Manufacturing Rev. **12**, 9 (2025)
© H.O. Velesaca and J.A. Holgado-Terriza, Published by EDP Sciences 2025
https://doi.org/10.1051/mfreview/2025003

Manufacturing Review
Available online at:
https://mfr.edp-open.org

MINI-REVIEW OPEN ACCESS

OPC-UA in artificial intelligence: a systematic review of the integration of data mining and NLP in industrial processes

Henry O. Velesaca[1,2,*] and Juan A. Holgado-Terriza[1]

[1] Software Engineering Department, Research Center for Information and Communication Technologies (CITIC-UGR), University of Granada, 18071, Granada, Spain
[2] FIEC, CIDIS, ESPOL Polytechnic University, Campus Gustavo Galindo, Guayaquil, 09-01-5863, Ecuador

Received: 27 January 2025 / Accepted: 24 February 2025

Abstract. This systematic literature review explores the integration of OPC-UA with Data Mining and Natural Language Processing (NLP) techniques within industrial environments. As industrial automation evolves, this integration faces challenges related to intelligence, autonomy, security, privacy, and interoperability—similar. The review evaluates current methodologies and applications aimed at addressing these challenges, particularly in areas like predictive maintenance, anomaly detection, process optimization, and others. Reviewing several primary studies, selected from high-impact scientific databases this paper identifies key strengths, weaknesses, opportunities, and threats in leveraging OPC-UA protocols for AI-based automation. Moreover, it highlights trends and future directions for improving decision-making processes and enhancing machine interoperability in data-driven industry.

Keywords: OPC-UA / industry 4.0 / control systems / data mining / natural language processing / NLP / artificial intelligence

1 Introduction

The arrival of Industry 4.0 has driven the adoption of advanced technologies to optimize industrial processes. Among these, OPC-UA (Open Platform Communications Unified Architecture) has become one of the main communication architectures used in the automation and control of industrial plants [1,2]. As industries generate massive amounts of data, the integration of Data Mining and Natural Language Processing (NLP) techniques with OPC-UA[1] has opened up new possibilities for improved decision making, anomaly detection, and intelligent automation, among other applications.

This literature review seeks to explore how these technologies converge to enable more intelligent and adaptive industrial systems. Data mining has become a key tool for extracting valuable information from large industrial data sets. Its ability to identify patterns, trends and correlations allows companies to optimize their operations and improve the quality of their products. In combination with OPC-UA, Data Mining can process and analyze data in real time, improving the responsiveness of automated systems and providing critical feedback for operational decision making [3,4].

On the other hand, natural language processing (NLP) has demonstrated its potential to interpret, understand and generate human language. In industrial environments, the use of NLP can facilitate interaction with control and monitoring systems, whether through voice commands, analysis of technical reports or the interpretation of textual descriptions of faults and problems in machinery. Integrated with OPC-UA, NLP can contribute to more intuitive and efficient control systems, reducing the need for direct human intervention and increasing efficiency in information management [5].

The integration of these technologies is not without challenges. The main difficulty lies in the heterogeneous and unstructured nature of natural language data, which requires sophisticated analysis and processing approaches. Furthermore, implementing Data Mining and NLP techniques in an industrial environment may face limitations related to cybersecurity, handling large volumes of data, and interoperability between systems. This review examines previous research addressing these issues and offers a critical view of proposed solutions.

* e-mail: hvelesac@espol.edu.ec

[1] https://opcconnect.opcfoundation.org/2024/06/opc-ua-for-ai-enhancing-automation- with-artificial-intelligence/

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. OPC-UA features and their alignment with NLP and data mining capabilities.

OPC-UA feature	Description	NLP applications	Data mining applications
Standardized Information Modeling	Object-oriented model representing complex data relationships and hierarchies	Text structure analysis, Document classification	Feature engineering, Relationship learning, Graph-based ML
Real-time Data Communication	Supports streaming and historical data access	Real-time text analytics, Stream processing	Online learning, Real-time predictions, Time series analysis
Semantic Data Description	Includes metadata and contextual information about data points	Natural language understanding, Semantic analysis, Information extraction	Semantic feature extraction, Context-aware modeling
Security and Authentication	Built-in encryption, authentication, and access control	Secure text processing, Privacy-preserving NLP	Secure ML model deployment, Protected data training
Information Modeling	Hierarchical data representation	Document structure analysis, Topic modeling	Hierarchical clustering, Deep learning architectures

In summary, this article presents a literature review that analyzes how integrating Data Mining and NLP with OPC-UA can improve intelligent industrial systems. The benefits, technical challenges and potential applications of these technologies in industrial settings are explored, with the aim of providing a solid foundation for future research and development in the field.

To address this work in detail, the manuscript is organized as follows. Section 2 details the process applied and the results obtained. Section 3 analyzes the models proposed to create intelligent industrial systems. Then, Section 4 describes the main strengths, weaknesses, opportunities and associated threats, along with the challenges and future lines of research. Also, Section 5 shows the most common state-of-the-art applications. Finally, conclusions are presented in Section 6.

2 Systematic review

A systematic literature review entails the identification, evaluation, and interpretation of the most relevant studies related to a specific topic or research question [6]. Its objective is to provide evidence that addresses questions emerging from prior research. To conduct this review, the methodological guidelines outlined by [6], originally applied in the context of Software Engineering, are followed and contrasted with an alternative methodological approach tailored for Computer Science studies proposed by [7]. The implementation of this process is detailed in the following sections.

2.1 OPC-UA

The first thing is to define what OPC-UA is and how it integrates with techniques such as NLP and Data Mining. OPC-UA is a platform-independent, service-oriented architecture for industrial automation and data communication. Table 1 show its core features and how they align with NLP and ML capabilities.

2.2 Definition of RQs

The scientific community has recently shown growing interest in the integration of Data Mining and Natural Language Processing (NLP) techniques with OPC-UA in industrial system architectures. This emerging approach has motivated our literature review to address several critical research questions that have not been fully explored in previous studies.

- $RQ_{1.}$: What are the current solutions that address the use of Data Mining and NLP techniques within OPC-UA-based industrial system architectures?
- RQ_2: What are the main Data Mining and NLP techniques used in industrial system architectures based on OPC-UA?
- RQ_3: What specific applications have been obtained or improved thanks to the use of Data Mining and NLP techniques in combination with OPC-UA? (For example, in quality control or predictive maintenance in industry).
- RQ_4: What are the main strengths, weaknesses, opportunities and threats (SWOT analysis) of using Data Mining and NLP techniques in OPC-UA based industrial system environments?
- RQ_5: What applications have been obtained or improved thanks to the use of Data Mining and NLP techniques in industrial system architectures?

2.3 Selection of information sources

The previously defined research questions guided the selection of key terms for the search process. These primary terms include (*opc-ua, opc ua*), (*NLP, natural language processing*), (*text mining, mining*), and (*text to speech, speech to text*). Using these terms, three search strings (SS) are strategically formulated to follow the structure described below:

- SS_1 :*(opc-ua OR "opc ua") AND (NLP OR "natural language processing")*;
- SS_2 :*(opc-ua OR "opc ua") AND ("data mining" OR "text mining" OR mining)*;

Table 2. Main primary information sources reviewed.

No.	Source	R. String	References	Type study	Year
E1	IEEE	SS2	Hastbacka et al. [8]	Conference	2014
E2	Scopus	SS2	Rix et al. [9]	Conference	2015
E3	IEEE	SS2	Srinivasan et al. [10]	Conference	2016
E4	IEEE	SS2	Fleischmann et al. [11]	Conference	2016
E5	IEEE	SS2	Sriyakul et al. [12]	Conference	2017
E6	Scopus	SS2	Cupek et al. [13]	Conference	2017
E7	IEEE	SS2	Gutiérrez et al. [14]	Conference	2017
E8	ACM	SS1	Hormann et al. [15]	Conference	2018
E9	Scopus	SS2	Kretschmer et al. [16]	Journal	2018
E10	ACM	SS2	Neu et al. [17]	Conference	2019
E11	IEEE	SS2	Bosi et al. [18]	Conference	2020
E12	IEEE	SS2	Qin et al. [19]	Conference	2020
E13	Scopus	SS2	Mathias et al. [20]	Conference	2020
E14	WOS	SS2	Vrana [21]	Journal	2020
E15	IEEE	SS2	Rubart et al. [22]	Conference	2020
E16	Scopus	SS2	Arevalo et al. [23]	Conference	2020
E17	ACM	SS3	Fuhrmann et al. [24]	Conference	2021
E18	Scopus	SS2	Soller et al. [25]	Conference	2021
E19	Scopus	SS1	Wu and Yang [26]	Conference	2022
E20	Scopus	SS2	Bakken [27]	Conference	2022
E21	IEEE	SS1	Tufek [28]	Conference	2023
E22	IEEE	SS1	Tufek et al. [29]	Conference	2023
E23	WOS	SS1	Bareedu et al. [30]	Journal	2024
E24	WOS	SS2	Hornsteiner et al. [31]	Journal	2024

– SS_3 :*(opc-ua OR "opc ua") AND (text AND to AND speech)*.

