different classification schemes by varying the number of quality categories to assess their impact on model performance. The dataset includes samples reflecting natural process variability and defects, which are quantitatively transformed into indices to serve as model inputs. Results indicate that the MLP model effectively distinguishes multiple quality levels with high accuracy, although the classification performance is influenced by the granularity of the defined classes. The study emphasizes the utility of classification techniques as efficient tools for process monitoring in injection molding, enabling early detection of states that may compromise the final product's quality. This research complements machine state analysis by linking process monitoring with product quality, demonstrating how classification models can support the maintenance of operational excellence in injection molding processes.

Aslantaş et al. [12] propose a machine learning model for identifying and predicting failures in plastic injection molding (PIM) machines. They research sensor data analysis to anticipate failures before they occur, enabling more efficient maintenance planning and preventing unexpected production downtimes. To achieve this goal, classification algorithms such as Random Forest (RF) and Extreme Gradient Boosting (XGBoost) were applied to data collected from three machines in a home appliance manufacturing plant. Additionally, the SMOTE technique was employed to balance data distribution and enhance model accuracy. This study introduces a machine learning model designed to predict

failure types in PIM machines based on sensor data. The model is constructed using classification algorithms to analyze sensor readings and forecast machine failures before they occur. Failure identification in PIM machines follows multiple stages, from raw data collection to classification. The estimation of Remaining Useful Life (RUL) plays a crucial role in identifying machines with a higher likelihood of failure. RUL, defined as the time interval between the present moment and the point of failure or maintenance requirement, is computed using historical maintenance and failure data. Several factors, such as clamping force, cycle time, and oil temperature, influence this interval. These data are obtained through sensors, process parameter logs, alarms, and maintenance records. Two classification algorithms were employed to develop predictive maintenance models: Random Forest and XGBoost. Raw sensor data alone do not sufficiently describe failure types, making feature extraction a crucial step. Extracted features include minimum, maximum, mean, skewness, kurtosis, and entropy. In this study, raw data were aggregated into 60-minute intervals, and features were automatically extracted for each interval. Data were collected from three plastic injection molding machines in a Turkish home appliance factory. The dataset sizes range from 205,000 to 913,000 records, covering the period from 2018 to 2021. A separate predictive model was constructed for each machine using variables such as clamping force, cycle time, and oil temperature. Experimental results indicate that RF and XGBoost, with and without SMOTE, demonstrated strong performances. XGBoost with SMOTE achieved the best performance, with an average accuracy of 98%. The highest F-score was also obtained using XGBoost with SMOTE (0.98), compared with RF (0.88) and other variations. This study estimated the types of failures in plastic injection molding machines, addressing the issue as a multi-class classification problem for predictive maintenance. Handling missing values and extracting relevant features were crucial steps in the process. The XGBoost and Random Forest (RF) algorithms were evaluated, with XGBoost demonstrating superior performance, especially when combined with the SMOTE technique. The achieved accuracy was 98% with XGBoost + SMOTE.

3.2. Maintenance and Process Optimization

Pierleoni et al. [13] discuss the evolution of maintenance strategies in industrial settings, emphasizing the importance of Predictive Maintenance (PdM) within Industry 4.0. Effective maintenance management is critical to avoid unexpected failures, which can result in high costs and negatively impact product quality and system reliability. Despite its importance, many companies have yet to adopt advanced strategies to optimize their maintenance budgets. The study focuses on the application of PdM in four electric plastic injection molding machines equipped for Industry 4.0. Analyzing failure patterns and monitoring process variables, a methodology was developed to predict component wear and prevent complete breakdowns. Data were collected from machine sensors, measuring variables such as injection pressure, plasticization volume, cycle time, temperature, and motor force. Since most companies follow preventive maintenance strategies that avoid obvious failures, obtaining real defect data was challenging. Instead, failures were inferred based on qualitative analysis of historical data and expert interviews. The collected data were labeled into two categories: 'optimal' operation and 'functional limit', allowing the development of predictive models for twelve different machine-product combinations. A key limitation of the study is the scarcity of real failure data, which affects the accuracy of adverse condition classifications.

Pérez-Mora et al. [14] propose a study of "Plastic Injection Molding Process Analysis: Data Integration and Modeling for Improved Production Efficiency". They present a data-driven framework aimed at improving the efficiency and quality control of plastic injection molding through the application of classification models. The study integrates real-time

data acquisition from injection molding machines using a custom IoT-based DAQ system, capturing key variables such as injection time, cycle time, and mold pressure. A thorough exploratory data analysis was conducted to identify patterns and correlations that influence product quality and process stability. The core of the methodology involves the application of supervised classification algorithms, specifically logistic regression and random forest, to detect anomalies and predict process behavior. These models were trained to classify whether a production cycle met optimal operational conditions, with the random forest model achieving a remarkably high accuracy of 99.5%, and logistic regression reaching 97%. Feature importance analysis revealed that cycle time was the most critical variable influencing classification performance. The models were embedded into an intelligent agent system capable of real-time monitoring and adaptive decision-making. A graphical interface was developed to facilitate operator interaction, providing visual analytics and actionable recommendations based on model outputs. The integration of classification modeling with real-time feedback mechanisms demonstrated tangible improvements in production reliability, energy efficiency, and defect reduction. This study underscores the potential of machine learning classification in enabling predictive, responsive, and intelligent manufacturing systems.

