Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

DEMAND PREDICTABILITY EVALUATION FOR SUPPLY CHAIN PROCESSES USING SEMANTIC WEB TECHNOLOGIES USE CASE

Nour Ramzy
Philipp Ulrich
Lancelot Mairesse
Hans Ehm

Supply Chain Innovation, Semantic Web
Infineon Technologies AG
Am Campeon 1-15
Neubiberg, 85579, GERMANY

ABSTRACT

Semantic Web technologies provide the possibility of a common framework to share knowledge across supply chain networks. We explore Semantic Web technologies to evaluate processes' demand predictability through a use case. First, we create an ontology describing the relevant domain concepts and data and define competency questions based on the need of the use case. Then, we map the data to the ontology in a knowledge graph. We design a chain of SPARQL queries to retrieve and insert information from the knowledge graph to answer the competency questions. We calculate the underlying demand for supply chain processes using aggregations and the created semantic description. We successfully computed a pre-defined metric for demand predictability for different time scopes and process groups: the mean of yearly coefficient of variation. Using this approach, one could perform predictability evaluations for relevant indicators of end-to-end supply chains by further data integration in the semantic framework.

1 INTRODUCTION

The idea of describing an enterprise's knowledge using ontologies and knowledge graphs has been explored for more than twenty years. Knowledge is one of the critical resources of an organization, and its management impacts all operations within the enterprise (Loucopoulos and Kavakli 1999). Knowledge implies a contextualized and reusable understanding of associated information, which can then be used as a framework to assess and address new situations. Data is supposed to be factual, but how it is inherited, processed, structured, and interpreted is prone to subjective bias. Biases are functional and necessary to explore meanings of data and construct information but are possible pitfalls when creating reusable and well-defined knowledge. The same data can lead to inconsistent or conflicting information as the inherited knowledge is conditioned by all biases within the information chain.

Historically, information systems for organizations were aligned with activities and solution-oriented. The specificity of each use case and technological diversity led to significant heterogeneity of information systems across organizations and departments. The acceleration of globalization and the internet has led to complex, global end-to-end supply chains. Establishing and optimizing global networks in supply chains requires larger coordination scopes, quantity, and precision. Many tools and data support decisions along supply chains, but an overall lack of visibility and overview exists. Information is partitioned between different groups of the same end-to-end supply chain. Industry 4.0 refers to the fourth industrial revolution, characterized by growing trend in automation and data exchange in the manufacturing field. It aims at

978-1-6654-7661-4/22/$31.00 ©2022 IEEE

optimizing cross-departments and cross-organization cooperation. Challenges can not be considered as only distinct entities in such interconnected systems but also need to be met on a global layer. These challenges raise the need for data integration and contextualization in unprecedented scopes.

Data sources can have semantic, schematic, or syntactic heterogeneity (Bishr 1998). Data integration aims at combining heterogeneous data from multiple sources into a unified system shared by all users. One can share, access, and re-use the information within one integrated view. Multiple approaches have been developed over time and combined, depending on the challenges and accessible technologies (Ziegler and Dittrich 2007). Data integration is enhanced by data contextualization. Abowd, Dey, Brown, Davies, Smith, and Steggles (1999) formulated a vastly adopted definition of context in the computing field:" Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." Contextualization enables the integration of heterogeneous data with a reduced loss of information and can improve data processing efficiency (Yavari, Jayaraman, Georgakopoulos, and Nepale 2017). One can even gain additional information with the inferred knowledge.

Semantic data integration can handle both data integration and contextualization by combining isolated and heterogeneous data sources and make them accessibly in a unified way. With semantics, data is given a well-defined meaning which allows both humans and computers to make meaningful interpretations of the provided information. Semantic Web technologies enable building a formal and inclusive framework describing concepts, relations, and associated underlying data. Relevant data silos within an enterprise and external relevant shared data can be integrated into knowledge graphs. One can also access external data end-points and use the shared data without integrating, storing, and administering it.

