Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 53-3 (2020) 173–178

End-to-End Industrial IoT Platform for Actionable Predictive Maintenance

Ioannis T. Christou*, Nikos Kefalakis*,
Andreas Zalonis*, John Soldatos*, Raimund Bröchler*

*INTRASOFT Intl. S.A., 2b Rue Nicolas Bové, L-1253
Luxembourg (Tel: +352 44 10 12 2000; e-mail: Ioannis.Christou@intrasoft-intl.com, Nikos.KEFALAKIS@intrasoft-intl.com, Andreas.ZALONIS@intrasoft-intl.com, John.Soldatos@intrasoft-intl.com, Raimund.BROECHLER@intrasoft-intl.com).

Abstract: Predictive maintenance is one of the most prominent use case of smart manufacturing in Industry4.0. Nevertheless, the development of predictive maintenance systems is still challenging as a result of the need to integrate multiple fragmented data sources, to research and apply advanced predictive analytics, and to close the loop to the field in order to provide actionable intelligence. The paper presents the architecture, design and practical implementation of an end-to-end system that addresses these challenges. The system has been successfully deployed in two factories and is positively evaluated in terms of its ability to reduce unscheduled downtimes and to provide increased Overall Equipment Efficiency (OEE).

Keywords: Predictive maintenance, Artificial intelligence, Predictive Control, Big Data, Industrial IoT.

1. INTRODUCTION

The effectiveness of assets' maintenance, service, and repair operations is a core element of an industrial organization's competitiveness. Enterprises are striving to put in place effective maintenance operations that reduce downtimes and help them get the most out of their assets. In principle, there are three ways used by enterprises for maintaining assets: (i) **Reactive maintenance**, where repairs take place when the asset has already broken down, and it is associated with equipment breakdowns; (ii) **Preventive maintenance**, which is performed at regular intervals as a means of preventing equipment failures and unexpected downtimes; it relies on known information about the expected lifetime of the asset and (iii) **Predictive maintenance**, which is based on monitoring of the performance and the condition of the asset, as a means of identifying the best time to maintain it before it breaks down.

Preventive maintenance offers benefits over reactive maintenance, as it provides a good basis for avoiding costly production stops and unscheduled maintenance operations. However, its performance is not optimal, as it takes place much in advance before the asset's breakdown. On the other hand, predictive maintenance alleviates this limitation of preventive maintenance and provides a basis for optimal asset utilization and increased OEE (Overall Equipment Efficiency), when compared to other maintenance models.

The advent of Industry4.0 and digital manufacturing is recently driving manufacturers to lean towards the predictive approach. This is largely due to the advent of cloud computing and Big Data in manufacturing, which facilitates the collection and analytics of large volumes of maintenance data from a variety of different sources. The execution of advanced analytics functions enables the extraction of predictive insights on the assets, which is the foundation for implementing Predictive Health Management (PHM) and Condition-based Maintenance (CBM) techniques [Cachada et al. (2018)].

1.1 Related Work

The benefits of predictive maintenance have been acknowledged for more than a decade as illustrated in [Hashemian et al. (2011)], which provides an overview of predictive maintenance systems implementations. Predictive maintenance systems collect data from a variety of systems and devices including sensors, machine PLCs, communication protocols and business information systems. Accordingly, the datasets are analysed by means of data mining and machine learning techniques, in order to derive predictive insights on the lifetime of assets, based on parameters like RUL (Remaining Useful Life) and EoL (End of Life) [Paolanti et. al (2018)]. A variety of different machine techniques have been used in predictive maintenance systems, such as neural networks [Wu et. al. (2007)], linear regression systems and LSTM (Long Short-Term memory) approaches [Graves (2012)], [Sherstinksi et al. (2018)]. These algorithms vary in terms of their performance, which depends on the systems and datasets at hand.

The practical deployment of predictive maintenance systems is usually based on well-structured architectures, which make use of state of the art digital enablers like cloud computing, Internet of Things (IoT), advanced data analytics and augmented reality [Cachada et al. (2018)]. While several Industrial IoT systems for predictive maintenance have been implemented, there are still challenges including:

-Data fragmentation and lack of interoperability across different datasets: Many plants are already collecting

10.1016/j.ifacol.2020.11.028

various maintenance-related datasets. Nevertheless, these datasets are in most cases fragmented across different systems and form isolated data islands ("silos"), which feature different semantics and formats, are hardly interoperable and cannot be easily combined in the scope of a PdM approach.