Khatir et al. [15] propose a hybrid methodology combining analytical models and Artificial Neural Networks (ANN) to predict the deflection behavior of tapered cellular steel beams. The study aims to overcome the limitations of purely analytical or numerical methods by integrating data-driven learning with classical structural mechanics. The authors construct an ANN model trained on a dataset generated from parametric simulations and validated against analytical formulations. Key input features include geometric parameters, material properties, and loading conditions. The ANN is optimized to estimate mid-span deflections under varying configurations with high accuracy. Results demonstrate that the proposed ANN model achieves excellent predictive performance while significantly reducing computational time compared with traditional finite element approaches. The study highlights the potential of machine learning to enhance structural performance assessment and support engineering decision-making in complex design scenarios. In further advances, Khatir et al. [16] present a novel hybrid method combining Frequency Response Function (FRF) inputs with an ANN, optimized by the Reptile Search Algorithm (RSA) for damage detection in steel beams. The central hypothesis is that FRF-based features carry sensitive information about structural degradation, which can be more effectively exploited when the neural network training is optimized via a metaheuristic algorithm to avoid local minima and enhance convergence. Experimental modal testing provided FRF datasets under controlled notch-induced damage scenarios, which were validated using finite element modeling. These features were fed into an ANN whose training was fine-tuned by RSA, and compared against alternative hybrid models involving ANN + Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), and Genetic Algorithms (GA). Results show that the ANN–RSA hybrid outperforms comparison models, achieving regression values approaching 1.0, reducing prediction errors by approximately 15%, and consistently maintaining high accuracy (error margins below 1%) across damage types. In addition, the RSA-optimized network converged in fewer iterations, demonstrating both higher precision and computational efficiency. This study underlines the effectiveness of integrating bio-inspired optimization algorithms like RSA with vibration-based ML methods to improve structural health monitoring performance, suggesting potential applicability in industrial machine fault diagnosis frameworks. Additionally, Khatir et al. [17] present a novel hybrid optimization method, PSO YUKI, for detecting double cracks in Carbon Fiber Reinforced Polymer (CFRP) cantilever beams. The study investigates the hypothesis that combining PSO with a novel YUKI algorithm enhances sensitivity and accuracy

when identifying multiple crack locations in composite structures. Experimental vibration data were collected from beams with artificially induced double cracks under controlled conditions. These modal response measurements served as input to the hybrid PSO-YUKI algorithm, which optimized a machine learning-based classifier to accurately detect and locate cracks. The results show that the PSO-YUKI hybrid approach achieves significantly higher detection accuracy and reliability compared with standalone PSO or traditional crack detection techniques. Moreover, the methodology offers improved convergence speed and robustness in handling complex damage patterns involving multiple cracks. This work underscores the potential of bio-inspired and hybrid optimization algorithms to enhance structural fault diagnosis, suggesting applicability to broader contexts, including industrial machinery maintenance and real-time condition monitoring.

4. Data and Methodology

In this study, an unsupervised learning approach was used. This is a type of machine learning that analyzes data without human supervision. Unlike supervised learning, it works with unlabeled data, allowing models to discover patterns and insights independently. This approach was employed to analyze the dataset, using Principal Component Analysis (PCA) for dimensionality reduction and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) for clustering. This method is particularly useful for organizing large volumes of information into clusters and identifying previously unknown patterns. The methodology followed a structured workflow involving data preprocessing, feature extraction, and clustering to uncover meaningful patterns in the data. Figure 2 illustrates the steps of the process that was followed.

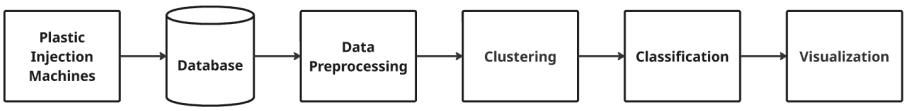


Figure 2. Diagram illustrating the steps of the process used followed.

4.1. Dataset Description and Data Preprocessing

The dataset consists of sensor values recorded by the machine over time. It contains failure data, although it does not show what kind of failure it was. Due to its high-dimensional nature, a preprocessing step was necessary to reduce the dataset size while preserving crucial information to improve clustering performance.

This dataset contains records from April 2024 to January 2025, covering a sampling period of 274 days. The records were sampled at a frequency of seconds, based on a *best effort* approach. There is a total of 48 million records across 129 variables and seven machines. The subset for analysis was subsequently reduced to 19 critical variables, containing 3,242,214 records from a single machine, Machine 76. Table 2, as well as Figures 3 and 4, show the statistics of the variables used for clustering and the plots of the records for two of the variables utilized.

Table 2. Description of variables with their respective statistical values.

Variable	Number of Values	Mean	Max	Min	Std. Dev.
Specific_injection_pressure_peak_value	170,645	1710.18	2071.20	0.0	363.77
Switchover_volume_actual_value	170,642	159.31	502.76	0.0	29.71
Specific_pressure_at_switch_over	170,644	1631.39	2000.60	0.0	338.75
Specific_holding_pressure_peak_value	170,642	1629.60	1997.90	0.0	337.83
Material_cushion_smallest_value	170,642	35.22	495.08	0.0	23.65
Material_cushion_after_holding_pressure	170,643	55.38	697.64	0.0	25.63
Material_cushion_end_holding_pressure	170,642	36.23	538.11	0.0	28.06
Shot_volume	170,644	991.61	1096.76	0.0	161.45
Injection_time	170,644	3.72	20.50	0.0	0.42
Speed_peak_value	170,642	0.43	0.65	0.0	0.06
Specific_back_pressure_peak_value	170,644	78.77	301.30	0.0	10.78
Plasticizing_volume	170,642	1049.19	1132.39	0.0	161.97
Plasticizing_time	170,644	14.83	330.21	0.0	2.76
Clamping_force_peak_value	170,644	5077.88	5255.10	0.0	465.12
Mold_opening_stroke_peak_value	170,642	644.68	652.20	0.0	57.50
Cycle_time	170,644	59.11	2335.19	0.0	25.31
Cooling_time	170,642	24.48	333.04	0.0	3.57
Cycle_time_holding_pressure	170,641	14.41	20.00	0.0	2.13
Barrel_Temperature_Zone_Actual_Temperatures_INJ1_Z01	170,641	200.39	280.20	22.90	27.33

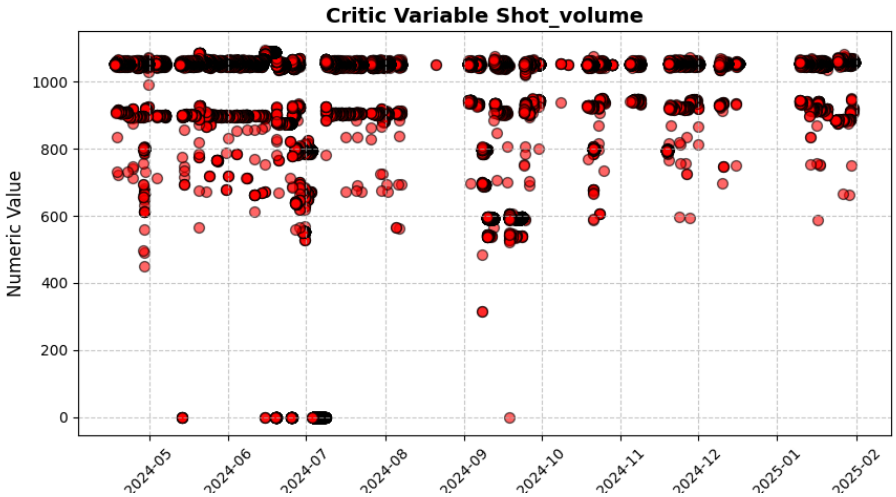


Figure 3. Critical variable analysis of shot volume for Machine 76. It is evident from the chart that there are discrepant values, different states and also missing values.