After defining the search strings (SS), the next step is to identify the information sources for executing the search queries. This is done to retrieve the most relevant studies on the integration of OPC-UA with techniques based on Data Mining and NLP, published in high-impact scientific journals. In line with the methodological frameworks outlined by [6,7] for conducting systematic reviews, four primary information sources (PIS) are selected. These included three specialized digital libraries in the field of Computer Science (PIS1) and one documentary database (PIS2, and PIS3).

– PIS_1: Association for Computing Machinery (ACM) Digital Library, available at: https://dl.acm.org/
– PIS_2: Institute of Electrical and Electronics Engineers (IEEE) Digital Library, available at: http://ieeexplore.ieee.org/.
– PIS_3: Scopus Database, available at: https://www.scopus.com/.
– PIS_4: Institute for Scientific Information (ISI) Web of Science Database, available at: https://webofknowledge.com/.

2.4 Selection of studies

The search process, conducted in December 2024, involved querying the titles of relevant publications using each of the specified search engines. This approach enabled us to filter out numerous studies related to OPC-UA that did not specifically focus on the integration of Data Mining and NLP techniques within industrial systems. Many of the excluded studies either addressed Data Mining or NLP in isolation or mentioned OPC-UA only in the state-of-the-art section, making them irrelevant to our investigation. Our primary focus is on studies where Data Mining and NLP techniques are actively applied within industrial system architectures for process automation.

The exploration and discovery process resulted in a total of 94 studies. After eliminating duplicates, 84 unique studies remained. Upon reviewing the titles and abstracts, 59 articles are deemed irrelevant to the focus of this study, leaving 24 studies for in-depth analysis. As a result, 24 studies are considered primary sources for the analysis in this systematic review. The details of these studies are presented in Table 2, and the outcomes of the selection process are summarized in Table 3.

Table 3. Search results based on the defined strings and information sources.

Search strings	PIS_1	PIS_2	PIS_3	PIS_4	*Sum studies*	*Delete studies*	*Repeat studies*	Unique *studies*
SS_1	7	5	11	1	24	15	4	5
SS_2	31	13	19	2	65	41	6	18
SS_3	5	0	13	0	5	4	0	1
Sum studies	43	18	30	3	94	60	10	24
Delete studies	41	4	14	0	–	–	–	–
Repeat studies	0	1	1	0	–	–	–	–
Unique studies	2	13	15	3	–	–	–	–

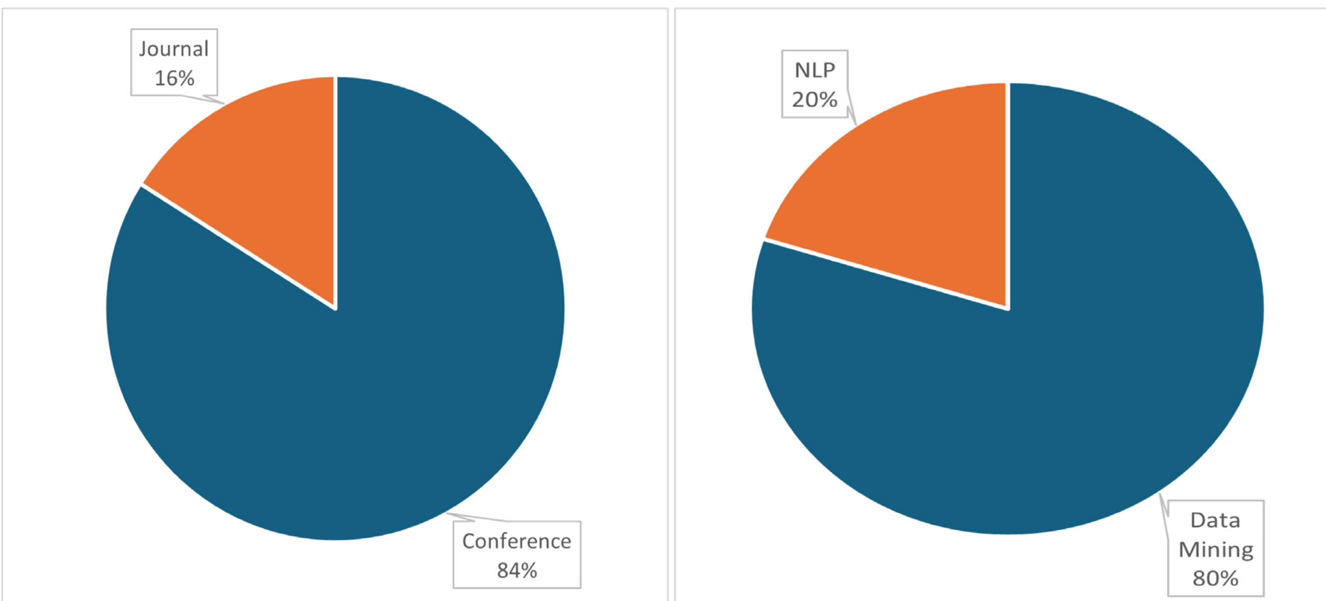


Fig. 1. (left) Distribution of the type of studies analyzed. (right) Distribution of the use of NLP and Data Mining in studies analyzed.

3 Results

In this section, the primary studies listed in Table 2 are analyzed and discussed, with the objective of extracting relevant information on the trends in the integration of Data Mining and NLP techniques with OPC-UA in the last decade, as well as its impact on solving real problems in industrial environments. Figure 1 *(left)* shows the distribution based on the type of studies, showing that 84% of the publications are conference type, while 16% represent journal articles. On the other hand, Figure 1 *(right)* shows that 80% of the articles analyzed are related to Data Mining techniques while 20% are focused on NLP techniques.

The results, shown in Figure 2, show the techniques based on Data Mining which appeared for the first time in 2014 and had growth in 2020 with 34% of the studies considered in this period for this type of technique, after the year 2020 a normalization of the curve is shown indicating a decrease in articles related to this topic.

On the other hand, the results, illustrated in Figure 3, show a constant growth in the number of publications related to the integration of NLP and OPC-UA techniques. Although the first research on the combination of these technologies began to emerge in 2021, the greatest impact occurred in 2023, with 50% of the studies considered primary data sources published during this period.

Table 4 shows a clear trend towards the use of Data Mining techniques to improve efficiency and quality in industrial environments, with a particular focus on predictive maintenance and process optimization. Of the works analyzed, the vast majority (80%) use Data Mining to analyze and extract patterns from industrial systems. This suggests that production systems are adopting advanced technologies to predict failures and optimize operations. An example of this trend is the first work presented by Hastbacka et al. [8] (E1), which uses semantic analysis and pattern recognition for predictive maintenance, while others such as Fleischmann et al. [11] (E4) focus on monitoring energy and temperature to improve process quality.

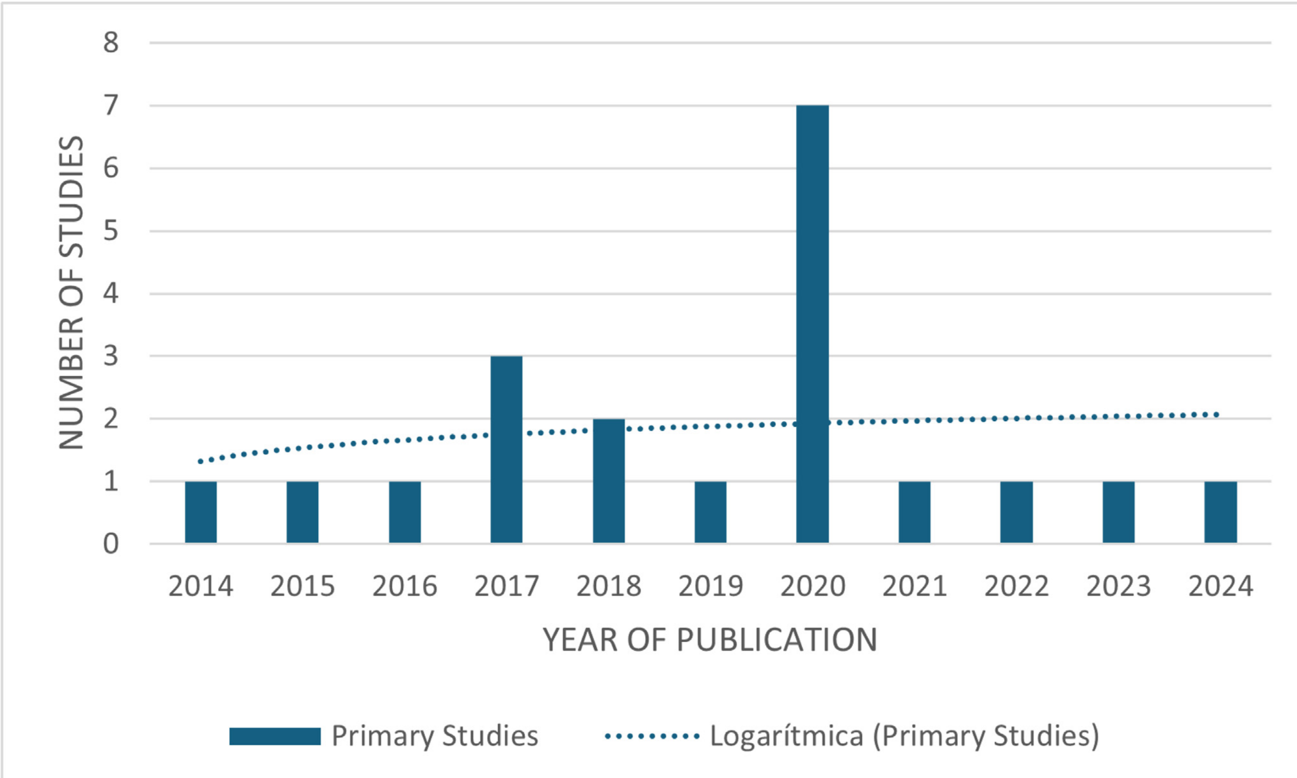


Fig. 2. Trends in the use of data mining techniques and OPC-UA based on current scientific publications.