-Inability to combine multiple maintenance modalities based on advanced analytics: Most maintenance activities are based on a single modality for detecting maintenance issues, such as vibration analysis, oil analysis, or even thermal imaging. Recent advances in BigData analytics (including machine learning, statistics, and deep learning) provide the means for exploiting large amounts of data from different sources, yet they are still rather poorly exploited in maintenance solutions.

-Lack of flexibility in reconfiguring production processes (i.e. closing the loop to production): Manufacturers are concerned not only about deriving PdM insights, but also about exploiting them as actionable intelligence. State-of-the-art platforms for predictive maintenance are usually focused in the monitoring part of PdM, while also exploiting the data in order to change the state of business information systems (e.g., production schedules in ERP (Enterprise Resource Planning) Systems). Hence, they do not fully exploit the potential of predictive insights for automated optimization and control of production processes.

1.3 PROPHESY-PDM: Value Propositions

Driven by these challenges in this paper we introduce a digital platform for predictive maintenance (PdM), which provides end-to-end support for the implementation of PdM applications. The platform reduces the effort required to implement predictive maintenance applications, based on middleware functions that facilitate data collection and consolidation, data analytics, as well as closing the loop to the field in order to configure other systems based on predictive insights about the assets. Hence, the term end-to-end refers to the provision of support for both data processing and industrial automation functionalities.

To support these functionalities, the platform comprises:

- A data collection sub-system, which enables the collection of maintenance-related data from a variety of different sources, including sources with high-ingestion rates (i.e. streaming sources).
- A data interoperability sub-system, which makes sure that the data sources adhere to a common data model. The latter boost the interoperability across diverse data sources.
- Configurable data analytics functionalities, which enable the implementation of various techniques for deriving predictive insights, such as techniques for RUL calculation.
- A set of (customizable) dashboards that can visualize predictive insights.
- Interfaces to automation, as a means of closing the loop to the field.

The introduced platform has been deployed and validated in different industrial environments (factories), over different datasets and using various data mining techniques for extracting maintenance-related knowledge. The latter techniques include a novel family of data mining algorithms for detecting rate events with adequate accuracy, which are called QARMA (Quantitative Association Rule Mining Algorithms). The QARMA algorithms are briefly outlined in this paper, but their detailed operation can be found at earlier publications (e.g., [Christou et al. (2019)]). Note that the presented platform has been partially developed and validated in the scope of the H2020 PROPHESY platform, which is the reason why it is conveniently called PROPHESY-PDM.

The paper is structured as follows: Section 2 following this introduction outlines the architecture of the platform. Section 3 discusses the digital modelling approach of the project, while Section 4 is devoted to the presentation of the algorithms used. Some validation results from the actual deployment of the platform are discussed in Section 5, while Section 6 concludes the paper.

2. PREDICTIVE MAINTENANCE PLATFORM ARCHITECTURE

The architecture of our predictive maintenance platform is depicted in Fig. 1.

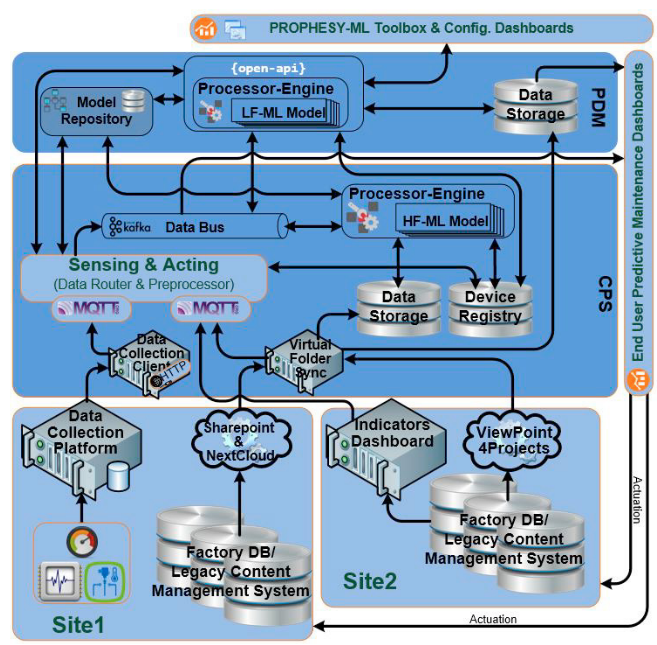

Fig. 1. Anatomy of the Predictive Maintenance Platform.

