

zExplAIIn, AI explainability applied to the manufacturing industry

By David Monzo, Tyris.ai

Project Details and Motivation

Artificial Intelligence (AI) provides multiple benefits to the industry. However, usually human operators (also in industry) managing AI systems have scarce knowledge on how they make predictions because of the “black-box effect”, as Machine Learning (ML) and Deep Learning (DL) by nature do not provide hints on how and why a decision has been made. Thus, fears of undetected bias, mistakes, and miscomprehension made by the systems grow among manufacturers, decreasing their trust in them.

Explainable AI (XAI) is a novel field aimed to give rationale on how black box decisions of AI systems are made, inspecting and attempting to understand the steps and variables involved in decision making, in order to increase human trust and provide additional useful information about the predictions. Explainability is particularly important to identify potential bias in the training data and provides information on how the different inputs of the system and the own AI systems influence the decision-making process, on how they are designed, and on what is the rationale for deploying them. This is key to tear down the barriers of a more general adoption of AI.

In this sub-project, Tyris AI will develop zExplAIIn, a new explainable-AI-based zApp within the ZDMP platform. zExplAIIn will provide rationale and insights to the factory operators on the origin and root causes of the anomalies detected within ZDMP, acting as an analytical layer that will run on top of the AI Analytic Runtime module to expand the results already provided by other AI-based zApps, such as zAnomalyDetector or XRay Monitor. zExplAIIn will complement these black-box algorithms currently used within ZDMP with an additional dedicated surrogate model (hybrid model approach) to provide explanations that will help understand the predictions. Regarding its assessment, zExplAIIn will be validated within ZDMP Use Case 1.2 (Defect prediction in machining operations) and the viability to generalise to further scenarios using ZDMP dataset will be studied.

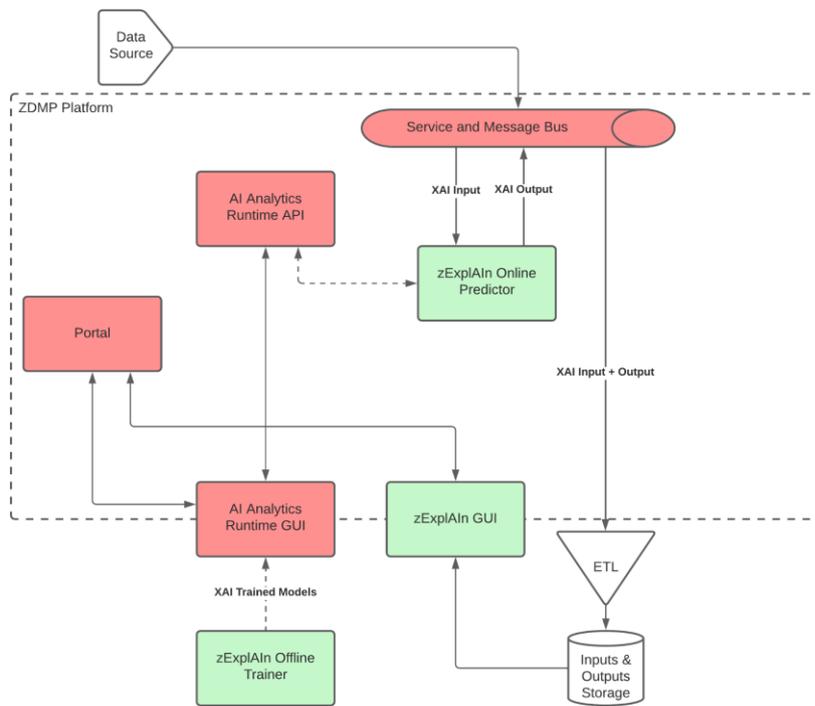
The main novelty of this new zApp is the inclusion of Explainable Artificial Intelligence, bringing a useful feature to the ZDMP platform, so that the predictions made by other zApps (eg detections of anomalies based on process signals), will be explained in different ways:

- Determining the **relevance of each input variable** on the predictions
- Explaining the **root causes determining the decisions** of the system.

ZDMP Fit

zExplAIIn will be implemented as a complementary module to the zAnomalyDetector in the UC1.2, therefore, the architecture and the way this zApp is interacting has been replicated. It is important to mention that this application will work OnPemise at Ford Valencia Engine Plant.

The architecture and the interaction with ZDMP components is shown in figure 1. Two different types of components are shown: (1) the High-Level Architecture of zExplAIIn (green modules) and (2) ZDMP components to which zExplAIIn will have to interact (red modules).