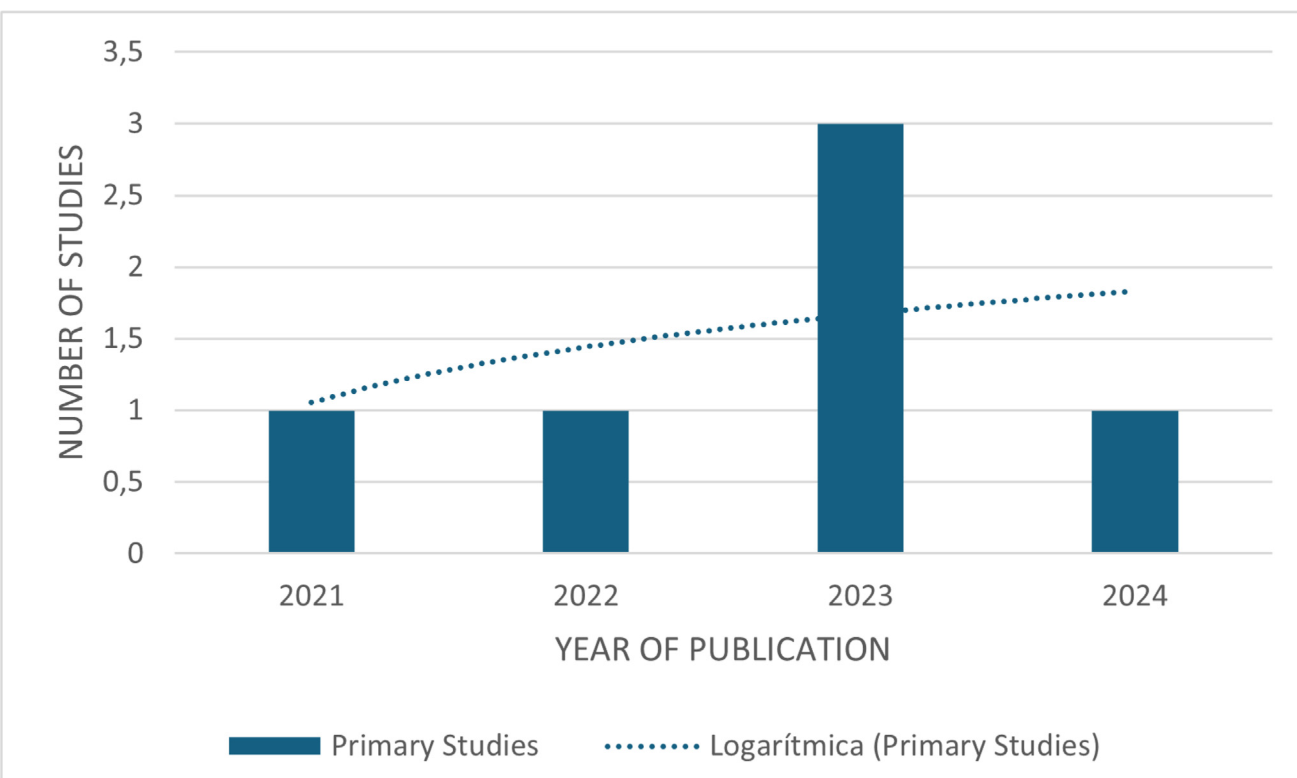


Fig. 3. Trends in the use of NLP techniques and OPC-UA based on current scientific publications.

Table 4. Main primary information sources reviewed.

No.	References	Application	Type	Cls.	Techique
E1	Hastbacka et al. [8]	Predictive maintenance	DM	TP	Semantic analysis of control systems and pattern recognition
E2	Rix et al. [9]	Predictive maintenance, quality monitoring	DM	PR	Predictive data analysis, integration of open standards
E3	Srinivasan et al. [10]	Industrial processes optimization	DM	PR	Predictive data analysis, pattern recognition
E4	Fleischmann et al. [11]	Predictive maintenance and quality monitoring	DM	RE	Data analysis of energy and temperature in crimping processes
E5	Sriyakul et al. [12]	Predictive maintenance and real-time product quality monitoring	DM	RE	Estimating unknown parameters, evaluates variables (energy, temperature) to detect wear conditions
E6	Cupek et al. [13]	Monitoring of energy consumption in discrete production lines	DM	CL	K-means clustering
E7	Gutiérrez et al. [14]	Automatic network configuration in real-time	DM	SP	Extraction of traffic parameters and learning in networks
E8	Hormann et al. [15]	Processes optimization and intrusion detection	DM	VI	Feature extraction, text processing using regular expressions
E9	Kretschmer et al. [16]	Data management	DM	LM	Persistent industrial data storage
E10	Neu et al. [17]	Security in industrial networks	DM	CA	J48 algorithm [37] in WEKA, traffic behavior analysis
E11	Bosi et al. [18]	Improving operational efficiency, Predictive maintenance, data integration	DM	VI	Feature extraction and analysis
E12	Qin et al. [19]	Quality monitoring of teaching	DM	NN	Pattern extraction and evaluation of educational data using GRU neural networks
E13	Mathias et al. [20]	Quality inspection of welding processes through the analysis of electrical signals	DM	CL, CA	Clustering algorithms, time series analysis and multi-label classifiers
E14	Vrana [21]	Nondestructive Evaluation (NDE), Industrial Internet of Things	DM	TP	Pattern extraction with digital twins and statistical models
E15	Rubart et al. [22]	Industrial processes optimization	DM	TP	Data analysis and semantic annotations
E16	Arevalo et al. [23]	Fault detection	DM	AR	Failure Mode and Effects Analysis for knowledge extraction
E17	Fuhrmann et al. [24]	Control and monitoring of injection molding machinery in industrial production environment	NLP	TOK, PAR	Voice interaction and intent recognition using RASA framework
E18	Soller et al. [25]	Predictive maintenance, anomaly detection in production	DM	OD, NN	One-Class Support Vector Machine, Isolation Forest and Autoencoder
E19	Wu and Yang [26]	Industrial Internet of Things, improving automation of data subscription in smart factories	NLP	TOK	Word2Vec model to calculate text similarity, collaborative filtering to evaluate similarity between subscribers and topics
E20	Bakken [27]	Sensor data extraction and processing in the electrical domain	DM	TP	Data extraction for analysis, using a language designed to simplify access to time series data

Table 4. (continued).

No.	References	Application	Type	Cls.	Techique
E21	Tufek [28]	Extracting and formalizing semantics from industrial standards to enhance machine interoperability	NLP	NER	Rule classification using NER and Machine Learning
E22	Tufek et al. [29]	Automate the extraction of compliance rules from OPC-UA companion specifications	NLP	NER	Semantic rule extraction using NER
E23	Bareedu et al. [30]	Semantic validation rules from industrial standards	DM, NLP	DT, TOK	Extract and analyze text, patterns and rules from structured and unstructured data
E24	Hornsteiner et al. [31]	End-of-line process of an automotive supplier (robotic inspections, laser engraving and cleaning)	DM	AR	Rule and model based techniques using network traffic

Furthermore, a growing integration of natural language processing (NLP) techniques in industrial systems is observed in the same Table 4, with emphasis on semantic extraction and automation. Recent research, such as that of Fuhrmann et al. [11] (E17) and Wu and Yang [26] (E19), highlight the use of NLP models for interaction with machines and improving automation in smart factories. In particular, the Word2Vec technique applied by Wu and Yang [26] (E19) helps calculate similarities between texts in smart factories, facilitating automatic data subscription. These advances suggest that natural language processing is gaining traction in the industry to improve machine interoperability and the automation of complex processes.

On the other hand, Table 4 selecting the records with *Type* equal to Data Mining shows in the column *Cls.* the classification of techniques defined using [32–34], below shows a small description and code of each one.