It comprises the following components:

-Virtual folder synchronization component is responsible to collect measurements and system data from the cloud providers. For example, as part of the platform's deployment in the scope of the PROPHESY project, interfaces to two cloud platforms (i.e. Microsoft SharePoint Platform and the ViewPoint 4Projects) which are used to share data. The measurements and system data are provided in different proprietary formats.

-Sensing and Acting component where the initialization of the CPS (Cyber Physical System) instance is taking place. The configuration of this component is mandatory for the pre-processing of the different data formats and file types to a common repository, transforming them into a common data model. The component provides the means for interfacing to the field i.e. performing actuation functions using the low-level capabilities of CPS systems and IoT devices.

-A data bus infrastructure (developed based on Apache Kafka) is responsible to circulate the information to the different components. For example, the observations retrieval (pre-processed data), used by the ML-Engine (Data Bus → Processor-Engine), as well as the output (predictions) of the ML-Engine (Processor-Engine → Data Bus) are published in different Kafka topics and is consumed by the relevant subscribing components.

-A Machine Learning (ML) Toolkit is used to support the integration and execution of different ML algorithms over the industrial data of the platform and in module fashion. The toolkit consists of two core components which are the PDM Processor-Engine (PDM-PE) and the CPS Processor-Engines (CPS-PE). The CPS Processor-Engines resides within a CPS system and is responsible for processing low-level data streams with the use of HF-ML (High-Frequency Machine Learning) Models collected from the CPS. The PDM Processor-Engine resides within the CPS (Cloud) tier and is responsible for processing data coming from the different CPSs. It, therefore, provides more complex and consolidated analytics from the whole infrastructure with the use of LF-ML (Low-Frequency Machine Learning) Models collected from multiple CPSs. They both follow the same principles and provide common functionality but have different scopes. These Processor-Engines act as “wrappers” of the ML Models described above hiding the complexity of the Management and configuration from the algorithm designer.
The “Wrappers” concept is inspired by earlier work of the authors on distributed data analytics for the Industrial Internet of Things [Kefalakis et al. (2019)]. The Processor Engines are facilitated by a centralized interface where the different components can be controlled which is called Processor Engine Open API. This Open API is used from the PROPHESY-ML Toolbox & Configuration Dashboards. Hence, it acts as the configuration User Interface of the ML Toolbox. One equally important asset of the ML Toolkit is the ML Models. ML Models are the heart of the Analytics system where ML Models instances can be combined from the PDM-PE and CPS-PE in order to provide a complex Analytics solution. This facilitates modularity in testing and deploying different ML algorithms.

-End user Predictive Maintenance Dashboards are responsible for the visualization of analytics results such as Remaining Useful Life (RUL).

-Field Actuation provides proprietary interfaces from the Predictive Maintenance Dashboards to the field. This enables the user to interact with the field to control actuators that would perform urgent maintenance procedures like stopping a production. This feature is built upon availability of such interfaces from the field and the security access level of the Predictive Maintenance Dashboards and its Users. The direct connection is allowing the near real time response of the actuation in critical cases.

3. DIGITAL MODELS FOR INTEROPERABLE DATA COLLECTION AND CONFIGURABLE ML ANALYTICS

To enable different ML algorithms to leverage data from diverse sources (e.g., CPS systems, IoT devices), our platform leverages on a number of digital models that support the representation of measurements from the monitored machines and enables their configurable data routing and pre-processing (e.g., from one or more data sources to one or more algorithms that consume them). The models of the platform have been derived from the data models specified in [Kefalakis et al. (2019)], based on their proper enhancement and customization for predictive maintenance applications. The main entities of these data models include:

-Data Source Definition (DSD): Defines the properties of a data source on the shop floor, such as a data stream from a sensor or an automation device.

-Data Interface Specification (DI): The DI is associated with a data source and provides the information need to connect to it and access its data, including details like network protocol, port, the network address and more.

-Data Kind (DK): Specifies the semantics of the data source. The DK can be used to define virtually any type of data in an open and extensible way.

-Data Source Manifest (DSM): Specifies a specific instance of a data source in-line with its DSD, DI and DK specifications. Multiple manifests (i.e. DSMs) are therefore used to represent the data sources that are available in the factory in the scope of the predictive maintenance platform.