This paper presents the possibility of using Semantic Web technologies to support a concrete business need: gain knowledge about demand predictability for supply chain processes. For an enterprise, demand forecasting provides critical knowledge. All the necessary processes and assets (human resources, raw materials, fabrics) essential to supply the demand can be impacted by under-evaluated supplies or a waste of resources for an evaluated market. In fields such as production planning and asset management, many decisions will aim for production capacity optimization versus the demand. However, demand forecasting can be very complex, and misleading predictions can lead to wrong decisions. Accuracy of forecasts can vary significantly between products or over time, as demand planning can face a tremendous amount of factors that can impact the demand of all products or only some. The underlying demand for each process required to manufacture all enterprise products can vary in predictability.

Evaluating the predictability of demand forecasting across different resources and product views has multiple benefits in price agreements, strategic and tactical planning, contractual commitments for resources, and outsourcing strategy. Demand predictability provides information to understand better the limitations of the methodology employed to forecast demand and the usability of the inherited knowledge. It also enables insights to decide which processes or resources should be internal or external to the company and on which terms. For contractual commitments, one would consider first short contractual commitment for a group of resources if the underlying demand is unpredictable. One can also evaluate how accurate forecasts have been depending on when they were generated or any other correlation considered.

However, evaluating the predictability of demand forecasting across different resources and product views requires knowing the relations between products for which demand is forecasted and the necessary resources (processes, assets). The required data is often part of data silos. This problem is amplified for companies having complex and worldwide end-to-end supply chains. A pre-study was performed within the Infineon Supply Chain Innovation Team to define an approach and a metric, among many possible, to evaluate the predictability of processes. The yearly mean coefficient of demand variation for a given process was used. This statistical approach has been validated and followed using Semantic Web technologies. The statistical analyses required to gain this knowledge are performed using multiple data sources generated by different agents in the company and several process views (sales product, process groups, procurement processes).

2 RELATED WORK

Semantic Web is a field of interest in supply chain management and enabler for various use cases across industries.

Pal (2017) presented components of an ontology-based Semantic Web Service architecture and then implemented a prototype for material procurement systems of a manufacturing supply chain. Another architecture that can facilitate the process of autonomous supplier discovery in distributed environments has been proposed by Ameri and Patil (2012).

Landolfi et al. (2018) introduced an ontology-based semantic data model supporting a *Manufacturing as a Service platform*. All supply chain stakeholders are considered, including customers, suppliers, regulators, and consumers. Platforms allow matching by production capacity, know-how and by-product. Other functionalities are certifications management, innovation management, and ecosystem optimization.

Our work is related to *Business Intelligence*. We use Semantic Web technologies to perform statistical analyses on time series to support business decisions. For these needs, a typical solution for business intelligence is data warehousing combined with on-line analytical processing (OLAP) technology. (Chaudhuri and Dayal 1997)

Semantics has already been implemented in OLAP approaches. Quamar et al. (2020) proposed an ontology-driven approach to conceive a BI conversational system. It uses the conventional OLAP model to represent and process multidimensional data. The data cube vocabulary (Cyganiak, Reynolds, and Tennison 2014) is used to transcribe data cubes definitions in RDF format. The semantic framework allows the user to generate structured queries dynamically, explore their data and create charts by natural language querying. This demonstrates the benefits of semantics in enhancing the flexibility and expressivity of the end-user solution.

