Historical and Real Time Analysis of Endurance Test of Diesel Injectors

Christian Machado do Nascimento
Gustavo Henrique Scherpinski
Fernando de Souza Gonçalves Cassiano
Márcio José Moraes
Robert Bosch Ltda.

ABSTRACT

The durability test for diesel injectors is of major importance to ensure the product's quality throughout its lifespan. The test is conducted on a test bench that replicates the conditions of a diesel engine, simulating the real usage conditions of the injectors. During the testing period, failures may occur in the injector or the test bench. The behavior of these failures is only verified at the conclusion of the test or through analysis conducted by product and bench specialists, demanding unnecessary time and effort.

To make those analyses more frequent and reliable this paper will address the study and implementation of a DAQ system, it will consist in two separate data flows occurring asynchronously: in stream and batch modes. This implementation enables continuous test analysis, for example, through dashboards, which provide a detailed evaluation of the injectors' behavior. The objective is to utilize data, such as the temperature of each cylinder, to verify not only the behavior of the failure but also its development. This analysis facilitates the transmission of signaling alarms to bench specialists, thereby enabling immediate interventions in the event of errors. Additionally, it enables the study of cause-and-effect relationships, ensuring the delivery of more reliable technical reports based on data, rather than solely on post-test product analysis. This approach ensures product quality, facilitates historical analyses and future implementations of machine learning.

INTRODUCTION

Diesel injectors play a pivotal role in the automotive industry, as they are instrumental in ensuring the proper functioning of diesel internal combustion engines. The efficiency, precision, and endurance of the engine are critical factors in the performance of the engine, fuel economy, and emission reduction. The PROCONVE P8 Phase of the Air Pollution Control Program for Motor Vehicles, utilized since January 1, 2022, establishes the requirements for the PROCONVE Control Program based on EURO 6, which is designed to regulate the emissions of polluting gases and noise from new heavy-duty vehicles intended for road use. Furthermore, the program implements a variety of additional provisions [1]. Vehicles are subject to a series of restrictions on NOx, CO, and NMHC emissions from exhaust, as well as homologation requirements, testing procedures, and a multitude of other requirements for internal combustion vehicles. The advent of this new standard underscores the imperative for the enhancement of validation and testing methodologies for diesel injectors. This imperative is predicated on its potential to catalyze innovation and promote continuous development within the sector.

Historically, the endurance tests of diesel injectors have relied on conventional methods for measurement, such as the continuous maximum usage of the injectors' capacities in test benches or the simulation of real use in production engines. The process ensures that the parameters of a new injector meet the requirements established by international and national standards. This allows for the mass production of the injector for use with numerous brands of diesel vehicles.

After the execution of the test with a predetermined time, a thorough evaluation is conducted on both the internal and external components of the injector body to ascertain their conformance to the established criteria. These methods pose significant challenges in the diagnosis of potential injector failures, including the lack of data indicating the failure occurrence time during the test period, the potential failure or substandard performance of the testbench, and any other potential issues that may arise during the test. This results in delays in the analysis process, potentially compromises the quality of the technical report for customers and increases costs due to tests being conducted while failures are not being diagnosed in real time.

In the face of the challenges posed by conventional methods, there is an imperative to establish systems for the acquisition of both real-time and historical data. Due to the big amount of data that will be stored, frameworks for data-driven decision-making (D3M) will assume major importance in the identification of trends, insights, and the visualization of improved opportunities through the meticulous analysis of data [2]. These systems facilitate the continuous monitoring of test performance, enabling expeditious responses when necessary. The historical data can assist in the analysis of the injector's behavior and the creation of a technical report for the client. Furthermore, these systems enable the comparison of numerous tests conducted over extended periods, ranging from months to years, thereby facilitating the assessment of the quality of new products or those in production.

DEVELOPMENT

The solution to these problems was the implementation of a pipeline for acquiring, transforming, and normalizing data, as well as loading it between frameworks. The pipeline under consideration has the capacity to facilitate asynchronous data streaming and storage.