- Classification (CA): Used to categorize data based on predefined labels (e.g. J48, multi-label classifiers).
- Clustering (CL): Groups similar data without the need for labels (e.g. K-means, clustering algorithms).
- Tracking Patterns (TP): Detects patterns in data to obtain insights (e.g. semantic analysis, pattern extraction).
- Regression (RE): Analyzes relationships between independent and dependent variables (e.g. energy and temperature analysis).
- Outlier Detection (OD): Identifies anomalies that do not follow common patterns (e.g. One-Class SVM, Isolation Forest).
- Sequential Patterns (SP): Finds temporal or sequential relationships between data (e.g. learning traffic parameters).
- Prediction (PR): Combination of techniques to predict future events (e.g. predictive analysis).
- Association Rules (AR): Relates events or patterns between variables (e.g. FMEA, traffic-based rules).
- Visualization (VI): Helps visualize patterns and trends in the data (e.g. feature extraction).
- Neural Networks (NN): Uses neural networks to learn complex patterns (e.g. GRU, Autoencoder).
- Long-term Memory Processing (LM): Analyzes large volumes of historical data (e.g. persistent storage).

Similar to the previous paragraph, Table 4 selecting the records with *Type* equal to Natural Language Processing shows in the column *Cls.* the classification of techniques defined using [35,36], below shows a small description and code of each one.

- Tokenization (TOK): Breaking text into meaningful elements such as words or phrases.
- Parsing (PAR): Analysis of the grammatical structure of sentences to extract meaning.
- Named Entity Recognition (NER): Identification of entities such as proper names, places, organizations, etc.

Figure 4 *(left)* shows a word cloud based on the title of the 24 articles analyzed for this literature review. In this image it can be shown that the 5 words that attract attention without considering the basic words of the industrial context are *"subscription", "rules", "mining", "semantics" and "extraction"*. On the other hand, Figure 4 *(right)* shows the word cloud that contains one of the summaries of the analyzed articles. In this image you can see that the 5 words that catch your attention the most are *"semantics", "maintenance", "monitoring", "documents" and "quality"*.

4 Analysis

This section addresses the analysis of the data presented in Section 3 first the analysis of the research questions is performed and then a SWOT analysis is performed based on the articles collected in this study.

4.1 Analysis of research questions

This section answers each of the research questions posed at the beginning of this study, each of the answers is based on the 24 articles collected in this study.

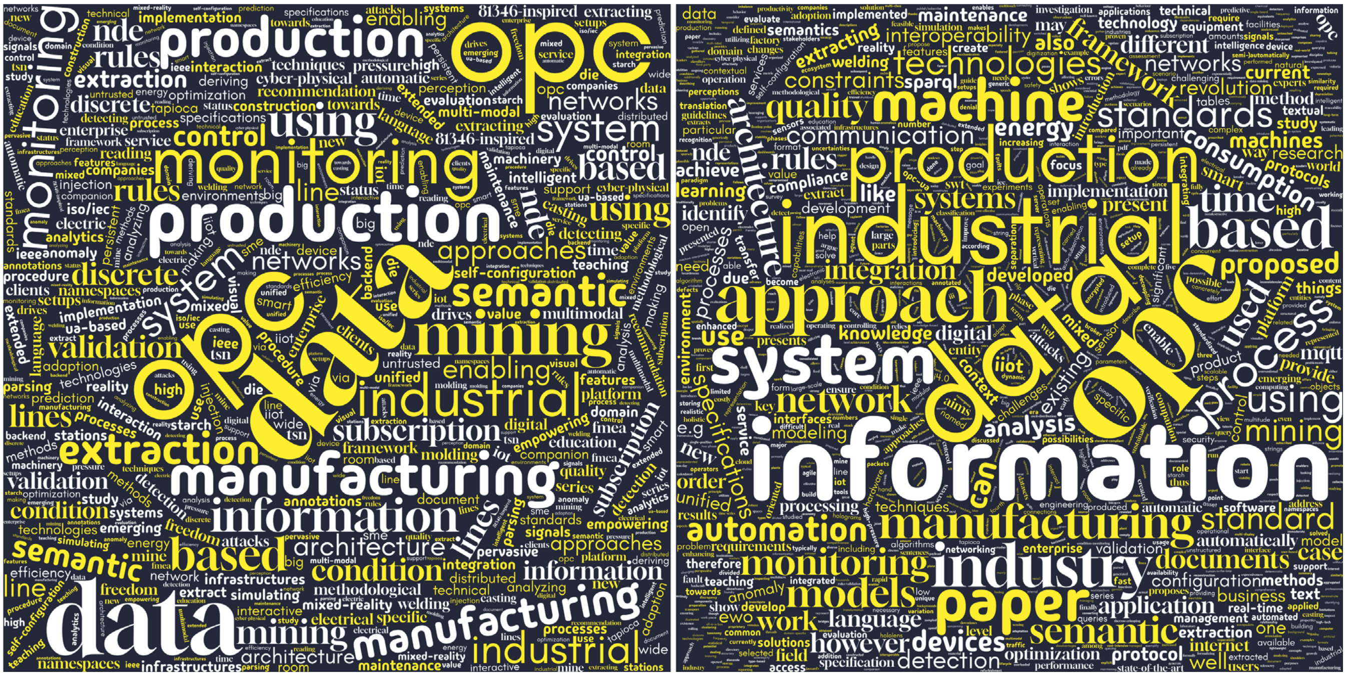


Fig. 4. (left) Word Cloud based on title. (right) Word Cloud based on abstract.

$RQ_{1.}$: **What are the current solutions that address the use of Data Mining and NLP techniques within OPC-UA-based industrial system architectures?** Currently, various solutions integrate Data Mining and NLP techniques within industrial system architectures based on OPC-UA. A significant example is the use of Data Mining techniques for the extraction and processing of sensor data, as seen in the work of Bakken [27] (E20), where data is analyzed in the context of industrial engineering. This approach allows companies to optimize data-based decision-making, thus improving their operational efficiency.

Furthermore, solutions have been developed that apply NLP for the extraction and formalization of semantics, as detailed in the study by Tufek [28] (E21). These techniques focus on improving machine interoperability through rule classification using Named Entity Recognition (NER). Automation of the extraction of OPC-UA [29] (E22) specification compliance rules has also been observed, representing a crucial advance in the standardization and optimization of industrial processes. These applications not only make data management easier but also promote more effective integration in complex environments.

RQ_{2}**: What are the main Data Mining and NLP techniques used in industrial system architectures based on OPC-UA?** Predominant Data Mining techniques in OPC-UA-based industrial architectures include predictive analytics and pattern recognition, as detailed in the works of Srinivasan et al. [10] (E3) and Hormann et al. [15] (E8). These techniques make it possible to identify anomalous behavior and predict failures, which is essential for process optimization as also evidenced in the work of Rix et al. [9] (E2) and Soller et al. [25] (E18). Furthermore, the use of clustering, as in the case of Cupek et al. [13] (E6), is essential to segment data and improve the efficiency of energy consumption in production. These techniques allow companies to anticipate problems and optimize processes, resulting in greater efficiency and reduced operating costs.

On the other hand, NLP techniques play a crucial role in improving communication and understanding between systems. Named entity recognition (NER) is used to extract semantic information from documents, as described in the research of Tufek et al. [28,29] (E21, E22). Likewise, text similarity models, such as Word2Vec, are used to evaluate the similarity between texts, thus facilitating the analysis of relevant information in industrial contexts [26] (E19). These techniques not only help in data management but also allow for more natural interaction between humans and machines.

$RQ_{3.}$: **What specific applications have been obtained or improved thanks to the use of Data Mining and NLP techniques in combination with OPC-UA?** (For example, in quality control or predictive maintenance in industry) The applications that have been obtained or improved through the use of Data Mining and NLP in industrial contexts are numerous. For example, predictive maintenance has benefited significantly, with studies showing how data analytics can predict failures and improve product quality [8,12,25] (E1, E5, E18). These techniques allow organizations to anticipate problems before they occur, resulting in reduced downtime and repair costs.

In addition, quality monitoring has improved significantly thanks to the application of Data Mining and semantic analysis techniques. Research such as that of Qin et al. [19] (E12) and Mathias et al. [20] (E14) illustrate how these tools are used for the inspection of industrial processes, such as welding, through the analysis of electrical signals. This approach not only guarantees the quality of the final product, but also optimizes the production process, contributing to greater efficiency and effectiveness in the industry.

$RQ_{4.}$: **What are the main strengths, weaknesses, opportunities and threats (SWOT analysis) of using Data Mining and NLP techniques in OPC-UA based industrial system environments?** The use of Data Mining and NLP techniques in OPC-UA-based industrial environments presents several significant strengths. Among them, the ability to extract valuable information from large volumes of data stands out, which improves informed decision making and allows companies to be more proactive in their operations. This analysis capability also helps increase operational efficiency and reduce costs, which is crucial in a competitive environment. However, there are also weaknesses that must be considered.

Implementing advanced techniques can be complex and require specialized technical skills, which can be a barrier for some organizations. Furthermore, the effectiveness of these techniques depends largely on the quality of the data used. Despite these challenges, there are numerous opportunities on the horizon, especially with the growth of the Internet of Things (IoT) and the digitalization of the industry. However, companies must be alert to threats, such as cybersecurity risks and regulatory changes that could impact the implementation of these technologies.