We use forecasts time series and the coefficient of variation to evaluate demand predictability in our work. The coefficient of variation is a commonly used measure to evaluate demand volatility. Evaluating volatility helps to understand demand patterns better and apply relevant strategies in supply chain management. (Packowski (2013))

A higher coefficient of variation is associated with more uncertainties in demand behavior and causes difficulties in forecasting (Huang, Chang, and Chou (2008)). Abolghasemi, Mahdi and Gerlach, Richard and Tarr, Garth and Beh, Eric (2019) use this metric to measure the volatility variations in demand. They divided demand series into three volatility groups based on their coefficient of variation and compared the accuracy of different forecasting standards models depending on their volatility groups. The results show that the volatility of demand may significantly change forecasts accuracy. Then they empirically investigated the most suitable forecasting model depending on the coefficient of variation of the time series. In the conclusion of their article, they address the possibility of analyzing why some models work better than others in certain conditions. A semantic framework might be the best suited to accomplish this task as all relevant data can be integrated and well defined in the framework. Abolghasemi, Mahdi and Gerlach, Richard and Tarr, Garth and Beh, Eric (2019) used Semantic Web technologies to quantify the semantic content of words in news articles about a company and use this as a predictor of its stock volatility. The use case showed better results than standard models based only on historical volatilities.

3 USE CASE AND MOTIVATION

In this section, we explain the relevance of demand predictability evaluation in the semiconductor industry as well as the use case and preliminary analysis on which we base our work.

3.1 Demand predictability for supply chain processes

The semiconductor industry is characterized by high capital intensity; a manufacturer has to invest substantially in its assets to generate revenue. A new factory can cost more than 1 billion euros. Furthermore, the Covid-19 pandemic has accelerated the digitization trend and created a push in demand, creating competition for foundry capacity. Demand forecasting and planning support the constant trade-off between production capacity and financial risk. However, semiconductor demand is challenging to forecast. Short product life cycle and technological innovation induce complexity in asset management. Semiconductor manufacturers are positioned early in the value chains for products sold to the final consumer and thus have limited demand visibility. Hence, they are particularly affected by the Bullwhip Effect, which increasingly distorts the demand as one moves upstream along the supply chain; from end consumer to early suppliers. Identified causes are demand uncertainty, complexity in supply chain processes, and communication restrictions.

One way to increase flexibility in the capacity is to rely on outsourced processes. The use of an outsourced process and the associated terms and conditions are chosen with many factors considered. One of the critical factors is how well the underlying demand for the Sales Product requiring an outsourced process is predictable. As discussed in the introduction, evaluating the predictability of demand forecasting across different processes and product views has multiple benefits. In the use case context, demand predictability has been defined as crucial knowledge to apply different foundry management strategies in topics such as order management and obligation management. The first focus is on outsourced management strategy, defining why and how an external organization should carry out a process required to manufacture products. This can, for example, be translated into more or less flexible contracts.

Three approaches were considered for high volatility identification:

1. High Fluctuations in customer demand for technologies using outsourced foundry processes
2. Forecasts frequently change in customer demand
3. Low forecast accuracy

In previous work at Infineon, a first study was conducted to define a metric for evaluating front-end outsourced processes predictability. The chosen metric is the mean of the yearly coefficient of variation of the sum of forecasted demand for sales product per *Procurement Processes Category* (PPC). This quantifies the average demand volatility for an outsourced process over a year, as illustrated in Figure 1. *PPC* is the first categorization of Infineon outsourced front-end processes. This key, composed of multiple data points, provides the first relevant granularity for studying demand predictability for outsourced processes.

Python was used to create a mapping table from the original data sources, distribute time series to create new ones, and compute the selected metric. *PPC* were then sorted by their mean yearly coefficient of demand. The group's mean and medians were considered possible separators between what would be regarded as predictable or unpredictable processes. Two evaluations were performed, one with the entire datasets and another with outliers removed. In this context, outliers are mostly caused by exceptional and rare changes in demand that infer misleading conclusions from the overall dataset evaluation.

The results were presented to Infineon supply chain experts, who could reflect on the relevance of the calculated metric and the resulting demand predictability status. Based on this feedback, the metric obtained validation.

3.2 Motivation

The coefficient of variation is a viable metric to evaluate demand predictability and is a significant factor in determining an adapted forecasting method (Abolghasemi, Gerlach, Tarr, and Beh (2019)).