SYSTEM REQUIREMENTS – The requirements were specified based on customer demand and analysis of the difficulties they faced. Based on this information, the following prerequisites for a robust solution were formulated:

1. The process of acquiring data (DAQ) is facilitated by a Programmable Logic Controller (PLC) via input/output (I/O) ports from Thermocouple Sensors embedded on the injectors of each cylinder;

2. Temporary storage of data on a platform that specializes in acquiring real-time data from event sources (e.g. sensors and software applications) for subsequent consumption. This category of platform is designed as an event streaming platform [3];

3. Upon the arrival of new data arrives on the event streaming platform, Python is responsible for consuming, transforming, and injecting the data into two databases asynchronously. One database specializes in time series data, defined as the analysis of one or more variables over time [4]. The second database has the capacity to store large volumes of data in any format (e.g., structured, semi-structured, and unstructured) from any data source. This concept is referred to as a "data lake" [5]. A platform comprising dashboards with real-time refresh capabilities is employed to access the database of time series, thereby facilitating remote monitoring;

4. The dashboard application is designed for the analysis of historical data stored in a data lake.

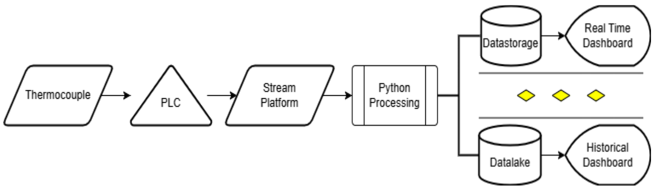

Figure 1. UML Diagram of Pipeline Workflow.
Source: Author

The thermocouples transmit data in millivolts; however, the final value must be expressed in degrees Celsius [6].

COLD-JUNCTION COMPENSATION – Cold-Junction Compensation is a method used to compensate for unwanted parasitic thermocouple effects. The process of creating an electrical junction between the thermocouple and another material, such as copper with thermocouple wire of a different material, is essential for the functioning of the thermocouple. This phenomenon is attributed to the presence of voltage differences, which are a consequence of disparate temperatures along the wire. That is to say, one extremity of the wire is exposed to higher temperatures than the other. The thermoelectric EMF is defined as the electrical potential difference across a thermoelectric device, and its direction is determined by the flow of electrons. When electrons move from a higher temperature (hot) side to a lower temperature (cool) side, the EMF is negative [7]. Conversely, when electrons move from a lower temperature (cool) side to a higher temperature (hot) side, the EMF is positive. This phenomenon, known as the Seebeck effect, was first observed by the renowned scientist Thomas Johann Seebeck (1770–1831) [8].

Thermocouple Law of Intermediate – The law stipulates that the incorporation of a third metal within a thermocouple is permissible, provided that the junctions with the third metal remain isothermal. This principle is elucidated in Figure 2.

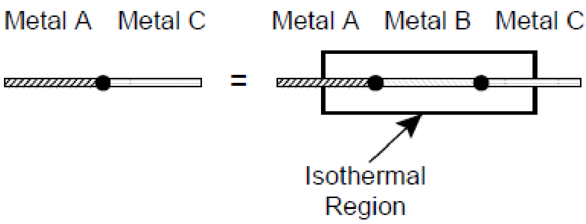

Figure 2. Thermocouple Law of Intermediate Metals [7]

As illustrated in Figure 3, the J2 junction, which corresponds to the T_{ref}, along with the J4 and J3 junctions, is situated within the Isothermal Region. This implies that the three junctions maintain a uniform temperature, thereby precluding any contributions to the ultimate output. Consequently, only the J1 junction remains outside the Isothermal Region, resulting in an imbalance that gives rise to a positive or negative voltage output.