-Observation: Models and represents the actual dataset that stems from an instance of a data source that is represented through a DSM, as can be seen in Fig. 2. Hence, it references a DSM, which drives the specification of the types of the attributes of the Observation in-line with the DK that facilitates the discoverability of the data. An Observation is associated with a timestamp and keeps track of the location of the data source in case it is associated with mobile (rather than a stationary) edge node. An Observation has a location attribute (virtual or physical), which identifies the placement of the data source. The value type of observation is a complex object which is described with the DK entity that an Observation references. Hence, an observation can depict multiple raw measurements coming from a machine or a single value (i.e. the number of cycles/m of a rotor) or even an Analytics Processor result (i.e. the calculated RUL of a machine).

-Edge Gateway: Models an edge gateway of an edge computing deployment of the predictive maintenance platform. In the scope of deployment of the platform, data sources are associated with an edge gateway. This usually implies not only a logical association but a physical association as well, i.e. an edge gateway is deployed at a station and manages data sources in close physical proximity to the station.

Based on the above entities it is possible to represent the different data sources of a digital shop floor in a modular, dynamic and extensible way. This is based on a repository (i.e. registry) of data sources and their manifests, which keeps track of the various data sources that register to it (see the Device Registry and Model Repository components in Fig.1). The predictive maintenance platform includes such a registry, which provides dynamicity in creating, registering and using data sources in the industrial plant.

In order to facilitate the management and configuration of analytics functions and workflows over the various data sources, the PROPHESY digital models specify various analytics-related entities, including:

-Analytics Processor Definition (APD): Specifies a processing function to be applied on one or more data sources. Three processing functions are defined, including functions that pre-process that data of a data source (i.e. Pre-Processors), functions that store the outcomes of the processing (i.e. Store Processors) and functions that analyse the data from the data sources (i.e. Analytics Processors). These three types of processors can be combined in various configurations over the data sources in order to define different analytics workflows.

-Analytics Processor Manifest (APM): Represents an instance of a processor that is defined through the APD. The instance specifies the type of processor and its actual logic through linking to a programming function. In the case of PROPHESY, the latter is a class/programme implemented in the Java language.

-Analytics orchestrator Manifest (AM): Represents an entire analytics workflow. It defines a combination of analytics processor instances (i.e. of APMs) that implements a distributed data analytics task. The latter is likely to span multiple edge gateways and to operate over their data sources.

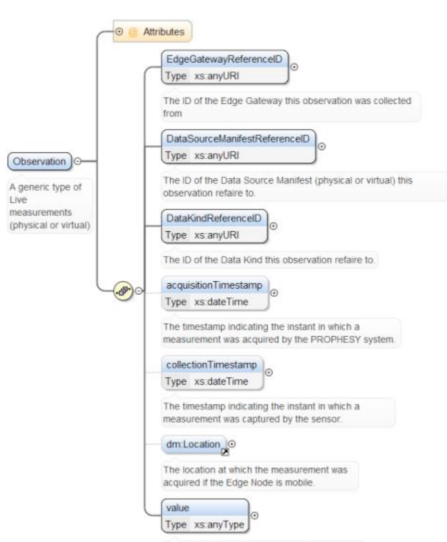

Fig. 2. Snapshot of the Digital Models Schemas Used in the Predictive Maintenance Platform.

The platform provides also files/schemes that specify the configuration of predictive maintenance analytics jobs. These include:

-The Processor Definition (PD) schema, which specifies a processing function to be appliedto one or more data sources. It can be used to set up a data routing flow and to utilize analytics algorithms.

-The Processor Manifest (PM) schema, which represents an instance of a processor that is defined through the PD. The instance specifies the type of processors and its actual logic through linking to a programming function.

-The Processor Orchestrator (PO) schema, which represents an entire data routing workflow. It defines a combination of data processor instances (i.e. of PMs) that implements a distributed data routing task. The latter is likely to span multiple edge gateways (i.e., CPSs) and to operate over their data sources.

The Processor Orchestrator provides the configuration of a runtime environment (Processor-Engine) that controls and executes analytics instances, which are specified in the APM (manifest) format. The Processor-Engine can parse and forward the specified configurations identified in a Processor Manifest. Processor Manifests are stored in a database, a processor manifest has always the same structure and the same semantics. The latter is ensured by the use of the above-listed digital models.

As part of its operation, the processor Engine is capable of instantiating multiple ML Models for the purposes of executing an analytics instance that is described through a Processor manifest. Each analytics instance holds the attributes and sequences to set up the required processors in order to serve one analytics instance.