As we can see from these two graphs Figures 3 and 4, the data sampling period together with where the data have the highest frequency, and the areas where black stands out are due to the high number of points. The dataset had areas of missing records in all the critical variables used. As can be seen in Figure 3, it is possible to observe the absence of records for some time intervals. The values were replaced using the forward-fill (ffill) method in the Python pandas library, which drags the last valid value known until a new valid value is found. This approach was chosen because sensor values often remain constant during operation, making the previous value a reasonable estimator.

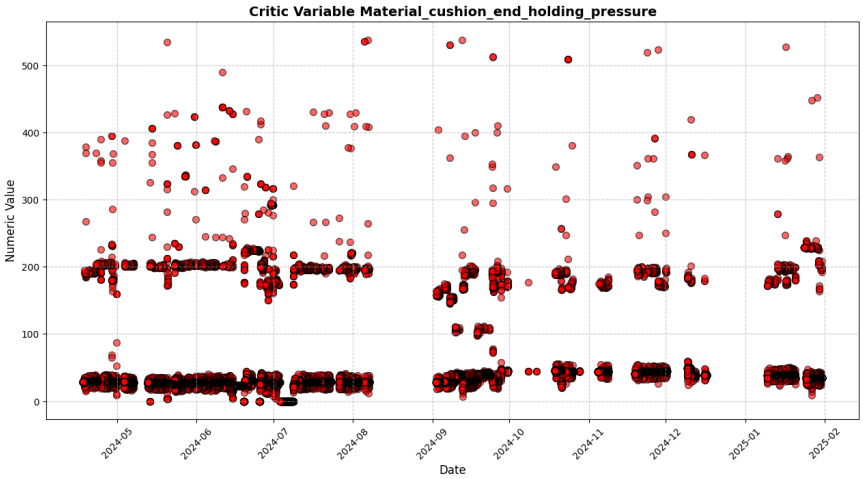


Figure 4. Critical variable analysis of material cushion end holding pressure for Machine 76.

4.2. Dimensionality Reduction Using PCA

Principal Component Analysis (PCA) is a statistical technique that transforms high-dimensional data into a lower-dimensional space while retaining most of its variance. This process is particularly useful in clustering, as it removes noise and redundant information, leading to improved performance. PCA identifies directions, known as principal components, along which data exhibits the most variance, and projects the data onto these components.

Since PCA is sensitive to scale, all numerical features were standardized using z-score normalization, ensuring a mean of zero and a standard deviation of one. This step prevented features with large magnitudes from dominating the analysis.

As the dataset contains a large number of records and the DBSCAN clustering model uses a large amount of memory, it was necessary to reduce the dataset, preserving as much of the variance as possible. For this reason, PCA was tested with *variance threshold* values between 0.8 and 0.995, which are meant to have between two and four components.

4.3. Rationale and Objectives for Clustering

Since the dataset does not contain any labels indicating the machine's operational state, it is necessary to apply an unsupervised learning approach to infer these states from the data. Specifically, we use clustering to group similar observations together based on the values of the available features.

The rationale for employing clustering lies in its ability to uncover natural groupings within the data without prior knowledge of the categories. By doing so, we aim to identify distinct operational states of the machine. Each cluster is expected to correspond to a different state, such as normal operation, early signs of degradation, or imminent failure.

Although this technique is quite interesting, it presents certain challenges. Most notably, the fact that the groups identified in the data may not reflect real-world groupings. To address this, several tests needed to be conducted using different parameters, followed by a meeting with company stakeholders to determine the number of clusters that best fit the context. This gave greater reliability to the study.

4.4. Clustering with DBSCAN

DBSCAN is a density-based clustering algorithm that groups data points based on their density, making it well-suited for datasets with irregular cluster shapes. Unlike K-Means, DBSCAN does not require specifying the number of clusters beforehand. Instead, it identifies clusters as dense regions separated by areas of lower density.

DBSCAN assigns each data point to a cluster or labels it as noise if it does not meet the density criteria, that is, if the number of neighboring points within a given radius is less than a predefined minimum. A core point is one that has at least this minimum number of neighbors within the radius. Points that do not have enough neighbors are classified as noise.

This model contains two parameters that can be changed to obtain different results. These parameters, min_samples and eps, are responsible for defining the minimum number of points within the radius to form a cluster and for defining that radius, respectively. The values used for this study for min_samples are between 50 and 300 and for eps between 0 and 2. These parameter values were chosen based on the dataset values so that no single cluster or several irrelevant clusters are created.

4.5. Classification with PyCaret

PyCaret is an open-source, low-code machine learning library in Python 3.12.9 that automates machine learning workflows. It was designed and implemented to simplify the work of data scientists. With just a few lines of code, it automates the entire machine learning workflow, from data preparation and model comparison to hyperparameter tuning, cross-validation, and final model selection.

In this study, PyCaret was also used for easy and balanced splitting of the dataset across all clusters, ensuring the same proportion of data from each cluster was used for both training and testing. Additionally, PyCaret trains multiple machine learning models and tests various hyperparameter configurations for each one. The best-performing model is then selected, saved, and used in the application. It also offers a variety of visualizations to assess classification performance and detect potential issues such as overfitting or underfitting.

Before carrying out the classification tests to determine the most effective classifier, the data were divided equally, percentage-wise, among all the clusters. This approach aims to ensure that, regardless of the number of samples per cluster (for example, one cluster with 1000 samples and another with only 100), the proportion of data allocated for training and testing remains uniform. The data_split_stratify parameter, provided by PyCaret, ensures that.

Figure 5 illustrates the steps followed during the data analysis process.



Figure 5. Diagram showing the developed work process.

5. Clustering and Classification Results

Clustering and classification results are presented and discussed below.

5.1. DBSCAN Results

After optimization, DBSCAN identified a total of six clusters for the machine, plus noise. Noise in the present case can be discrepant samples caused by electromagnetic noise, spurious phenomena, or a machine malfunction. In the present study, after careful analysis, the noisy samples were considered not relevant for further analysis.