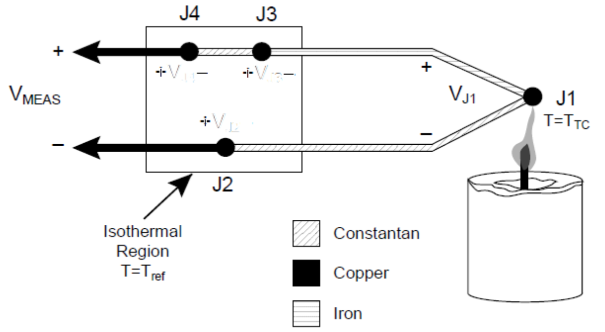

Figure 3. Combination of Thermocouple Law of Intermediate Metals with new Junction [6].

Assuming $V_{jx}(T_y)$ as the voltage, junction as J_x and temperature as T_y, the thermocouple linear formula is as follows:

$$V_{MEAS} = V_{J1}(T_{TC}) + V_{J3}(T_{ref}) \tag{1}$$

- V_{MEAS} – Voltage measured;
- T_{TC} – Temperature of the thermocouple at J1;
- T_{ref} – Temperature of the thermocouple at reference junction.

A thorough examination of the Figure 3 reveals that J3 and J1 are of the same type, as evidenced by the congruence of their values, despite their opposing directions. Given that the junction J1 is currently under test, its voltage will be designate as V_{TC}. The following equation, known as the true open-circuit equation [6], is to be used in the subsequent analyses.

$$V_{MEAS} = V_{TC}(T_{TC}) - V_{TC}(T_{ref}) \tag{2}$$

DAQ WITH PLC – The PLC is equipped with the capacity to retrieve data from six thermocouples and transfer it to the data storage of the event streaming platform. This facilitates the provisional storage of data in real time, thereby enabling subsequent live and remote monitoring. The communication protocol employed is industrial Ethernet, a LAN that utilizes numerous protocols (e.g., TCP/IP, Profibus, Fieldbus, etc.). Perry S. Marshall and John S. Rinaldi asserts that Ethernet has become the global de facto standard for establishing connections between computers. [9]. The protocol under discussion is also widely utilized within industrial contexts. It facilitates communication with an enterprise's internal network, ensuring robust security measures are in place to safeguard data.

PYTHON FOR SOFTWARE COMPENSATION – The utilization of software for cold-junction compensation is employed to eradicate the parasitic thermocouple that replicates the Seebeck Effect. The thermocouple is directly connected to the I/O ports of the PLC, and the ETL process performs the requisite transformations in accordance with the *Direct Voltage Addition Method for Software Cold-Junction Compensation* [6].

<u>Direct Voltage Addition Method for Software Cold-Junction Compensation</u> – Equation 2 demonstrates the true open-circuit voltage with a reference junction at 0°C; however, the necessary information is J1, i.e., V_{TC}. In order to obtain this value, it is necessary to rearrange the quantities in order to obtain the $V_{TC}(T_{TC})$. The following equation is thus derived:

$$V_{TC}(T_{TC}) = V_{MEAS} + V_{TC}(T_{ref}) \tag{3}$$

As previously referenced, the correct implementation of software compensation necessitates the direct measurement of the temperature of T_{ref} with a thermometer. If necessary, the value must be converted from degrees Celsius (°C) to millivolts (mV). Subsequently, the values from the thermocouple must be converted from mV to °C, as the subsequent steps illustrate.

<u>Inverse polynomial function</u> – This equation is responsible for the correct transformation of the thermocouple value from millivolts (mV) to degrees Celsius (°C) [10].

$$t_{90} = \sum_{i=0}^{n} d_i \times E^i \tag{4}$$

- t_{90} is the thermocouple temperature according to ITS90 in Celsius degree (°C);
- n is the order of the polynomial;
- d_i is the i^{th} coefficient of the polynomial;
- E is EMF value in microvolts (μV).

With the assist of Table 1 the equation for quantity transformation can be assured.