We want to use Semantic Web technologies to integrate the relevant data and perform a demand predictability evaluation. Domain models in the Semantic Web are both understandable for humans and machines. Humans can visualize concepts and related data, document precisely where the inherited knowledge comes from, and share a common terminology. We think that a semantic framework based on open standards offers new and extensible possibilities for communication in semiconductor-related supply chains. The predictability of key indicators is a subject of interest for all.

To our knowledge, no previous work was conducted to distribute time series and calculate the coefficient of variations of these distributions using SPARQL and knowledge graphs. By doing so, we can propose

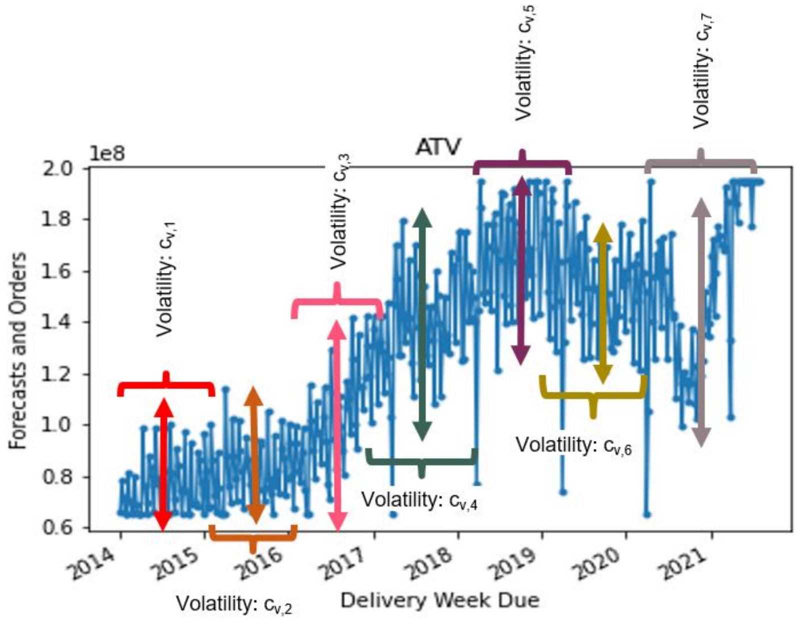

Figure 1: Yearly coefficients of variation to quantify average volatility of forecasts for a defined scope, based on previous work at Infineon.

not only a solution to our use case but also a reusable approach for predictability evaluations of supply chain indicators. Having a linked framework could allow us to explore data in unprecedented ways. We can integrate other relevant data, intern or extern, to combine approaches and envision a close feedback loop between methodologies and results. The Digital Reference developed by Infineon, including domains related concepts and relations between constituents of the semiconductor supply chain, could be used to enhance the quality of the predictability evaluation. Furthermore, the predictability evaluations could also support new use cases in the framework (Ehm, Ramzy, Moder, Summerer, Fetz, and Neau 2019).

4 IMPLEMENTATION

We use an OBDA (Ontology-Based Data Access) approach to implement the semantic data integration. The schema is given in terms of an ontology representing the formal and conceptual view of the domain (De Giacomo, Lembo, Lenzerini, Poggi, and Rosati 2018). Data sources are mapped and contextualized within a shared domain framework in the form of an ontology.

4.1 Ontology Modeling

The ontology definition provides good insights into the relations between concepts within the domain.

Domain and scope definition We list and define all the terms and data necessary to describe the domain accurately. We rely on competency questions. The knowledge graph (data set) derived from the ontology should allow answering the associated competency questions.

We define the scope of the work with the following competency questions (CQ).

- **CQ01**: Which *PPC* are predictable or unpredictable regarding the forecasted demand of their associated *Sales Products*?
- **CQ02**: Which *Process Groups* are predictable or unpredictable regarding the forecasted demand of their associated *Sales Products*?