Figure 6 shows a visualization of the results of the clustering algorithm. The figure illustrates how the algorithm identifies different clusters over time, for variable Shot Volume. Figure 7 shows the same for variable Injection Pressure. Those clusters were formed using four PCA components, which explain up to 99.5% of the variance in the dataset. Table 3

shows a description of each cluster. Points assigned label −1 are not included in any cluster, so they are considered noise. The description of the clusters and the respective assignment were discussed together with the provider of the dataset, and validated by the customer.

As the figures and the table show, DBSCAN was able to successfully separate the working states of the machine, with a silhouette score of 0.79. We used the silhouette score as a metric to assess the overall quality of the clustering results, but we focused more on the graphical analysis of the clusters and discussions with the company that provided the data. This collaborative approach helped us to identify the meaning of each cluster and to determine the number of clusters that best represented the machine's operating states. Points assigned labels −1 (noise), or 6, may require urgent attention because they refer to possible anomalous states. Cluster 5 refers to a state which also requires attention, because the machine is not operating in good condition.

Table 3. Description of clusters, their respective colors, and meanings. ID is the label of the cluster. Number −1 refers to noise points, which are isolated and do not form any cluster. Sil is the silhouette score.

ID	Color	Description	Sil.
−1	Light Blue	*Outliers*, possible noise points or very severe faults.	79%
0	Dark Blue	Reduced or anomalous operating condition.	79%
1	Light Green	Normal operation with a 4-cavity mold.	79%
2	Dark Green	Normal operation with a mold different from the light green one.	79%
3	Pink	Machine operating with fewer than 4 cavities.	79%
4	Red	Machine stopped.	79%
5	Yellow	Anomalous operation, capturing anomalies from the pink *cluster* (3 cavities).	79%
6	Orange	Severe anomaly, but with low occurrence, possibly irrelevant.	79%

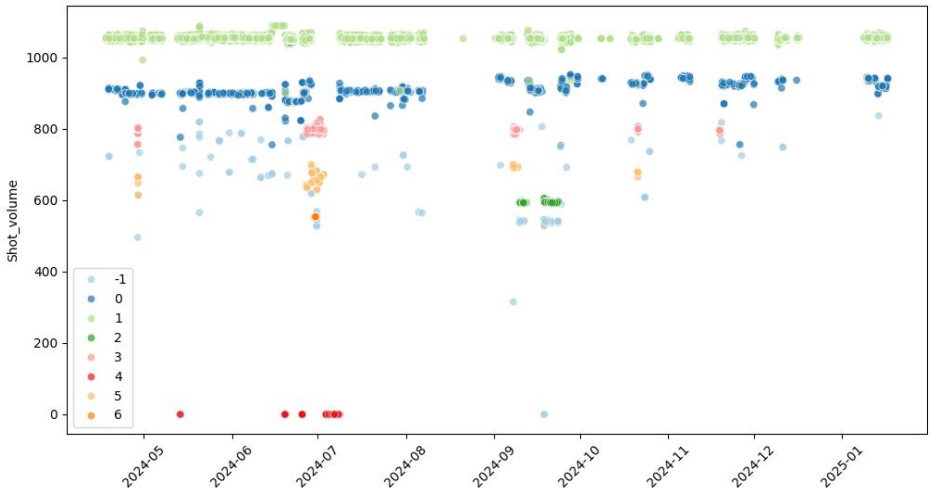


Figure 6. Illustration of DBSCAN clustering results for variable Shot Volume. There are 7 clusters with labels 0–6, plus noise with label −1. Cluster meanings and colors are described in Table 3.

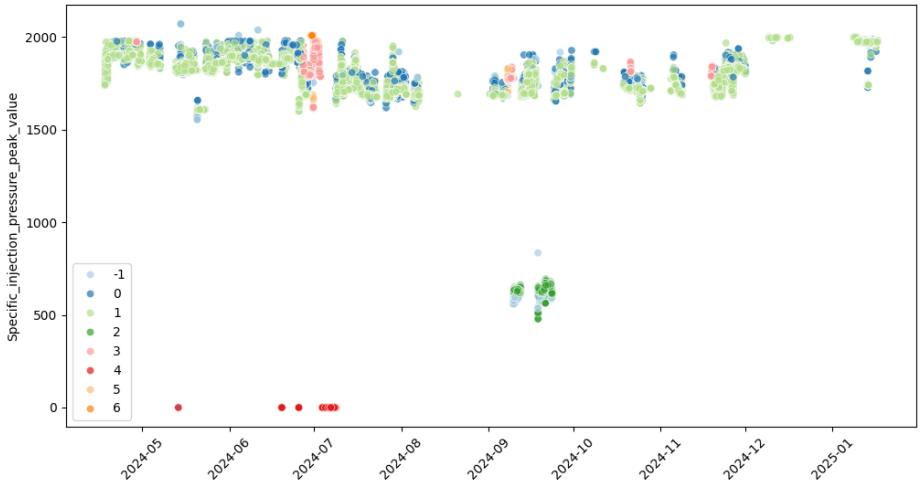


Figure 7. Illustration of DBSCAN clustering results for variable Specific Injection Pressure Peak Value. Cluster meanings and colors are described in Table 3.

5.2. Detailed Cluster Analysis

Since the machine operates with a four-cavity mold, a specific cluster immediately forms, represented in light green in Figure 6. If one of the cavities exhibits a defect or malfunction, the machine operates with three cavities. This operating condition forms a separate cluster. Another example is the replacement of the mold with one that has only a single cavity, represented in the graph by the dark green color.

The clusters with the highest potential to represent an anomalous machine state were identified as dark blue and yellow clusters. As shown in Figure 7, this cluster corresponds to values that fall above and below the ranges considered normal, thus indicating a potential irregularity in the machine's operation. Additionally, the light blue points, considered outliers, may be classified as anomalous points.

Through detailed analysis of the clusters with the dataset provider, it was possible to identify that one of them is directly related to the machine's idle state. This cluster, represented in red, corresponds to the moment when the machine is in standby mode. In this state, the machine is not performing any active operations, awaiting its next task or activation. Distinguishing this cluster is essential for understanding machine inactivity periods and effectively monitoring its operational cycles.

Table 4 shows some statistical properties of each cluster. The statistics are shown for the first two components of the Principal Component Analysis (PCA1 and PCA2): min is the minimum, max is the maximum, mean is the average value, and std is the standard deviation within the cluster. For noisy samples, the values are not shown, since noise does not form a cluster. A more detailed description of the clusters is given below.