<u>Note: In this solution only the range of 0°C to 1300°C is being used.</u>

Table 1. Type K Inverse Function Coefficients [11]

Polynomial Coefficient	Temperature range	
	0 °C to 500 °C (n=9)	500 °C to 1300 °C (n=6)
	0 µV to 20 644 µV	20 644 µV to 52 410 µV
d0	$0{,}000\ 000 \times 10^{0}$	$-1{,}318\ 058 \times 10^{2}$
d1	$2{,}508\ 355 \times 10^{-2}$	$4{,}830\ 222 \times 10^{-2}$
d2	$7{,}860\ 106 \times 10^{-8}$	$-1{,}646\ 031 \times 10^{-6}$
d3	$-2{,}503\ 131 \times 10^{-10}$	$5{,}464\ 731 \times 10^{-11}$
d4	$8{,}315\ 270 \times 10^{-14}$	$-9{,}650\ 715 \times 10^{-16}$
d5	$-1{,}228\ 034 \times 10^{-17}$	$8{,}802\ 193 \times 10^{-21}$
d6	$9{,}804\ 036 \times 10^{-22}$	$-3{,}110\ 810 \times 10^{-26}$
d7	$-4{,}413\ 030 \times 10^{-26}$	-
d8	$1{,}057\ 734 \times 10^{-30}$	-
d9	$-1{,}052\ 755 \times 10^{-35}$	-
Error (°C)	**Max = 0,033 | Min = –0,047**	**Max. = 0,054 | Min. = –0,046**

In the Python algorithm employed for the purpose of compensation, the values from the reference joint and the thermocouple are continually being gathered. The aforementioned information facilitates the configuration of the Python function, which facilitates the integration of the measured value from the thermocouple with the reference joint and the inverse polynomial function to effect its conversion [11]. This is a critical step in the process of structuring the data for subsequent visualization and analysis.

REAL TIME VISUALIZATION – In the realm of modern engineering, the real-time dashboard has emerged as a pivotal instrument for the extraction of insights from metrics, KPIs, and other critical information that necessitates constant monitoring. This instrument facilitates expeditious actions, enabling prompt adjustments within the testbench [12]. This dashboard plays a pivotal role in facilitating uninterrupted monitoring. In the event of system failures, the dashboard is designed to trigger an alert and automatically generate a comprehensive report, which is then disseminated via email to the relevant parties. During the course of the test, there is a possibility that failures may occur in two distinct ways: either in the injector body or in the structure of the testbench. The result of these failures would be temperature deviations in the values measured by the thermocouples.

As illustrated in Figure 4, the dashboard presents a concise yet pivotal array of data. The data presented herein illustrate the values measured by the thermocouples over the course of time. An hour meter is included to ascertain the duration of the test, including any periods of inactivity. In the event of an alarm being triggered, this information is displayed above the time series graph.

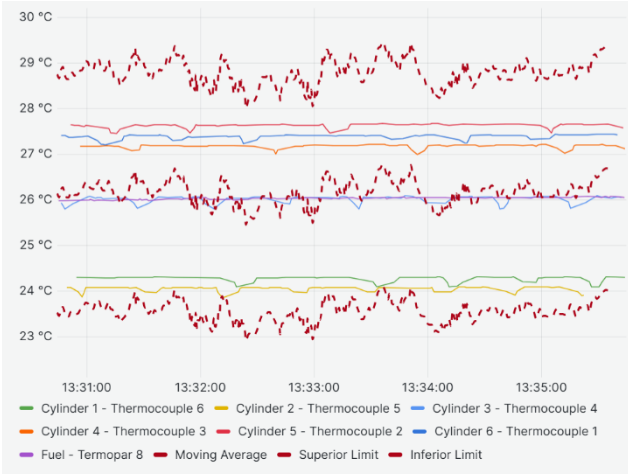

Figure 4. Real Time Dashboard of the Testbench. Source: Author.

The display provides a representation of the temperature of each cylinder and fuel. The temperature of the cylinders is shown to have a moving average according to the equation 5, and an inferior and superior limit that is 10% of the moving average according to equation 6. The dashed lines represent the calculated limits and moving averages.