4. ANALYTICS FOR RUL PREDICTION AND RCA

The platform enables the modular integration of different analytics algorithms and techniques. In practice, our modular approach has been validated with different algorithms, notably predictive analytics algorithms that are suitable for forecasting predicate maintenance related parameters such as RUL. The first predictive algorithms that were tested, were logistics regression techniques, which are proven to yield decent results in quite simple cases when not large datasets are available.

The first set of algorithms that have been implemented are Recurrent Neural Networks (RNNs) [Graves (2012), Sherstinsky (2018)]. RNNs were introduced in the late '90s to solve exactly prediction problems, as they typically use their internal state (memory) to process sequences of inputs. Presently, there are two well-established variants that have proved very robust and efficient in a wide variety of tasks:

Long Short-Term Memory Networks (LSTM), which are a variant of classical RNN with so-called forget gates that allow for some "forgetting" of old or non-important data during training.

Attention-based Networks, which is a more recent variant of RNN that avoid to a greater degree the so-called "curse of dimensionality" of long sequences. They have been particularly successful in sentiment analysis in social networks but also in protein structure analysis tasks.

We have also integrated the QARMA family of algorithms and its variants e.g. the R4RE system recently published for detecting rare events in Predictive Maintenance settings [Christou et al. (2018), Christou (2019)]. The QARMA models represent a drastically different approach to identifying and predicting when the next defect will happen in the scope of sequential datasets (i.e. parts made in a time-sequence). Instead of training a (possibly deep) network of hidden layers of neurons that activate via some kind of step-wise activation function so that eventually the network learns to predict the right target values, QARMA implements a Data Mining inspired approach in which sets of features that appear frequently in the dataset together are collected together, and then each feature is quantified (its value restricted in a numerical interval) with the goal of deriving conditions that imply that a target variable among the features takes on a desired value. A relevant Graphical User Interface (GUI) is illustrated in Fig. 3.

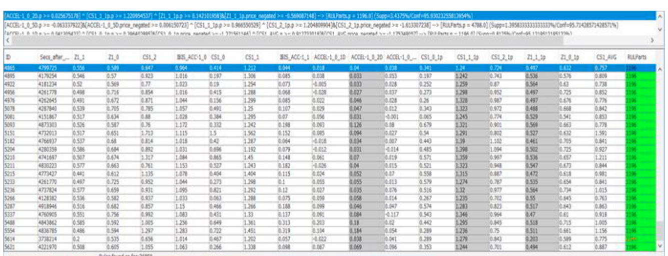

Fig. 3. QARMA GUI resembles a spreadsheet with appropriate filters applied, so that one column (the green one) satisfies a certain condition.

5. VALIDATION RESULTS

The predictive maintenance platform has already been deployed and validated in two different factories, as part of the H2020 PROPHESY project. In one of the demonstrators a dataset of sensor readings was collected (based on CPS deployed in the factory), together with information about the stroke-counter of the sensor equipment collecting the data. This resulted in a dataset containing approximately 6000 attributes and a total of 212 different rows, corresponding to different points in time. Given the stroke-counter available, we assumed that the largest value of this counter corresponds to the end-of-life of the component under scrutiny. Given this assumption, it was obvious how to estimate for each different row, the remaining number of strokes to end-of-life. Since each of the 6000 different attributes is a numeric value, a simple Linear Regression model (LR) was developed to fit the data in a linear equation as best as possible, for the sole purpose of demonstrating how a trained model can be integrated into the PROPHESY ecosystem. The LR model resulted in an absolute MSE error of less than 10% and used all of the independent variables except a single one! The model was implemented as a simple Java program that reads data (sensor readings) from the standard input (stdin) and writes its prediction (as a decimal number) onto standard output (stdout). This Unix-like interface (and Unix-like philosophy of interfacing) sufficed to provide a component that was directly pluggable in the demonstrator of the platform.

The QARMA algorithms (R4RE) were applied on a second dataset originating from the second factory. This second dataset has a much smaller number of sensors than the first one, in particular, 8 independent sensor values were recorded at each reading. An additional 9 derived attributes (the difference between current sensor value and the previously recorded sensor value) were also maintained for a total of 17 independent attributes plus the target attribute measured as parts remaining to be made before the next failure. Despite the much smaller number of features, the second dataset has a much larger number of instances available; more than a year's worth of data made up a total of more than 12,000 rows, making it much more amenable to training without the need for previous feature selection processes that often require much heavier human intervention. The application of QARMA on the reduced feature space yielded much better performance on many different metrics including Mean Absolute Percentage Error (MAPE) than simple regression, and in fact, it performed better than every other ML/DM method we tested on this second dataset by significant margins (see Fig. 4.)