Table 4. Statistical data of the principal components of each cluster and noise. ID is point label assigned by DBSCAN.

ID	PCA1				PCA2				Description
	min	max	mean	std	min	max	mean	std	
−1	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	Noise. Possible Severe Faults
0	−105.31	506.25	217.51	135.43	−300.29	238.49	−43.62	77.99	Reduced or Anomalous Operation
1	−37.88	540.91	229.67	137.53	−312.02	173.47	−48.17	89.29	Normal Operation with 4−Cavity Mold
2	−1880.73	−1614.96	−1688.68	24.43	688.59	860.92	735.20	16.29	Normal Operation with 1-Cavity Mold
3	−167.88	370.22	134.89	97.70	−154.71	241.31	−20.66	52.02	Normal operation of a 4-cavity mold with certain cavities blocked
4	−5411.40	−5409.07	−5410.74	0.72	−2640.40	−2632.49	−2635.88	2.67	Machine Stopped
5	−210.95	332.50	107.44	131.40	−149.69	250.29	−20.42	76.00	Anomalous Operation
6	428.60	435.49	429.91	1.03	−218.35	−203.47	−213.67	3.25	Severe Anomaly

- ***Points labeled* −1: *Outliers*/Possible Severe Faults**. Those points are extremely dispersed. Considered *outliers*, they potentially represent noise in the data or severe machine failures.
- ***Cluster* 0: Reduced or Anomalous Operation**. There is lower variability compared with *outliers*, with PCA1 values from −105.30 to 506.25. This indicates reduced or anomalous machine operation, possibly associated with an alert condition or low operational efficiency.
- ***Cluster* 1: Normal Operation with 4-Cavity Mold**. Well-concentrated values, with a PCA1 mean of 229.67 and controlled dispersion. Represents normal machine operation with its most common 4-cavity mold, indicating smooth process running.
- **Cluster 2: Normal Operation with 1-Cavity Mold**, PCA1 values concentrated in a negative region (−1880.73 to −1614.95) with low standard deviation. Indicates normal machine operation with a different mold compared with Cluster 1. Useful for classifying different operational modes of the machine.
- ***Cluster* 3: Operation with less than 4 Cavities**. Moderate variability, PCA1 values from −167.87 to 370.21. Represents an operational state where the machine runs with fewer than four cavities, expected in certain production cycles. This occurs when one of the cavities has a defect and is covered up, keeping the machine working in a normal state, but with only three cavities, reducing the number of products produced.
- ***Cluster* 4: Machine Stopped**. Extremely concentrated values in PCA1 (−5411.39 to −5409.07) and PCA2 (−2640.39 to −2632.49), with negligible standard deviation. Indicates that the machine is stopped, with no significant operational variations.
- ***Cluster* 5: Reduced or Anomalous Operation**. Intermediate values, PCA1 ranging from −210.95 to 332.50, with controlled dispersion. Suggests anomalous machine operation, representing low production or a failure state associated with specific conditions.
- ***Cluster* 6: Possible Anomaly**. Values concentrated in PCA1 (428.60 to 435.49), slightly negative variation in PCA2 (−218.35 to −203.47). Low presence in data, considered irrelevant for overall analysis. May represent a rare anomalous situation.

5.3. Classification Results

Once the clustering process was completed and the clustering-labeled dataset prepared, PyCaret was applied to proceed to automatic classification model selection with the best performance based on predefined metrics. Initially, several models were automatically tested using PyCaret; however, none demonstrated true robustness, with the highest accuracy reaching approximately 69%. Consequently, the XGBoost model was implemented due to its well-known robustness in similar cases. Using tree-learning algorithms, extreme gradient boosting (XGBoost) is one of the most efficient classification models available. However, it is still not available in PyCaret, so it was applied separately.

For XGBoost training and validation, a dataset with about 890,000 records was used, which included only data from seven clusters.

A detailed graphical analysis of the time series shows that there are some instantaneous transitions, which are but noise, for it is not possible for the machine to transition from one state to another and then immediately back to the previous state. Hence, the time series was filtered using a median filter with a window of variable width. This approach replaces each data point with the median value of its neighbors within a window of a given size. A window of size 5, applied specifically to the cluster labels, showed the best performance. The median filter reduces the impact of noise and outliers by smoothing the data, thus enhancing the quality of the training set.

Applying the median technique resulted in an improved accuracy of 83.26%, as can be seen in Table 5, a value considered acceptable for reliable classification. The confusion matrix shown in Figure 8 for the evaluated classification model demonstrates strong accuracy for class 1, with 617,605 correct predictions, suggesting that this class is well defined and easily distinguishable. However, notable classification errors are observed: a significant number of instances of class 2 (108,717) are misclassified as class 1, and similarly, many samples of class 3 (15,991) are also predicted as class 1. This pattern indicates that class 1 shares overlapping characteristics with neighboring classes, especially class 2 and class 3, leading to confusion for the model.

It is worth noting that, although class 2 achieves a high precision of 0.956, its recall is significantly lower at 0.412. This indicates that while the model is highly reliable when it predicts class 2, it fails to detect a substantial portion of its actual instances. This can be explained by the fact that class 2 corresponds to a different mold that was used only during a short period of time and differs from class 1 in the number of cavities. As a result, the model had limited exposure to examples of this class, making it more difficult to generalize and detect all occurrences reliably.

Table 5. Classification metrics by class.

ID	Precision	Recall	F1-score	Support
0	0.599	0.566	0.582	14,394
1	0.831	0.981	0.900	629,724
2	0.956	0.412	0.575	187,125
3	0.695	0.617	0.654	50,564
4	0.865	0.957	0.908	7427
5	0.720	0.512	0.598	2119
6	0.915	0.789	0.847	1217
Accuracy		0.833 (Total support: 892,570)		
Macro avg	0.797	0.690	0.724	892,570
Weighted avg	0.846	0.833	0.812	892,570