<u>Note: The equation under consideration incorporates only the cylinder temperature, given the distinct behavior exhibited by the fuel.</u>

$$average = \frac{\sum_{i=6}^{6} Cylinder_i}{6} \quad (5)$$

Equation 5 is merely an application of the normal average formula to the most recent values measured of each cylinder.

$$limits = average \pm (10\% \times average) \quad (6)$$

It has been determined that the behavior of each cylinder differs slightly from that of the others. This discrepancy can be attributed to the positioning of two cylinders along the periphery of the engine, a configuration that has been demonstrated to facilitate enhanced refrigeration. As demonstrated in Equation 6, the application of a 10% range of the average is intended to ensure a safe zone for work.

To facilitate optimal visualization, Figure 5 illustrates the hour meter display, while Figure 6 details the cylinder alarms display. Both are situated above the time series graph to facilitate the visualization of the current state of the testbench.

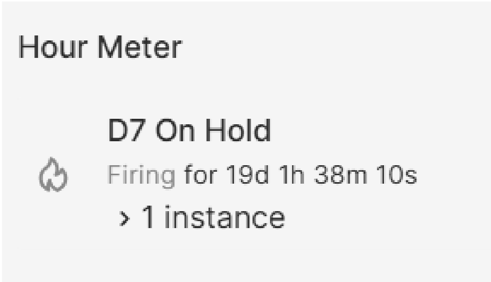

Figure 5. Hour Meter Display. Source: Author.

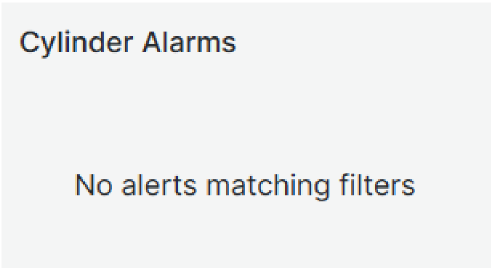

Figure 6. Cylinder Alarm Display. Source: Author.

The cylinders' alarms have been configured to assess the status of the testbench at 30-second intervals and to measure the elapsed time with a precision of one minute, as indicated by the hour meter. This temporal window is characterized by optimal conditions for fast action interference and low processing usage.

HISTORICAL VISUALIZATION – The historical visualization is employed to develop detailed technical reports for clients, ensuring the quality of the analysis and its data-based foundation due to the comprehensive documentation of all test data. As previously mentioned, the duration of each test can range from 8 to 1,500 hours, with the most prevalent being a 500-hour test. The Figure 7 represents an implementation of the D3M enabler showing a week of data from the testbench.

As illustrated in Figure 7, there is a discernible tendency for the temperature to rise and fall in response to variations in ambient temperature. This instrument facilitates exhaustive data analysis by enabling the comparison of data from disparate tests. Through the implementation of data filtration, the range of data and cylinder can be selectively visualized, thereby enhancing the comprehension of injector behavior or failure mechanisms.