1. High-Level Architecture of zExplAIIn

The High-Level Architecture of zExplAIIn are divided into 3 modules: zExplAIIn Offline Trainer, zExplAIIn Online Predictor and zExplAIIn GUI. The functionalities of each module are detailed below:

- **zExplAIIn Offline Trainer:** This module is responsible for the training of explainability models. Although this module will be part of the final zExplAIIn app, it is expected that the user uses it to train the models using a historical dedicated dataset offline, which means without connection to the ZDMP platform. After the training, the model is ready to be integrated into the platform via AI Analytics Runtime
- **zExplAIIn Online Predictor:** This module carries out the predictions of the models
- **zExplAIIn GUI:** This frontend module will display the outputs made by the zExplAIIn Online Predictor so to allow the end-user to analyse not only the predictions but also the explanations

2. ZDMP components to which zExplAIIn will have to interact

Following the proposed architecture, the components from ZDMP Platform that will be interacting with zExplAIIn are:

- **Portal:** This component is responsible for accessing the ZDMP functionalities. It is in charge to allow the access to other components such as the Marketplace to buy and install different zApps. Concretely, the Portal will allow to connect to the AI Analytics Runtime GUI and the zExplAIIn GUI
- **AI Analytics Runtime:** This component has been conceived with the mission of simplifying the provision of the infrastructure to run AI models automatically. Through a graphical user interface, the user will be able to update a trained model and the needed metadata (e.g. its library requirements) and scripts. After that, the component will create a Docker image of the model and deploy it, integrating and providing a series of functionalities to interact with other components
- **Service and Message Bus:** This component manages the communication between components. First, it allows communication through different component APIs. Another way to manage the communication is via a publish/subscribe messaging in MQTT and AMQP messaging protocols that allows the components to broadcast information (the publication part) and other components to listen for receiving information

(the subscription part). To allow the communication between ZDMP and zExplAIIn, the publish and subscribe messaging option will be employed

The way zExplAIIn provides explanations from the training of the models to the visualisation of the results starts in **zExplAIIn Offline Trainer** module, which is not integrated directly into the system working on-premise. This module will provide zExplAIIn a catalogue of different trained XAI Models to be used by the zExplAIIn Online Predictor. Once a XAI model is trained, it is ready to be integrated into the ZDMP on-premise system at Ford. This will be done through the AI Analytics Runtime GUI (after being logged via the ZDMP Portal component) provided by the ZDMP ecosystem, which will dockerize and deploy this model.

On the analytics side, the deployed XAI model provided by the AI Analytics Runtime is ready to receive the data from the Message Bus. This data will be used internally by the **zExplAIIn Online Predictor** to make explainable inferences. The results of these inferences, that is the XAI outputs will be sent back to the Message Bus.

After the sending of the XAI outputs to the Message Bus, the XAI Inputs and Outputs will be stored in a dedicated database in ElasticSearch, after being manipulated by an ETL process to adapt the message to the proper format of the database.

Finally, the stored inputs and outputs will be displayed through the zExplAIIn GUI (after being logged via the ZDMP Portal component) so as to provide different kinds of explanations, depending on the outputs provided by the various models delivered by the zExplAIIn Models Catalogue.

Results to Date

To date, progress has been made on several fronts of the projects:

- The explainability state of the art has been checked, which allows us to know the different explainability techniques that can be developed in zExplAIIn
- The different components that compose the ZDMP platform have been studied. This allowed choosing which of them were necessary for the development of zExplAIIn

Participant Details

- **Company:** Tyris.ai (www.tyris.ai)
- **Contact Person:** David Monzo (david.monzo@tyris.ai)
- **Profile:** Tyris.ai is a company specialised in Machine Learning and Predictive Analytics, with more than seven years of experience (previously as part of its parent firm, Tyris Software). One of the valuable resources of tyris.ai is its highly specialised team, formed by a solid combination of PhDs. and engineers in Artificial Intelligence, Communications and Software Development. The company stands out through its wide experience in Artificial Intelligence for Industry 4.0 and currently applies its own predictive models to optimise industrial and manufacturing operations.

Environment

The ZDMP – Zero Defects Manufacturing Platform – is a project funded by the H2020 Framework Programme of the European Commission under Grant Agreement 825631 and conducted from January 2019 until December 2022. It engages 31 partners (Users, Technology Providers, Consultants and Research Institutes) with a mission to “Provide the platform, components, services, and marketplace to achieve the right product, at the right time, with the right conditions using the right resources.”. Further information can be found at www.zdmp.eu. ZDMP channels 3.2M€ of SME orientated funding to subprojects, such as this one to both facilitate SMEs with their innovations and increase the value of the ZDMP ecosystem.

Links

| | |
|---------------------------|--|
| • Primary Partner: | david.monzo@tyris.ai |
|---------------------------|--|

| | |
|--|--|
| ● Secondary Partner: | |
| ● Sub project website/blog | www.tyris.ai |
| ● Architecture Component/App(s) | zExplAI |