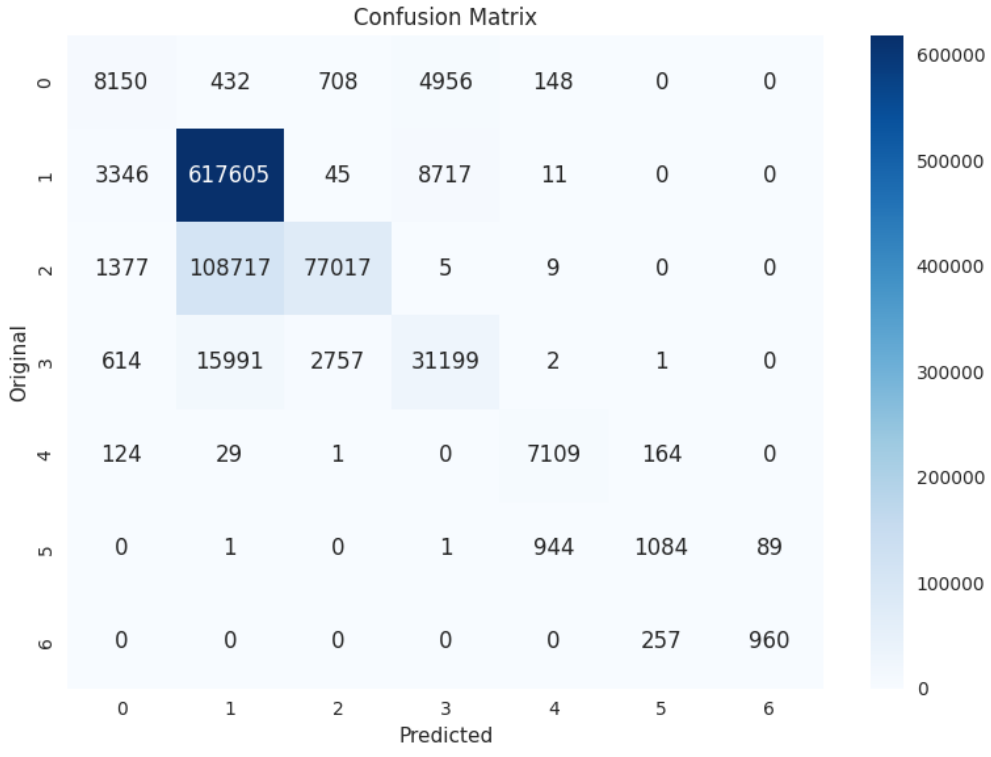
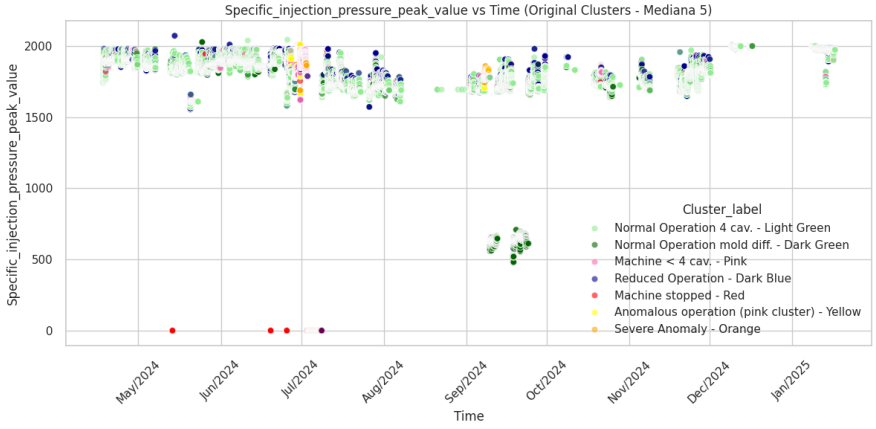
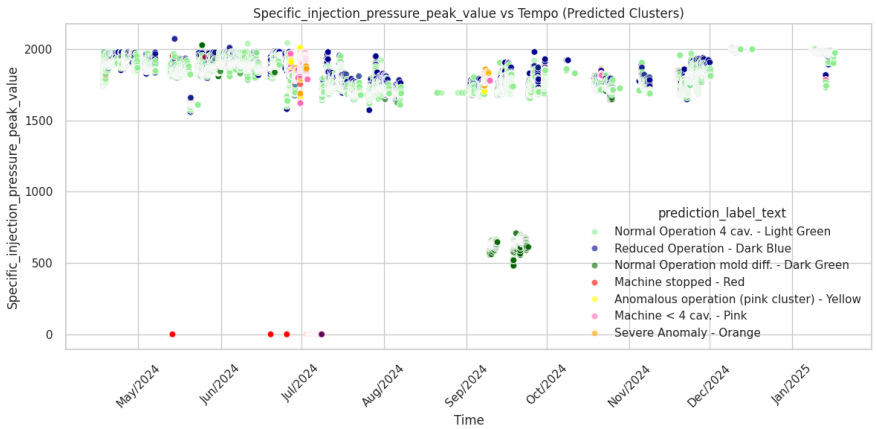


Figure 8. Confusion matrix.

Figure 9a,b presents the time series of the variable Shot_Volume over a given period. Data points are colored according to their respective cluster assignments. Figure 9a displays the ground-truth cluster labels, while Figure 9b shows the cluster predictions generated by the classification model. True Labels–Ground Truth Clustering: This figure represents the original clustering of the dataset, based on the true labels. Each color corresponds to a distinct cluster, providing a clear view of the actual data distribution according to the known classification. Predicted Clusters–Classification Model Output: This figure shows the predicted clusters assigned by the classification model. When compared with the ground truth (Figure 9a), the predicted clusters exhibit a similar overall pattern, suggesting that the model effectively captured the underlying structure of the data; minor discrepancies between the two figures may indicate regions of misclassifications or areas where clusters overlap.



(**a**)



(**b**)

Figure 9. Comparison of true labels, clustering predictions and XGBoost predictions. (**a**) Classification result from the DBSCAN clustering algorithm. (**b**) Classification result from the XGBoost algorithm.

6. Discussion

The results obtained from the DBSCAN clustering model provide valuable insights into the operational states of the plastic injection molding machine. A comparison with state-of-the-art methods, such as Predictive Maintenance by Pierleoni et al. [13] and Aslantaş et al. [12], represents an important first step. This is because, with the classification model, it becomes possible to determine when machines require maintenance, detect anomalies, or identify human errors. For example, if a mold is changed but an operator forgets to adjust the settings, this method can detect the inconsistency, triggering an alarm.

One of the main contributions of this study is the identification and differentiation of machine states, which will assist industrial companies operating plastic injection machines to detect anomalies more easily. Unlike conventional analyses that rely on artificially generated data or data not originating from real contexts, this study successfully identified distinct machine states using a real dataset covering several months of operation. This distinction is crucial, as it increases the likelihood that the clustering step will be more accurate and effective in real-world applications. Utilizing Principal Component Analysis to reduce data dimensionality and integrating the Silhouette Score metric to assess clustering quality, our methodology ensures robust segmentation of machine states, thereby supporting improved monitoring and predictive maintenance strategies.