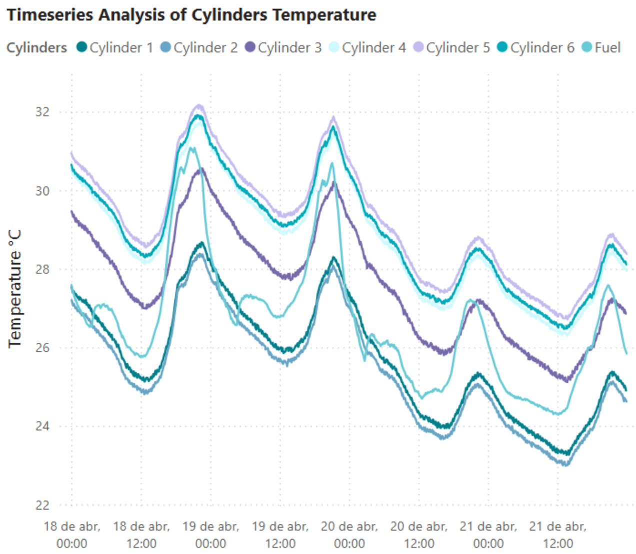

Figure 7. Historical Dashboard of the Testbench. Source: Author

CONCLUSION

The continuous evolution of standards for reducing exhaust emissions, exemplified by the PROCONVE P8 implemented since January 1, 2022, necessitates the validation of diesel injector technology at an appropriate pace. The central objective of this thesis is to provide a comprehensive framework for the validation of novel or existing injectors. The viability of the solution is contingent upon the implementation of a data pipeline that utilizes Type K thermocouples with proper calibration.

Calibration was achieved using software compensation in Python, incorporating the reference junction into the measured value and converting it from mV to °C using the Inverse Polynomial function and Table 1 with Type K Inverse Function Coefficients. This approach effectively eliminated the parasitic thermocouple, ensuring data accuracy and reliability. The data was stored in two distinct ways: asynchronously, in a data lake for historical analysis and in a database specialized in time series data for real-time visualization, each with its own dashboard, allowing remote monitoring during test bench trials, sending alert emails if occur any failure and providing a more comprehensive analysis of the process.

In addition, the integration of historical data stored in a data lake has been demonstrated to enhance the technical reports provided to clients through D3M. This integration facilitates the identification of new opportunities for enhancement and the development of a more efficient and durable injector.

As data accumulates over time in the data lake, it undergoes a transformation into a form of Big Data [13]. In this solution, the data is characterized as homogeneous, originating from a single source. However, this approach presents challenges in terms of updating the dashboard with conventional frameworks due to the substantial volume of data. Moreover, this method can result in increased storage costs, as data lake providers typically charge based on the amount of stored data and transactions. Furthermore, in some years, the data stored today will become obsolete due to innovation and improvement of the injectors, which will increase the cost of the data lake.

The subsequent steps for this solution entail the study and implementation of a machine learning model of a regression type, such as linear regression, random forest, or polynomial regression [14], on the Python program. The objective of this implementation is to predict failures before they occur. This will be achieved by turning off the testbench, thereby decreasing the chances of breaking some internal component, and enabling better efficiency in the technical report. This will result in the dismissal of the need for manual analysis of the failure. In addition, it is imperative to assess the validity of the data. This assessment will inform subsequent data refinement, leading to a reduction in storage costs.

This solution offers an effective approach to accompany ongoing tests in the testbench, facilitating real-time remote monitoring, alarms alerts, and historical analysis. Consequently, it contributes to the generation of superior technical reports for clients and interested parties. This implementation represents the inaugural deployment of the solution, with the potential for subsequent expansion to other test benches and adaptation to accommodate varying quantities and sensors. It enables the execution of new calibrations as required, fostering continuous enhancement. This includes as next steps the incorporation of machine learning models for failure prediction, cleaning algorithms to reduce data lake costs, and other future improvements as deemed necessary.

REFERENCES

[1] - Conselho Nacional do Meio Ambiente (CONAMA), "Resolution No. 490 of November 16, 2018 – Establishes phases L7 and P8 of the Air Pollution Control Program for Motor Vehicles (PROCONVE)," *Diário Oficial da União*, Brasília, Brazil, Nov. 20, 2018, Section 1, p. 155. (in Portuguese)

[2] - J. Lu, Z. Yan, J. Han and G. Zhang, "Data-Driven Decision-Making (D3M): Framework, Methodology, and Directions," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 3, no. 4, pp. 286-296, Aug. 2019, doi: 10.1109/TETCI.2019.2915813.

[3] - Apache Software Foundation, "Apache Kafka: Introduction," [Online]. Available: https://kafka.apache.org/intro. [Accessed: May 5, 2025].