$RQ_{5.}$: **What applications have been obtained or improved thanks to the use of Data Mining and NLP techniques in industrial system architectures?** Predictive maintenance has significantly benefited from the integration of Data Mining techniques with OPC-UA in industrial environments. Studies by Hastbacka et al. [8], Rix et al. [9], and Fleischmann et al. [11] demonstrate how combining OPC-UA's robust communication capabilities with advanced data analysis can predict equipment failures before they occur. These approaches typically involve analyzing patterns in sensor data, such as energy consumption and temperature, to detect early signs of wear or potential breakdowns. This proactive approach allows companies to reduce downtime and maintenance costs by addressing issues before they lead to critical failures.

Quality monitoring and control processes have also seen substantial improvements through the integration of OPC-UA with Data Mining and NLP techniques. Research by Sriyakul et al. [12] and Mathias et al. [20] showcases how real-time data analysis of production parameters can lead to enhanced product quality. These studies often employ clustering algorithms, time series analysis, and multi-label classifiers to detect anomalies or deviations in manufacturing processes. By enabling rapid interventions based on this analysis, companies can maintain high-quality standards and optimize their production processes. Additionally, the work by Tufek et al. [29] and Bareedu et al. [30] demonstrates how NLP techniques can be used to extract and formalize semantics from industrial standards, potentially streamlining compliance processes and improving overall operational efficiency.

4.2 SWOT analysis

The selected primary studies are analyzed, organized, compared, and contrasted to identify the key strengths (S), opportunities (O), weaknesses (W), and threats (T) associated with NLP and Data Mining techniques in the context of OPC-UA. The following sections provide a detailed description of each of these four components.

4.2.1 Strengths

Technologies based on OPC-UA, Data Mining and NLP provide a high degree of interoperability in industrial environments, which facilitates the integration of different devices and systems from various suppliers. This interoperability improves efficiency in data collection and analysis, allowing companies to make more informed decisions and optimize their operations. In particular, Data Mining allows large volumes of data to be analyzed in real-time, which is essential for predictive maintenance and early failure detection, helping to reduce downtime and improve productivity. Furthermore, the use of NLP in human-machine interaction improves accessibility and reduces the need for complex interfaces, facilitating personnel training and accelerating the implementation of these technologies in industrial plants.

Another notable aspect of these technologies is their ability to implement scalable and flexible solutions. Data Mining tools allow companies to detect hidden patterns in operational data, helping to predict future trends and make decisions based on historical data. In turn, NLP facilitates the processing of large volumes of technical documents and industrial standards, accelerating semantic validation and regulatory compliance verification. These strengths make OPC-UA and its associated technologies not only useful in optimizing industrial processes, but also in continuously improving the quality and safety of operations.

4.2.2 Weaknesses

Despite their strengths, the adoption of these technologies requires significant investments in infrastructure and training. Data Mining techniques, although powerful, require large volumes of historical data to be truly effective, which can present a challenge for companies that do not have adequate data infrastructure or are just beginning to digitize their operations. Furthermore, the use of NLP in industrial settings has limitations, especially when it comes to interpreting specific contexts or industrial technicalities that can vary significantly between sectors. This creates a steeper learning curve and potential delay in implementation.

Another important weakness is the dependence on experts in advanced technologies. Implementing systems such as OPC-UA, along with Data Mining and NLP, requires trained personnel to configure, maintain and update these systems. This requirement for highly specialized personnel can be expensive and is not always available, especially in regions with limited technological resources. Additionally, the high initial costs of adoption, including the necessary equipment

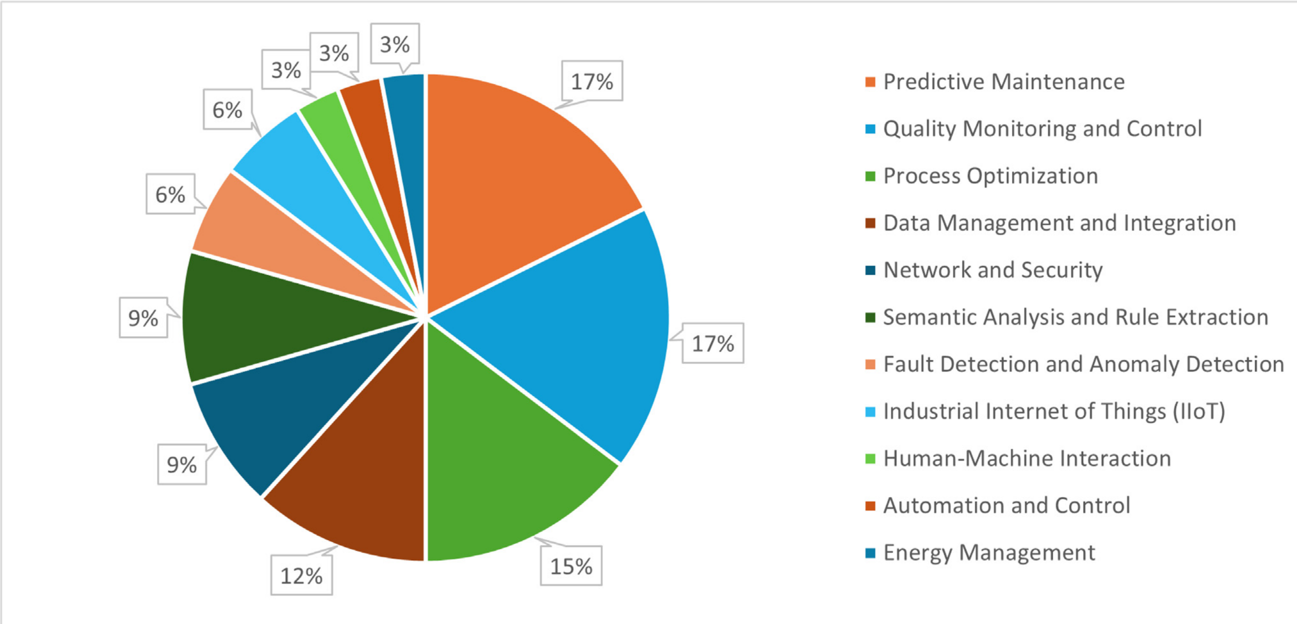


Fig. 5. Distribution of applications obtained or improved with the use of Data Mining and NLP techniques in industrial system architectures through the integration of OPC-UA.

and software, can be prohibitive for small and medium-sized businesses, limiting the reach of these technologies to large corporations with greater financial resources.

4.2.3 Opportunities

The rise of Industry 4.0 and smart factories provides a fertile environment for the implementation of Data Mining and NLP. As companies continue to digitize their processes, these technologies can be integrated to offer more automated and optimized solutions. The opportunities in predictive analytics are vast; Companies can use Data Mining to improve energy efficiency, predict failures before they occur, and optimize resource use. In sectors such as manufacturing and agriculture, Data Mining-based solutions can be instrumental in increasing productivity and reducing operating costs.

Furthermore, NLP has significant potential in improving human-computer interaction. The ability to interpret voice commands or process complex technical documents in real-time can facilitate the adoption of these technologies in a variety of industrial environments. There is also potential for NLP to be used to automate audit and compliance processes, improving accuracy and reducing administrative burden. The growth of cyber-physical technologies and the Industrial Internet of Things (IIoT) opens new opportunities to integrate Data Mining and NLP into industrial networks, improving connectivity and data analysis.

4.2.4 Threats

A significant threat is the increasing sophistication of cyber attacks, which affect industrial networks. Although OPC-UA provides advanced security mechanisms, attacks such as denial of service (DoS) and unauthorized access remain an ongoing concern. Security vulnerabilities in network-connected systems can compromise not only data integrity but also the secure operation of industrial machinery, potentially resulting in financial and reputational losses. Furthermore, the rapid evolution of technologies can quickly make current solutions obsolete, forcing companies to constantly update their systems.

Another threat lies in the lack of universal standardization in the use of Data Mining and NLP in the industry. Fragmentation in implementation approaches and different interpretations of standards can make interoperability between systems difficult, potentially leading to inefficiencies and compatibility issues. Additionally, increased automation and implementation of smart technologies may generate resistance from the workforce, as workers may perceive these technologies as a threat to their jobs, which could slow the adoption of these innovations.