- **CQ03**: Which *PPC* are predictable or unpredictable, regarding the forecasted demand of their associated *Sales Products*, for a specified time scope?

Data Sources We also define the underlying related concepts for each data point so that it can be easily mapped with the ontology in further steps. We selected the same original data source as previous work at Infineon did to implement the Semantic Web use case. The first source is called *Order_data* and is an excel file containing values of forecasted demand for all *Sales Products*. The second data source is *PG_SP_Table*, which maps *Sales Products* with their associated *Process Groups*. The last data source is *PPC_PG_Table*, a mapping table establishing which *Process Group* requires using one or multiple *PPC*. We replaced the original confidential data with mocked data as the platform used to create and query the knowledge graph is shared with other stakeholders. The created datasets reproduce the original data sources, with three exceptions. Identifiers of products and processes are not associated with real objects but are incremented and used to reproduce the same properties as original data sources. The population is relatively small, as this work does not cover query execution time and optimization.

Ontology Creation Then, we design a semantic framework, an ontology, to encode all necessary information using RDF triples. We model ontologies using the free, open-source ontology editor Protg (Musen 2015). Ontology editors provide an intuitive interface to define classes, relationships, axioms, constraints, and individuals of ontologies. Other functionalities and modules, such as reasoning and SPARQL query engines, are often provided. We further define the ontology by adding constraints to the classes and relations using the ontology editor. We also define the domains and ranges of the properties of the ontology. Object properties usually link instances of a specific class (domain) to instances of another (range). Specifying the expected domain and ranges of the properties allows an enhanced description of the domain and avoids errors when mapping the data sources to the created ontology. We constraint the associated datatypes using the W3C XML Schema Definition Language (XSD). Data mapped to the ontology must respect the schema components specified as the range in the ontology model. We use the Protg plugin ProtgVOWL (Lohmann, Negru, and Bold 2014) to generate a user-oriented visualization of the produced ontology. We then present the view to domain experts and the Semantic Web team at Infineon. We modify the ontology until validation is obtained from all parties. The result of the ontology modeling is a list of RDF triples containing all the formal and conceptual information necessary to populate and query the data sources to answer the previously defined competency questions.

The conceived ontology, displayed in Figure 2, can be divided into two groups.

The first group contains the triples necessary to integrate the data required to evaluate demand predictability for supply chain processes. The relations between the *Sales Products* and the associated supply chain processes are described to allow aggregating the underlying demand for these processes. The class *Process_Group* refers to aggregating similar process lines related to the same set of design rules. Each sales product is produced by one or multiple *Process Groups*. This relation is described by the object property *is_produced_by*. In addition, *Process Groups* can include one or multiple *Procurement Process Categories*, which are aggregations of outsourced processes. We also link *Sales Products* with their *Divisions*, representing the several big markets associated with a finished or unfinished product. *Sales Products* are uniquely identified by their *SP_SAP_Matnr*. They are linked to their associated forecasts. Each forecast is produced for a specific week in which an order is to be delivered, called *Delivery Week Due*. The *Diff Load Due Week* refers to the difference between the delivery *Delivery Week Due* and the week where the forecast was assessed. It indicates how far in advance a prediction is made.

The data value evaluated is called *forecast_and_orders* and quantifies the forecasted demand and confirmed products for a specific *Delivery Week Due* and a specific *Diff Load Due Week*.

The second group contains the triples necessary to save results from the SPARQL queries. Intermediate results and inherited information from the final evaluation are saved into the knowledge graph. Three classes are added: *Forecast_Aggregation* is used to save the calculated underlying demand; *Statistics* is used to save intermediate results and metrics; *Demand_Predictability_Evaluation* is used to save the final results of the evaluation.

Each instance is uniquely identified by a URI containing a specific fragment. For example, the URI of a sales product contains an identifier called *SAP_Mat_Number* as a fragment.