A key insight from our findings is the effective detection of anomalous states. The identification of outliers (label -1), or anomalous operation clusters (labels 0, 5, and 6), as potential severe malfunctions aligns with industry concerns regarding common failures in plastic injection molding machines [3], including injection device obstruction, mold cooling inefficiencies, and pressure or temperature variations. By clustering machine states based on real operational data, our method offers a data-driven approach to early and straightforward fault detection, complementing conventional preventive maintenance.

Segmenting machine states into seven distinct clusters provides a more granular understanding of operational behaviors. For instance, differentiating between normal operation with a four-cavity mold (light green cluster) and operation with the same mold but fewer active cavities (pink clusters) enables more precise monitoring of mold performance. Similarly, identifying idle states (red cluster) allows better assessment of machine downtime and potential maintenance periods.

In the study, different DBSCAN parameters were tested, and adjustments in PCA variance retention significantly improved clustering runtime and efficiency. Similar to Zhang and Alexander et al. [9], Principal Component Analysis was instrumental in retaining the most relevant features, enhancing clustering performance. The achieved Silhouette Score of approximately 80% indicates high-quality clustering performance, validating our model's effectiveness. Furthermore, the approval by the dataset supplier adds credibility and supports the real-world applicability of the results.

Regarding model generalization, the proposed approach can be extended to other plastic injection molding machines, provided that they operate with the same relevant variables and use identical or similar molds. This ensures that the learned clusters remain meaningful across different machines. In the current project, to promote robustness and reliable evaluation, the dataset was split using PyCaret, which performs stratified and randomized division, ensuring that each class is proportionally represented in both training and testing sets. This stratification functions similarly to cross-validation, mitigating bias and improving the model's generalization capability. While transfer learning or even trained model transfer may be possible, further training for each specific machine will probably be required.

From a deployment perspective, the model is ready for integration within the factory environment. The machines continuously record their process variables in a centralized database, and a Python program has been developed to fetch these data in real-time, applying the trained classifier to instantly categorize machine states. This enables timely alerts and operational decisions, thus supporting predictive maintenance in an automated manner.

However, since new molds or variables with different value ranges may be introduced over time, the model requires periodic retraining and fine-tuning with fresh data to maintain accuracy and adaptability. This continual learning process is essential to preserve the robustness and reliability of the predictive system as operational conditions evolve.

Another challenge is the early-stage nature of the data collection process, where inconsistencies and variable relevance shifts may occur. Therefore, the ongoing evaluation of variable importance and data quality is necessary to ensure the model remains aligned with industrial needs.

In addition to the results discussed, recent developments in machine learning and structural health monitoring (SHM) offer promising directions for enhancing the methodology proposed in this study. Khatir et al. [15] demonstrated that integrating artificial neural networks (ANNs) with analytical models can significantly improve prediction accuracy while reducing computational cost. Although their work focuses on deflection prediction in tapered steel beams, a similar hybrid approach could be adapted to plastic injection molding machines. For example, combining data-driven learning (such as XGBoost or ANN) with domain-specific process models (e.g., injection pressure curves, mold flow dynamics) could improve both the interpretability and generalization of the fault classification system.

Khatir et al. [16] further showed that using frequency response functions (FRFs) and optimizing ANN training with bio-inspired algorithms like the Reptile Search Algorithm (RSA) can enhance damage detection accuracy and convergence speed. While our current study uses only process parameters (e.g., pressure, cycle time, volume), incorporating vibration or acoustic signals into the dataset processed via RSA-optimized neural networks could increase sensitivity to mechanical faults such as wear in moving parts, misalignment, or structural fatigue. This would expand the detection capabilities beyond what is possible with process data alone.

Similarly, the hybrid PSO-YUKI approach proposed by Khatir et al. [17] for detecting double cracks in CFRP beams suggests potential advantages of using metaheuristic optimization to tune machine learning models in complex industrial systems. In our context, such optimization could be applied not only to improve the classification stage, but also to automatically tune DBSCAN parameters (eps and min_samples), a critical step currently conducted empirically. This could lead to more robust and reproducible clustering results across different machines or molds, particularly when deploying the method at scale in varied industrial settings.

These insights support the view that future versions of our predictive maintenance system could evolve toward a hybrid, multimodal approach, incorporating not only sensor data and machine learning but also signal processing, vibration analysis, and automated optimization. Such integration would align our system more closely with the broader goals of Industry 4.0 and intelligent manufacturing.

This research advances understanding of plastic injection molding machine behavior by introducing a clustering-based approach for state identification and classification to facilitate anomaly detection. Future work may involve retraining the classification model with new data and variables, integrating additional sensor inputs, and refining clustering techniques to further enhance predictive capabilities and operational efficiency.

In summary, the main objectives outlined in the introduction were fully achieved. The behavior of the plastic injection molding machine was thoroughly analyzed using both literature review and real-world dataset examination. Through the application of unsupervised clustering techniques, specifically DBSCAN combined with PCA, distinct machine operational states were successfully identified and characterized. Furthermore, these clusters served as a foundation for training a classification model capable of detecting anomalies and potential human errors in an industrial setting. The alignment of the results with real operational scenarios confirms the practical value of the proposed methodology and its potential for implementation in predictive maintenance systems.

7. Conclusions

This study presents a novel data-driven approach for identifying and monitoring operational states and anomalies in plastic injection molding machines using the DBSCAN clustering algorithm coupled with supervised classification. The methodology successfully uncovered seven distinct machine states, enabling a more granular understanding and timely detection of abnormal conditions that could lead to faults or inefficiencies.

The significance of this work lies in its practical applicability within industrial environments. By leveraging real operational data rather than simulated or synthetic datasets, the approach offers a robust foundation for predictive maintenance strategies that can reduce downtime, optimize production quality, and prevent costly failures. The implementation of a real-time data pipeline and classification system further demonstrates the feasibility of deploying such models in live factory settings, marking an important step towards intelligent, Industry 4.0-enabled manufacturing.

Nevertheless, certain limitations warrant consideration. The clustering and classification performance heavily depend on the quality, completeness, and representativeness of the data. Variations in machine configurations, mold types, and external environmental conditions could challenge the model's generalization capabilities. Moreover, the model's reliance on fixed variable sets and periodic retraining highlights the need for adaptive algorithms capable of accommodating evolving operational contexts without significant manual intervention.