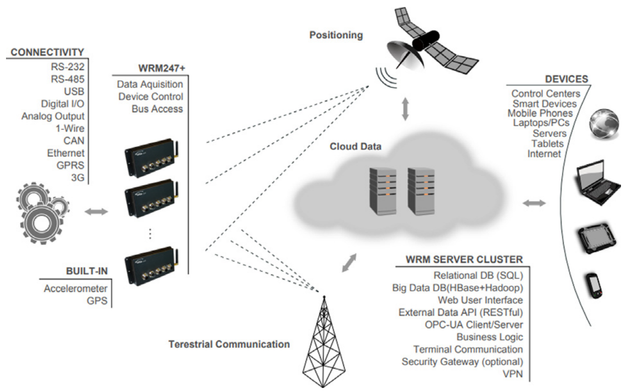
5 Applications

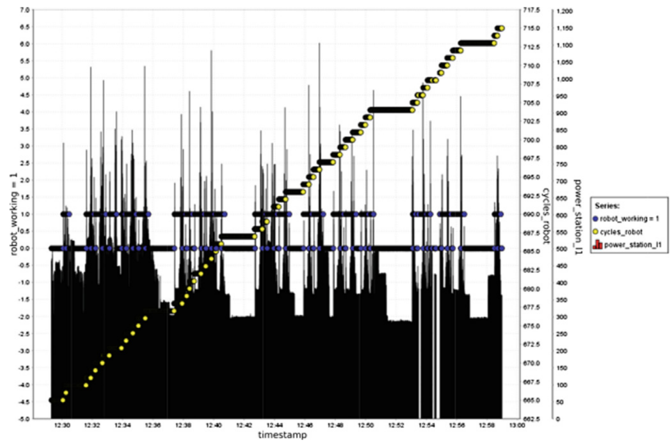
This section shows the applications found in the reviewed works of this work, the order shown is based on the number of publications that present the use of said applications. It is worth mentioning that the same application can appear in several publications. Figure 5 shows the percentage of applications analyzed in the present study. Additionally, Table 5 shows examples of different types of applications.

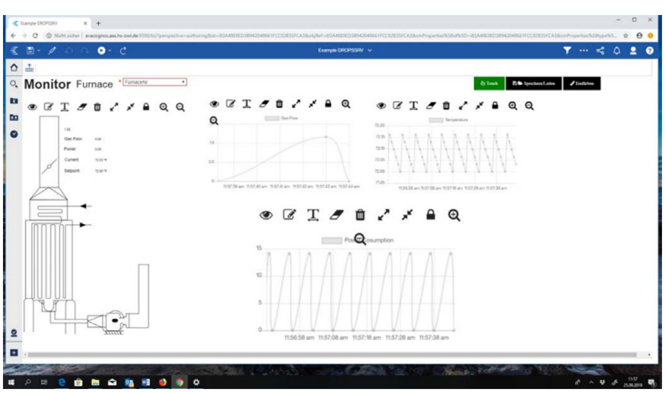
5.1 Predictive maintenance

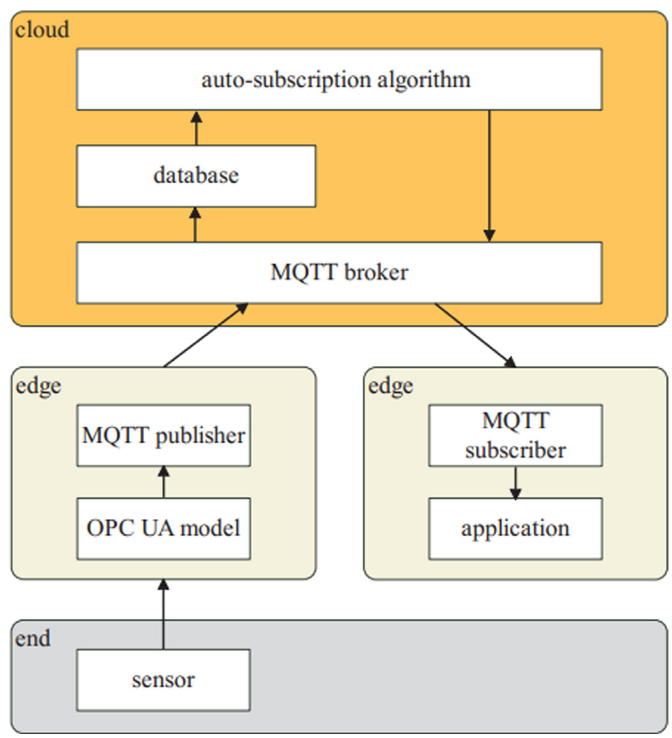
Predictive maintenance is a key application area where OPC-UA integration with Data Mining techniques shows significant promise. Studies like those by Hastbacka et al.

Table 5. Examples of applications found in the main primary information sources.

<table>
<tr><th>Application</th><th>Application</th></tr>
<tr><td>

Type: Predictive Maintenance
Title: Device status information service architecture for condition monitoring using OPC UA</td>
<td>

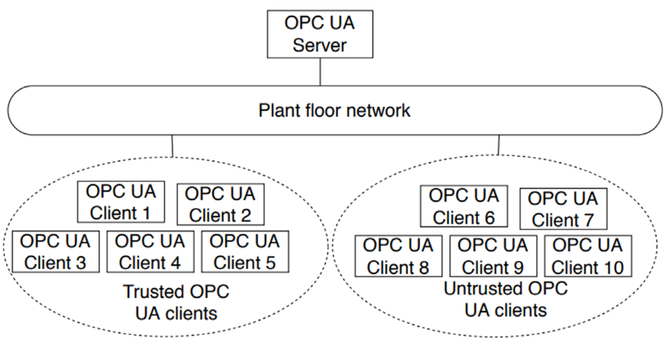
Type: Quality monitoring and control
Title: Data Mining Techniques for Energy Efficiency Analysis of Discrete Production Lines</td></tr>
<tr><td>

Type: Process optimization
Title: Predictive digital twin for offshore wind farms</td>
<td>

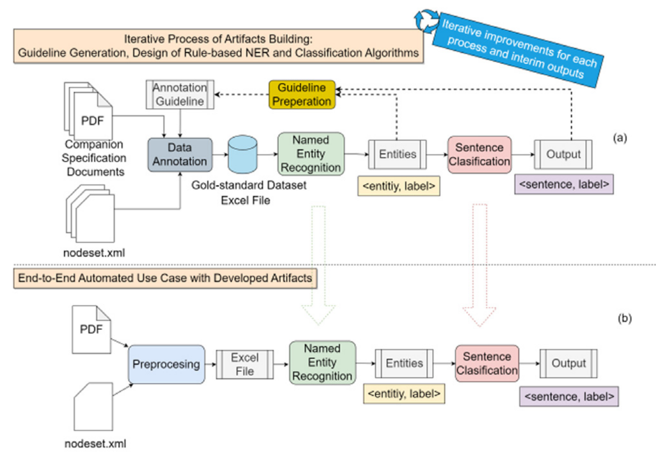
Type: Data management and integration
Title: Subscription Freedom: Automatic Industrial Data Subscription Based on Recommendation System</td></tr>
</table>

Table 5. (continued).

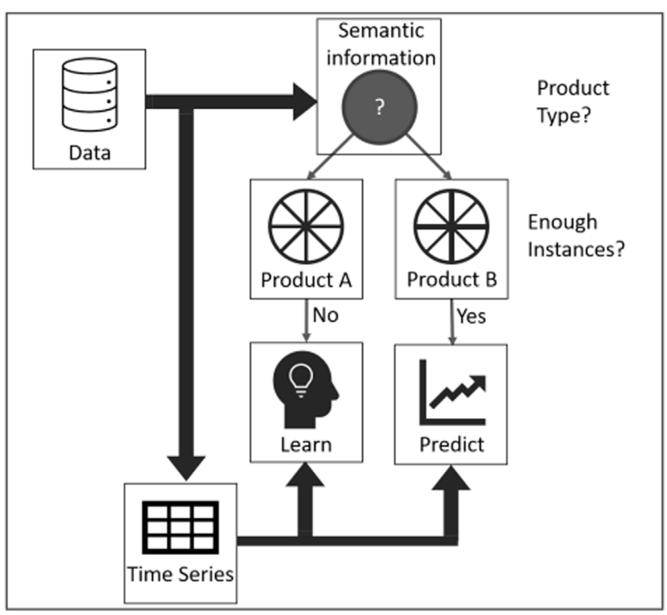
Application	Application



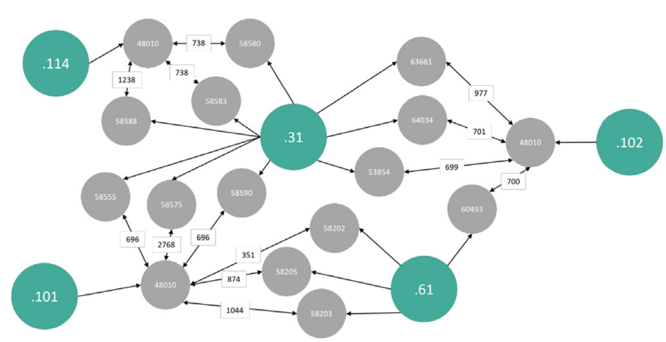
Type: Network and security
Title: Simulating and Detecting Attacks of Untrusted Clients in OPC UA Networks



Type: Semantic analysis and rule extraction
Title: Towards Extraction of Validation Rules from OPC UA Companion Specifications



Type: Fault detection and anomaly detection
Title: Evaluation and Adaption of Maintenance Prediction Methods in Mixed Production Line Setups based on Anomaly Detection



Type: Automation and control
Title: Reading between the Lines: Process Mining on OPC UA Network Data

[8], Rix et al. [9], and Fleischmann et al. [11] demonstrate how combining OPC-UA's robust communication capabilities with advanced data analysis can predict equipment failures before they occur. These approaches typically involve analyzing patterns in sensor data, such as energy consumption and temperature, to detect early signs of wear or potential breakdowns, thereby reducing downtime and maintenance costs.