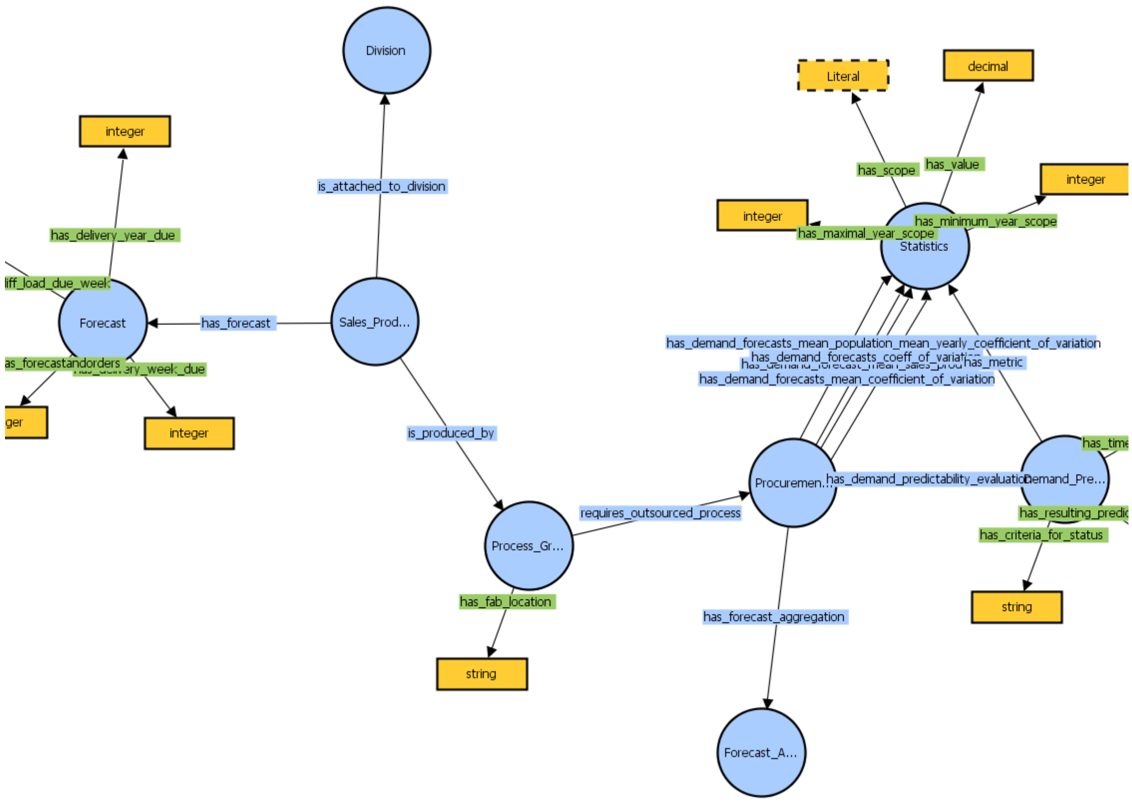

Figure 2: Ontology view with VOWL.

4.2 Output Knowledge Graph

To populate the ontology with its associated data sources in a new knowledge graph, we use the Eccenca Corporate Memory Tool from Eccenca GmbH.

Data sources are linked and transformed into RDF triples using transformation tasks and vocabulary from the previously created semantic framework.

Transformation tasks take an input dataset and execute a set of mapping and transformation rules to generate data for an output dataset. To perform the mapping, we connect the first mapped data (root mapping) to other data by adding values and object mapping.

We transform data by expressing instances of classes into URI (unique resource identifiers) using the URI pattern. The pattern is the standard part of the URI. A new class instance is added by generating a new URI composed of the dedicated pattern and a fragment respective to the instance. Custom transformations allow for the transformation of the integrated data using special operations. For example, we use the four characters of the *delivery_week_due* to inherit the *delivery_year_due*. The resulting linked data is integrated into a new knowledge graph.