Future research directions should focus on addressing these limitations by integrating multimodal sensor data (such as vibration and acoustic signals), developing automated hyperparameter tuning methods for clustering algorithms, and exploring hybrid models that combine data-driven learning with physical process knowledge. Additionally, extending this framework to a broader range of industrial machines and operational scenarios will be critical for validating its scalability and effectiveness in diverse manufacturing ecosystems.

In conclusion, this study contributes a valuable methodology for advancing predictive maintenance in plastic injection molding processes, with promising potential for enhancing manufacturing reliability, reducing operational costs, and supporting the transition toward smarter, data-centered industrial systems.

Author Contributions: Conceptualization: G.M., R.S., and M.M.; Methodology: G.M., R.S., J.B. and M.M.; Software, J.C. and R.S.; Formal analysis: G.M., R.S. and M.M.; Data Curation: J.C., G.M., R.S. and M.M.; Writing—original draft preparation: J.C., J.B. and M.M.; Writing—review and editing: G.M., R.S., J.B. and M.M.; Supervision: G.M., R.S., J.B. and M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work received financial support from the Polytechnic University of Coimbra within the scope of Regulamento de Apoio à Publicação Científica dos Trabalhadores do Instituto Politécnico de Coimbra (Despacho n.º 4654/2024).

Data Availability Statement: The datasets presented in this article are not readily available because they belong to an industry which only makes them available on contract. The authors accessed the data under a maintenance and research contract.

Conflicts of Interest: Authors Rui Silva and Gonçalo Martins were employed by the company Sinmetro LDA. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Daurenbayeva, N.; Nurlanuly, A.; Atymtayeva, L.; Mendes, M. Survey of Applications of Machine Learning for Fault Detection, Diagnosis and Prediction in Microclimate Control Systems. *Energies* **2023**, *16*, 3508. [CrossRef]
2. Costa, J.; Silva, R.; Martins, G.; Barreiros, J.; Mendes, M. Analysis of the state of a plastic injection machine. In Proceedings of the Proceedings of PAMDAS 2025-International Conference on Physical Asset Management and Data Science, Coimbra, Portugal, 17–18 July 2025.
3. TopStar Machine. Common Operating Problems and Solutions for Several Plastic Injection Machine. Plastic Injection Machine Play a Vital Role in Modern Injection Molding Manufacturing Processes. 2025. Available online: https://www.topstarmachine.com/common-operating-problems-and-solutions-for-several-plastic-injection-machine/ (accessed on 13 February 2025).
4. Abid, A.; Khan, M.T.; Iqbal, J. A review on fault detection and diagnosis techniques: Basics and beyond. *Artif. Intell. Rev.* **2021**, *54*, 3639–3664. [CrossRef]
5. Datta, S.; Sarkar, S. A review on different pipeline fault detection methods. *J. Loss Prev. Process Ind.* **2016**, *41*, 97–106. [CrossRef]
6. Md Nor, N.; Che Hassan, C.R.; Hussain, M.A. A review of data-driven fault detection and diagnosis methods: Applications in chemical process systems. *Rev. Chem. Eng.* **2020**, *36*, 513–553. [CrossRef]
7. Park, Y.J.; Fan, S.K.S.; Hsu, C.Y. A review on fault detection and process diagnostics in industrial processes. *Processes* **2020**, *8*, 1123. [CrossRef]
8. Yu, J.; Zhang, Y. Challenges and opportunities of deep learning-based process fault detection and diagnosis: A review. *Neural Comput. Appl.* **2023**, *35*, 211–252. [CrossRef]
9. Zhang, J.; Alexander, S.M. Fault Diagnosis in Injection Molding via Cavity Pressure Signals. In Proceedings of the IISE Annual Conference, Montreal, QC, Canada, 18–21 May 2004; Institute of Industrial Engineers: Peachtree Corners, GA, USA, 2004; pp. 1–6.
10. Kozjek, D.; Vrabič, R.; Kralj, D.; Butala, P.; Lavrač, N. Data mining for fault diagnostics: A case for plastic injection molding. *Procedia CIRP* **2019**, *81*, 809–814. [CrossRef]
11. Ke, K.C.; Huang, M.S. Quality classification of injection-molded components by using quality indices, grading, and machine learning. *Polymers* **2021**, *13*, 353. [CrossRef] [PubMed]
12. Aslantaş, G.; Alaygut, T.; Rumelli, M.; Özsaraç, M.; Bakırlı, G.; Bırant, D. Estimating Types of Faults on Plastic Injection Molding Machines from Sensor Data for Predictive Maintenance. *Artif. Intell. Theory Appl.* **2023**, *3*, 1–11.
13. Pierleoni, P.; Palma, L.; Belli, A.; Sabbatini, L. Using Plastic Injection Moulding Machine Process Parameters for Predictive Maintenance Purposes. In Proceedings of the 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 17–19 June 2020; pp. 115–120. [CrossRef]
14. Hernández-Vega, J.I.; Reynoso-Guajardo, L.A.; Gallardo-Morales, M.C.; Macias-Arias, M.E.; Hernández, A.; de la Cruz, N.; Soto-Soto, J.E.; Hernández-Santos, C. Plastic Injection Molding Process Analysis: Data Integration and Modeling for Improved Production Efficiency. *Appl. Sci.* **2024**, *14*, 10279. [CrossRef]
15. Khatir, A.; Capozucca, R.; Khatir, S.; Magagnini, E.; Le Thanh, C.; Riahi, M.K. Advancements and emerging trends in integrating machine learning and deep learning for SHM in mechanical and civil engineering: A comprehensive review. *J. Braz. Soc. Mech. Sci. Eng.* **2025**, *47*, 419. [CrossRef]
16. Khatir, A.; Capozucca, R.; Khatir, S.; Magagnini, E.; Cuong-Le, T. Enhancing Damage Detection Using Reptile Search Algorithm–Optimized Neural Network and Frequency Response Function. *J. Vib. Eng. Technol.* **2025**, *13*, 88. [CrossRef]
17. Khatir, A.; Capozucca, R.; Khatir, S.; Magagnini, E.; Benaissa, B.; Le Thanh, C.; Wahab, M.A. A new hybrid PSO-YUKI for double cracks identification in CFRP cantilever beam. *Compos. Struct.* **2023**, *311*, 116803. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.