5.2 Quality monitoring and control

The integration of OPC-UA with Data Mining and NLP techniques has proven valuable in enhancing quality monitoring and control processes. Research by Sriyakul et al. [12] and Mathias et al. [20] showcases how real-time data analysis of production parameters can lead to improved product quality. These studies often employ clustering algorithms, time series analysis, and multi-label classifiers to detect anomalies or deviations in manufacturing processes, enabling rapid interventions to maintain high-quality standards.

5.3 Process optimization

Process optimization is another crucial area benefiting from the convergence of OPC-UA and advanced analytics. Studies by Srinivasan et al. [10] and Rubart et al. [22] illustrate how predictive 5ata analysis and semantic annotations can be used to optimize industrial processes. These approaches often involve pattern recognition and feature extraction techniques to identify inefficiencies and suggest improvements, leading to enhanced productivity and resource utilization in manufacturing environments.

5.4 Data management and integration

OPC-UA's integration with Data Mining and NLP techniques is proving valuable in managing and integrating diverse industrial data sources. Studies like those by Kretschmer et al. [16] and Wu and Yang [26] demonstrate how these technologies can be used to create persistent industrial data storage solutions and improve data subscription automation in smart factories. These applications are crucial for handling the growing volumes of data in modern manufacturing environments and enabling more effective decision-making processes.

5.5 Network and security

The integration of OPC-UA with Data Mining techniques plays a crucial role in enhancing network configuration and security in industrial settings. Research by Gutiérrez et al. [14] and Neu et al. [17] shows how traffic analysis and machine learning algorithms can be used to automate network configurations and detect potential security threats. These approaches are essential for maintaining the integrity and efficiency of industrial communication networks in increasingly connected manufacturing environments.

5.6 Semantic analysis and rule extraction

Recent research, such as that conducted by Tufek et al. [29] and Bareedu et al. [30] focuses on using NLP techniques in conjunction with OPC-UA to extract and formalize semantics from industrial standards. These studies aim to enhance machine interoperability and automate the extraction of compliance rules from OPC-UA companion specifications, which could significantly streamline industrial automation processes and improve standardization efforts.

5.7 Fault detection and anomaly detection

Fault and anomaly detection represent critical applications of OPC-UA integration with Data Mining. Research by Arevalo et al. [23] and Soller et al. [25] showcases how techniques such as Failure Mode and Effects Analysis (FMEA) and machine learning algorithms like One-Class Support Vector Machines can be used to identify potential faults or anomalies in production processes. These approaches enable proactive interventions, reducing the risk of equipment failure and production disruptions.

5.8 Industrial internet of things (IIoT)

The application of OPC-UA in conjunction with Data Mining techniques is playing a significant role in advancing the Industrial Internet of Things. Studies by Vrana [21] and Wu and Yang [26] demonstrate how these technologies can be leveraged to create digital twins, enable smart data subscription, and facilitate seamless communication between diverse industrial devices. These applications are fundamental to realizing the vision of smart factories and Industry 4.0.

5.9 Human-machine interaction

While less common, the integration of OPC-UA with NLP techniques shows promise in enhancing human-machine interaction in industrial settings. The work by Fuhrmann et al. [24] illustrates how voice interaction and intent recognition can be implemented in industrial production environments, potentially improving operator efficiency and ease of control in complex manufacturing processes.

5.10 Automation and control

The integration of OPC-UA with Data Mining techniques is also being applied to enhance automation and control in specific industrial processes. For instance, the work by Hornsteiner et al. [31] demonstrates the application of rule and model-based techniques using network traffic analysis in end-of-line processes of automotive suppliers. Such applications can lead to more efficient and precise control of complex manufacturing operations.

5.11 Energy management

While less represented in the reviewed literature, energy management is an important application area. The work by Cupek et al. [13] demonstrates how OPC-UA can be combined with clustering techniques like K-means to monitor and optimize energy consumption in discrete production lines. This application has significant potential for improving energy efficiency in industrial settings, contributing to both cost savings and environmental sustainability.

6 Conclusions

Current applications of integrating OPC-UA with advanced Data Mining and NLP techniques have shown significant impact on improving energy efficiency, real-time monitoring and more informed decision making in industrial processes. Especially, predictive maintenance and production anomaly detection have benefited from these technologies, optimizing machine downtime and anticipating failures before they occur. Techniques such as network traffic analysis and semantic extraction are facilitating the management and persistent storage of industrial data, improving the ability of smart factories to manage large volumes of data.

Regarding future directions, it is essential to advance the standardization of semantic models and ensure greater interoperability between systems. Current challenges include the need to unify different industrial protocols and improve integration between machine control systems and artificial intelligence platforms. Additionally, the incorporation of more advanced machine learning models and the expansion of the use of NLP in industrial regulatory compliance will allow for more complete and precise automation of processes.

Current trends point towards widespread adoption of digital twins in combination with OPC-UA, opening up new possibilities for real-time process simulation and optimization. This trend promises to transform factories into fully automated environments, where continuous monitoring and immediate feedback allow automatic adjustments to manufacturing processes. Additionally, the use of emerging technologies such as recurrent neural networks and cloud computing for data analysis promises to increase the speed and accuracy of detecting industrial problems.

Finally, although the integration of OPC-UA with artificial intelligence and NLP is in an advanced phase, the challenge lies in continuing to improve the robustness of the algorithms and guarantee security in connected industrial systems. The evolution towards an ecosystem of fully interconnected and smart factories will require a continued focus on cybersecurity, ensuring that technological advances are accompanied by robust protection against potential vulnerabilities.

7 Future directions

Building upon the findings of this systematic review, several key areas emerge as promising directions for future research and development.

There is a critical need to advance the standardization of semantic models in industrial environments. This will ensure greater interoperability between systems and facilitate more fluid integration of various industrial protocols. Creating unified standards for the semantic representation of industrial data could significantly accelerate the adoption of smart technologies in manufacturing. Incorporating more advanced machine learning models, particularly in the realm of deep learning, could significantly improve the accuracy and efficiency of predictive maintenance and anomaly detection systems. The development of algorithms capable of handling the complexity and variability of industrial data in real time is a promising area of research.

Further exploration of NLP applications in industrial regulatory compliance and technical documentation analysis could lead to more automated and efficient processes in manufacturing environments. The ability to automatically extract and formalize compliance rules from technical specifications could revolutionize the way industries handle regulatory compliance. As industrial systems become more interconnected, developing robust cybersecurity measures tailored to OPC-UA and IIoT environments will be crucial to protect against evolving threats. Research into security techniques that can scale with the increasing complexity of industrial networks is a priority.

Research on the integration of OPC-UA and data analytics with emerging technologies such as 5G, edge computing and blockchain could open new avenues for industrial process optimization and data management. These technologies have the potential to improve the speed, safety and efficiency of industrial operations. Advancing research in voice interaction and intent recognition in industrial environments could lead to more intuitive and efficient human-machine interfaces in complex manufacturing processes. This could significantly improve the usability and accessibility of industrial control systems.

As the volume of industrial data continues to grow, research into scalable architectures and performance optimization for real-time data processing and analysis will be essential. Developing solutions that can efficiently handle large volumes of heterogeneous data is crucial for the future of smart industry. Exploring methods to transfer knowledge and models between different industrial domains could accelerate the adoption and effectiveness of these technologies in various manufacturing sectors. Creating transferable learning models could significantly reduce the time and resources required to implement smart solutions in new industrial contexts.

Acknowledgments

This research has been partially supported by University of Granada and the ESPOL Polytechnic University.

Funding

This research has been supported by the ESPOL project “Automatización del proceso de detección de fallas en piezas de hojalata usando visión por computador” (CIDIS-004-2023).

Conflicts of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

All data generated or analyzed during this study are included in this published article.

Author contribution statement

Henry O. Velesaca: Data Collection, Analysis and Writing—original draft preparation; Juan A. Holgado-Terriza: Conceptualization, Methodology, Supervision, Writing—review and editing; All authors read and approved the final manuscript.