4.3 Algorithm

The created knowledge graphs can then be queried within the Eccenca Corporate Memory tool or be downloaded and used in other tools. We use the current standard language for querying RDF data, SPARQL 1.1. [21]. https://www.w3.org/TR/sparql11-query/ We query and update the resulting knowledge graph within the Corporate Memory Tool for this work. The metric and the evaluation are calculated using a chain of INSERT queries with which each intermediate result is saved to the knowledge graph.

- **Query 01**: We retrieve all forecast instances and their properties for all *Sales Products* produced by a (*Process Group* that requires an outsourced process (*PPC*). This is achieved using a WHERE clause containing the pattern needed to be matched by queried data. We aggregate the sum of the forecasted demand by *PPC* and *Delivery Week Due*. This creates new time series to calculate the underlying forecasted weekly demand for each *PPC*. We then save this new aggregated time series in the form of triples in the knowledge graph.
- **Query 02** The second query goal is to calculate and save the yearly mean of the new time series inserted in query 01. We retrieve the results from query 01 and perform an average aggregation, grouping by year and *PPC*. We then save each yearly mean as an instance of the Statistics class and add properties to link the results to the linked data. By doing so, we save for which processes and which year the mean has been calculated.
- **Query 03** The third query uses the results of the first two queries to calculate the yearly coefficient of *forecast_and_orders* per *PPC*. The select and where clause of the query are displayed in Figure 3. We will use this figure as a sample to illustrate how we use fundamental SPARQL techniques to construct an algorithm capable of answering competencies questions.
  The variable *?yearly_mean_Forecast_PPC* is used to retrieve values of the calculated means in query 02 (a). ?sum is used to retrieve the values of the *forecast_and_orders* aggregations per *delivery_week_due*, per *diff_load_due* per *PPC* (b). Because the same variable is used to match solutions for the scope of calculated means and the *delivery_year_due* of the associated *forecast_aggregations*, both retrieved data are bound. This allows associating each forecast aggregation correctly with the appropriate yearly mean. A forecast produced for a product to be delivered in 2014 will be associated with the mean of the forecasts aggregations per *PPC* for all *delivery_week_due* in 2014. In the bind function, the difference between each aggregated forecast and their associated mean is calculated, which results in finding the associated squared mean deviations. The results are saved in the new variable *?squaremeandeviation*. (c)
  The new variable is aggregated in the select clause using the *AVG()* aggregator and a *GROUP BY* (e) to calculate the standard deviations per *PPC* per *delivery_year_due*. An external library is used to calculate the square root of each resulting value. The prefix name math refers to the base URI of the functions library ¡http://www.ontotext.com/sparql/functions/¿. The desired function is selected by adding the path *sqrt*. The results of previous operations are given as input of the *sqrt* function. The output is the calculated yearly standard deviations. Finally, the resulting values are divided by their associated yearly mean to calculate the yearly coefficient of variation.
  (d) A *BIND* clause is used to create a new URI for the newly created object automatically.
  These results quantify how the underlying demand for each PPC fluctuates within a year.
- **Query 04**: The average of the yearly mean of coefficient of variation is calculated per *PPC*. This quantifies how much the underlying demand for each *PPC* fluctuates within a year but on average for a defined scope of years. It corresponds to the metric considered for the first approach of demand predictability evaluation for supply chain processes. The scope of the query can be changed to generate multiple evaluations under different time scopes.
- **Query 05**: Now that the predictability metric to quantify demand predictability for *PPC* has been computed, we calculate in this query the mean of each coefficient as the average of the results for all *PPC* per scope. This value will be used in the demand predictability evaluation. It will be compared to the calculated metric values and used as a separator between what are considered predictable and unpredictable supply chain processes.
- **Query 06**: Query 06 retrieves inserted results from queries 04 and 05 to perform the final evaluation and attribute a predictability status to supply chain processes.