[4] - M. Menezes Reis, "Capítulo 4 –Análise de Séries Temporais," [Online]. Available: https://www.inf.ufsc.br/~marcelo.menezes.reis/Cap4.pdf. [Accessed: May 5, 2025]. (in Portuguese)

[5] - Microsoft, "What is a data lake?" *Microsoft Azure*, [Online]. Available: https://azure-microsoft-com.translate.goog/en-us/resources/cloud-computing-dictionary/what-is-a-data-lake?_x_tr_sl=en&_x_tr_tl=pt&_x_tr_hl=pt&_x_tr_pto=sge. [Accessed: May 5, 2025].

[6] - T. J. Claggett, R. W. Worrall, and B. G. Liptak, "Thermocouples," in *Temperature Measurement*, CRC Press, 2022, pp. 95–118.

[7] - B. Perrin, "The basics of thermocouples," *Circuit Cellar*, [Online]. Available: https://www.mikrocontroller.net/attachment/11655/thermocouplecircuit1.pdf. [Accessed: Apr. 28, 2025].

[8] - MOLKI, Arman. Simple Demonstration of the Seebeck Effect. **Science Education Review**, v. 9, n. 3, p. 103-107, 2010.

[9] - P. S. Marshall and J. S. Rinaldi, *Industrial Ethernet, 2nd ed.: How to Plan, Install, and Maintain TCP/IP Ethernet Networks: The Basic Reference Guide for Automation and Process Control Engineers*, 2nd ed., Burlington, MA: Newnes, 2013.

[10] - National Institute of Standards and Technology (NIST), "ITS-90 Thermocouple Database," [Online]. Available: https://its90.nist.gov/. [Accessed: Apr 28, 2025].

[11] - International Electrotechnical Commission (IEC), *IEC 60584-1: Thermocouples – Part 1: EMF Specifications and Tolerances*, 3rd ed., Geneva, Switzerland: IEC, 2013.

[12] - V. K. Choudhary, G. Kumar, S. Suman and A. Kumar, "A Comprehensive Review on Data Visualization Techniques for Real-Time Information," 2024 International Conference on Inventive Computation Technologies (ICICT), Lalitpur, Nepal, 2024, pp. 866-871, doi: 10.1109/ICICT60155.2024.10544996.

[13] - X. Wu, X. Zhu, G. -Q. Wu and W. Ding, "Data mining with big data," in IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97-107, Jan. 2014, doi: 10.1109/TKDE.2013.109.

[14] - M. R. Putri, I. G. P. S. Wijaya, F. P. A. Praja, A. Hadi and F. Hamami, "The Comparison Study of Regression Models (Multiple Linear Regression, Ridge, Lasso, Random Forest, and Polynomial Regression) for House Price Prediction in West Nusa Tenggara," 2023 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS), Bali, Indonesia, 2023, pp. 1-6, doi: 10.1109/ICADEIS58666.2023.10270916

CONTACTS

Christian Machado do Nascimento

- Email: christian.nascimento@br.bosch.com

Gustavo Henrique Scherpinski

- Email: gustavo.scherpinski@br.bosch.com

Fernando de Souza Gonçalves Cassiano

- Email: fernando.cassiano@br.bosch.com

Márcio José Moraes

- Email: marcio.moraes2@br.bosch.com

DEFINITIONS AND ABREVIATIONS

DAQ Data acquisition

Stream Continuous flow of data

Batch Scheduled portions of data

Dashboard Panel for data visualization

PROCONVE P8 Air Pollution Control Program from Motor Vehicles for heavy-duty vehicles

NOx Nitrogen oxides

CO Carbon monoxide

NHMC Non-Methane hydrocarbons

D3M Data-driven decision-making

PLC Programmable logic controller

I/O Input and Output ports from PLC

ETL Extract, transform and load

Python Program language

EMF Electrical potential difference across a thermoelectric device

LAN Local area network

µV Microvolt

°C Degree Celsius

KPI Key Performance Indicator

mV Millivolt