References

1. H.O. Velesaca, J.A. Holgado-Terriza, J.M. Gutierrez Guerrero, Optimizing smart factory operations: a methodological approach to industrial system implementation based on opc-ua, E3S Web Conf. **532** (2024) 02004 https://doi.org/10.1051/e3sconf/202453202004
2. H.O. Velesaca, D. Carrasco, D. Carpio, J.A. Holgado-Terriza, J. Gutierrez-Guerrero, T. Toscano, A. Sappa, Anomaly detection in industrial production products using opc-ua and deep learning, in *Int. Conf. on Data Science, Technology and Applications*. INSTICC, SciTePress (2024), pp. 505–512, https://doi.org/10.5220/0012812600003756
3. A. Dogan, D. Birant, Machine learning and data mining in manufacturing, Exp. Syst. Appl. **166** (2021) 114060, https://doi.org/10.1016/j.eswa.2020.114060
4. H.O. Velesaca, A.D. Sappa, J.A. Holgado-Terriza, A case study of anomaly detection in tinplate lids: Supervised vs unsupervised approaches, in *Int. Conf. on Automation, Robotics and Applications*. IEEE (2025), pp. 1–5
5. M.C. May, J. Neidhöfer, T. Körner, L. Schäfer, G. Lanza, Applying natural language processing in manufacturing, Proc. CIRP **115** (2022) 184–189, https://doi.org/10.1016/j.procir.2022.10.071
6. S. Keele et al., Guidelines for performing systematic literature reviews in software engineering. In *Technical Report, version 2.3 ESBE Technical Report*, Keele University and Durham University, 1–15 (2007)
7. A. Kofod-Petersen, How to do a structured literature review in computer science. Ver. 0.1. October, Copenhagen: Alexandra Institute, **1**:28 (2012)
8. D. Hästbacka, L. Barna, M. Karaila, Y. Liang, P. Tuominen, S. Kuikka, Device status information service architecture for condition monitoring using opc ua, in *Emerging Technology and Factory Automation* (2014), pp. 1–7, https://doi.org/10.1109/ETFA.2014.7005141
9. M. Rix, B. Kujat, C. Buscher, T. Meisen, S. Jeschke, A methodological implementation of a pervasive information system in high pressure die casting manufacturing. In Int. Conf. for Production Research, 1–8 (2015)
10. S. Srinivasan, D. Grobmann, C. Del Vecchio, V. Emila Balas, L. Glielmo, Enabling technologies for enterprise wide optimization (2016), pp. 434–439, https://doi.org/10.1109/ICIINFS.2015.7399051
11. H. Fleischmann, S. Spreng, J. Kohl, D. Kiskalt, J. Franke, Distributed condition monitoring systems in electric drives manufacturing, in *Int. Electric Drives Production Conference* (2016), pp. 52–57, https://doi.org/10.1109/EDPC.2016.7851314
12. H. Sriyakul, D. Koolpiruck, W. Songkasiri, S. Nuratch, Cyber-physical system based production monitoring for tapioca starch production, in *Int. Conf. on Information Science and Control Engineering* (2017), pp. 926–930, https://doi.org/10.1109/ICISCE.2017.196
13. R. Cupek, J. Duda, D. Zonenberg, U. Chopa's, G. Dziedziel, M. Drewniak, Data mining techniques for energy efficiency analysis of discrete production lines, in *Int. Conf. Computational Collective Intelligence*. Springer (2017), pp. 292–301
14. M. Gutiérrez, A. Ademaj, W. Steiner, R. Dobrin, S. Punnekkat, Self-configuration of ieee 802.1 tsn networks, in *Int. Conf. on Emerging Technologies and Factory Automation* (2017), pp. 1–8, https://doi.org/10.1109/ETFA.2017.8247597
15. R. Hormann, S. Nikelski, S. Dukanovic, E. Fischer, Parsing and extracting features from opc unified architecture in industrial environments, in Proceedings of the 2nd International Symposium on Computer Science and Intelligent Control, ISCSIC '18, New York, NY, USA (2018), https://doi.org/10.1145/3284557.3284741
16. F. Kretschmer, C. von Arnim, A. Lechler, A. Verl, Persistent data backend for opc ua namespaces in it infrastructures, Proc. CIRP **72** (2018) 174–178, https://doi.org/10.1016/j.procir.2018.03.233
17. C. Varlei Neu, I. Schiering, A. Zorzo, Simulating and detecting attacks of untrusted clients in opc ua networks, in Central European Cybersecurity Conference, CECC 2019, Association for Computing Machinery, New York, NY, USA (2019), https://doi.org/10.1145/3360664.3360675
18. F. Bosi, A. Corradi, G. Di Modica, L. Foschini, R. Montanari, L. Patera, M. Solimando, Enabling smart manufacturing by empowering data integration with industrial iot support, in *Int. Conf. on Technology and Entrepreneurship* (2020), pp. 1–8, https://doi.org/10.1109/ICTE47868.2020.9215538
19. F. Qin, W. Zeng, L. Li, R. Zhao, Construction of big data monitoring platform for teaching quality under intelligent education, in *Int. Wireless Communications and Mobile Computing* (2020), pp. 1594–1597, https://doi.org/10.1109/IWCMC48107.2020.9148224
20. S.G. Mathias, S. Schmied, D. Grossmann, Monitoring of discrete electrical signals from welding processes using data mining and IIoT approaches, in *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), Baltimore, MD, USA* (2020), 911–916, https://doi.org/10.1109/ICTAI50040.2020.00142
21. J. Vrana, Nde perception and emerging reality: Nde 4.0 value extraction, Mater. Evaluat. **78** (2020) 835–851, https://doi.org/10.32548/2020.me-04131

22. J. Rubart, B. Lietzau, P. Söhlke, Analyzing manufacturing data in a digital control room making use of semantic annotations, in *Int. Conf. on Semantic Computing* (2020), pp. 434–438, https://doi.org/10.1109/ICSC.2020.00084
23. F. Arevalo, D. Sunaringtyas, C. Tito, C. Piolo, A. Schwung, Interactive visual procedure using an extended FMEA and mixed-reality (2020), pp. 286–291, https://doi.org/10.1109/ICIT45562.2020.9067296
24. F. Fuhrmann, A. Weber, S. Ladstätter, S. Dietrich, J. Rella, Multimodal interaction in the production line – an OPC UA-based framework for injection molding machinery, in *Int. Conf. on Multimodal Interaction, ICMI' 21. Association for Computing Machinery, New York, NY, USA* (2021), pp. 837–838, https://doi.org/10.1145/3462244.3481300
25. S. Soller, B. Fleischmann, M. Kranz, G. Holzl, Evaluation and adaption of maintenance prediction methods in mixed production line setups based on anomaly detection (2021), pp. 520–525, https://doi.org/10.1109/PerComWorkshops51409.2021.9430995
26. Y. Wu, B. Yang, Subscription freedom: automatic industrial data subscription based on recommendation system. (2022), pp. 3614–3619, https://doi.org/10.1109/CAC57257.2022.10055861
27. M. Bakken, An iso/iec 81346-inspired domain specific language to extract time series data for analytics (2022), pp. 202–209, https://doi.org/10.1109/SII52469.2022.9708860
28. N. Tufek, Semantic information extraction from multi-modal technical document (2023), https://doi.org/10.23919/CISTI58278.2023.10211635
29. N. Tufek, A. Sai Sree Thuluva, V. Philipp Just, M. Sabou, Towards extraction of validation rules from opc ua companion specifications, in *Int. Conf. on Emerging Technologies and Factory Automation* (2023), pp. 1–8, https://doi.org/10.1109/ETFA54631.2023.10275371
30. Y.S. Bareedu, T. Frühwirth, C. Niedermeier, M. Sabou, G. Steindl, A. Saisree Thuluva, S. Tsaneva, N. Tufek Ozkaya, Deriving semantic validation rules from industrial standards: an opc ua study, Semantic Web **15** (2024) 517–554, https://doi.org/10.3233/SW-233342
31. M. Hornsteiner, P. Empl, T. Bunghardt, S. Schönig, Reading between the lines: process mining on opc ua network data, Sensors **24** 1–15 (2024), https://doi.org/10.3390/s24144497
32. S.-H. Liao, P.-H. Chu, P.-Y. Hsiao, Data mining techniques and applications-a decade review from 2000 to 2011, Exp. Syst. Appl. **39** (2012) 11303–11311, https://doi.org/10.23919/IConAC.2017.8082090
33. E.W.T. Ngai, L. Xiu, D.C.K. Chau, Application of data mining techniques in customer relationship management: a literature review and classification, Expert Syst. Appl. **36** (2009) 2592–2602, https://doi.org/10.1016/j.eswa.2008.02.021
34. N. Jain, V. Srivastava, Data mining techniques: a survey paper, Int. J. Res. Eng. Technol. **2** (2013) 2319–1163
35. P.M. Lavanya, E. Sasikala, Deep learning techniques on text classification using natural language processing (nlp) in social healthcare network: a comprehensive survey, in Int. Conf. on Signal Processing and Communication (ICPSC) (2021), pp. 603–609, https://doi.org/10.1109/ICSPC51351.2021.9451752
36. J. Sawicki, M. Ganzha, M. Paprzycki, The state of the art of natural language processing—a systematic automated review of nlp literature using nlp techniques, Data Intelligence **5** (2023) 707–749, https://doi.org/10.1162/dint_a_00213
37. L. Bilge, S. Sen, D. Balzarotti, E. Kirda, C. Kruegel, Exposure: a passive DNS analysis service to detect and report malicious domains, ACM Trans. Inf. Syst. Secur. **16** (2014) 1–28, https://doi.org/10.1145/2584679

Cite this article as: Henry O. Velesaca, Juan A. Holgado-Terriza, OPC-UA in artificial intelligence: a systematic review of the integration of data mining and NLP in industrial processes, Manufacturing Rev. **12**, 9 (2025), https://doi.org/10.1051/mfreview/2025003