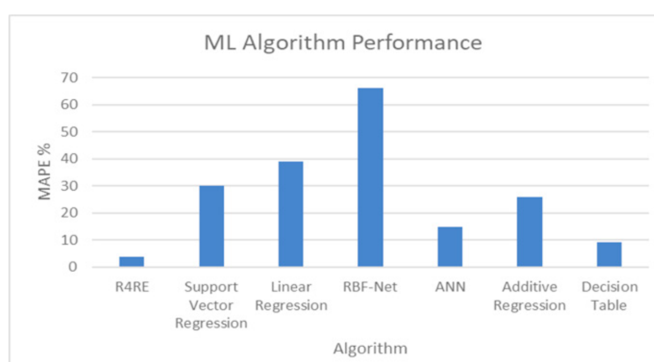

Fig. 4. Produced MAPE% on test data of various ML algorithms implemented in WEKA, compared to the R4RE algorithm (QARMA variant); smaller is better.

6. CONCLUSIONS

Despite the growing interest in Industry4.0 and Predictive Maintenance, there is still no easy way to implement modular, flexible reconfigurable, end-to-end solutions. State of the art approaches are usually tailored to the requirements of specific product lines and datasets; these are hardly generalizable to the needs of other plants. In this paper, we have introduced a digital platform for predictive maintenance applications, which eases the tasks of data collection, data interoperability, and data analytics while closing the loop to the field based on interfaces to automation platforms and business information systems. We have also presented the implementation of different data analytics algorithms, including the QARMA framework for the detection of rare events. The platform has been validated in several industrial deployments, which is an initial proof of its generality and modularity. Our development roadmap for the PROPHESY-PDM platform includes the implementation of a visual tool for configuring elements of the infrastructure, as well as the integration of some of the used algorithms with the OpenML infrastructure.

ACKNOWLEDGEMENTS

Part of this work has been carried out in the scope of the H2020 PROPHESY project (contract number 766994), which is co-funded by the European Commission in the scope of its H2020 programme. The authors acknowledge valuable help and contributions from all partners of the projects, including the plants where the PROPHESY-PDM infrastructure has been validated.

REFERENCES

Christou I.T, et al. “A Parallel/Distributed Algorithmic Framework for Mining All Quantitative Association Rules”, arXiv preprint arXiv:1804.06764, 2018.

Christou I.T, “Avoiding the Hay for the Needle in the Stack: Online Rule Pruning in Rare Events Detection” (2019), *Proc. IEEE Intl. Symp. On Wireless Comunication Systems, Sp. Ses. on IoT in Energy Systems and Industrial Environments*, Oulu, Finland, Aug. 27-30, 2019.

Cachada A., et al., "Maintenance 4.0: Intelligent and Predictive Maintenance System Architecture," 2018 *IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, Turin, 2018, pp. 139-146. doi: 10.1109/ETFA.2018.8502489.

Graves A. "Supervised Sequence Labelling with Recurrent Neural Networks”, Springer-Verlag, Berlin, 2012.

Hashemian H. M., Bean W. C., "State-of-the-art predictive maintenance techniques" IEEE Transactions on Instrumentation and measurement vol. 60 no. 10 pp. 3480-3492 2011.

Kefalakis, N., Roukounaki, A., Soldatos, J., Configurable Distributed Data Management for the Internet of the Things. Information 2019, 10, 360.

Paolanti M., Romeo L., A. Felicetti, Mancini A., et al., "Machine Learning approach for Predictive Maintenance in Industry 4.0", *Proc. 14th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, Oulu, Finland, 2018, pp. 1-6. doi: 10.1109/MESA.2018.8449150

Sherstinksi A., “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network”, arXiv:1808.03314 [cs.LG], 2018.

Wu S.-J., Gebraeel N., Lawley M. A., et al., "A neural network integrated decision support system for condition-based optimal predictive maintenance policy" IEEE Transactions on Systems Man and Cybernetics-Part A: Systems and Humans vol. 37 no. 2 pp. 226-236 2007.