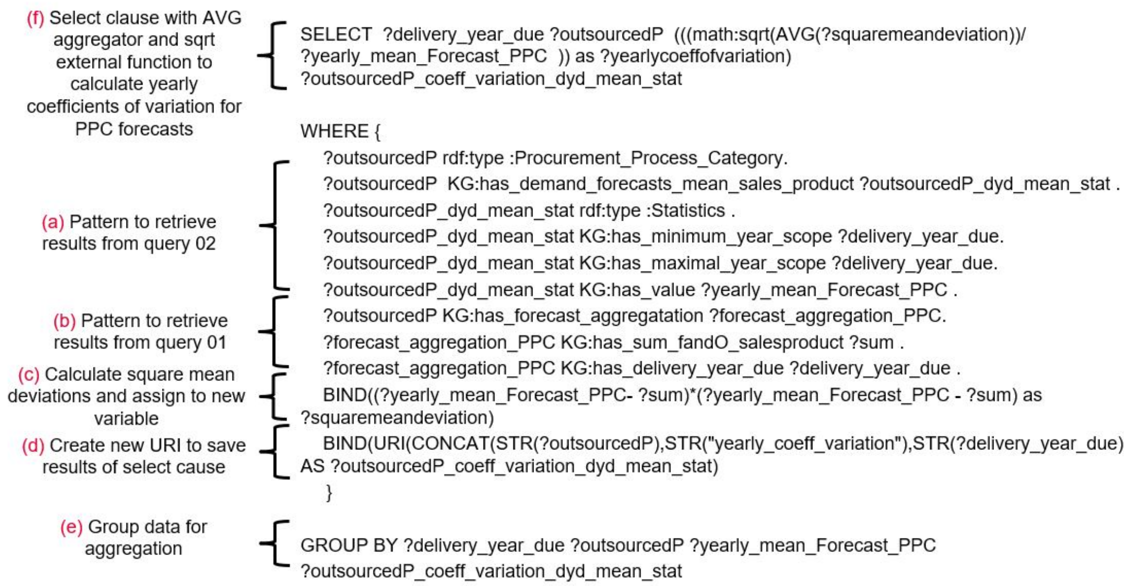

Figure 3: SELECT and WHERE clause from query 03.

The calculated demand predictability values (query 04) are compared to their associated separator (query 05). If the average of the yearly mean of coefficient of variation for a *PPC* is inferior to the average of the group for the same defined scope, the supply chain process underlying demand is evaluated as predictable. If not, it is evaluated as unpredictable. We then save the predictability status in the knowledge graph. We also save the metric to quantify the predictability, the criteria to give a demand predictability status, and the associated separator.

5 EVALUATION

In this section, we evaluate the results of our work and discuss the implications.

5.1 Use Case Evaluation

The results are the answers provided for all competency questions of the ontology. First, we evaluate the accuracy of the computed values during the SPARQL queries. SPARQL queries should accurately compute the expected values necessary to evaluate demand predictability of supply chain processes, following the pre-defined metric. We chose as an evaluation methodology unit testing. We verify the proper functioning of code, proceeding by step. Each step is assessed independently, allowing for faster determination of the source of possible errors. We compared results from both approaches and validated the computed values of each SPARQL query in the sequential order of the overall evaluation.

Figure 4 displays the results for the competency questions 01 and 03. The resulting knowledge graph provides the first evaluation in the form of a predictability status associated with each evaluated process. The metric is the mean of the yearly coefficient of variation of forecasts, and the average of the group is used as a separator between processes with demand evaluated as predictable or not. The methodology to quantify demand predictability, the time scope of the analysis, and the criteria to choose a predictability status are also saved in the knowledge graph.

The coefficient of variation provides the first metric to compare the variability of the distributed demand. For competency question one, predictability is calculated for a full-time scope (2012-2015). Demand for *PPC* 05, 06, and 11 are evaluated as predictable as the mean yearly coefficient of variation is inferior to the average of the evaluated group. Demand for all other *PPC* is evaluated as unpredictable.

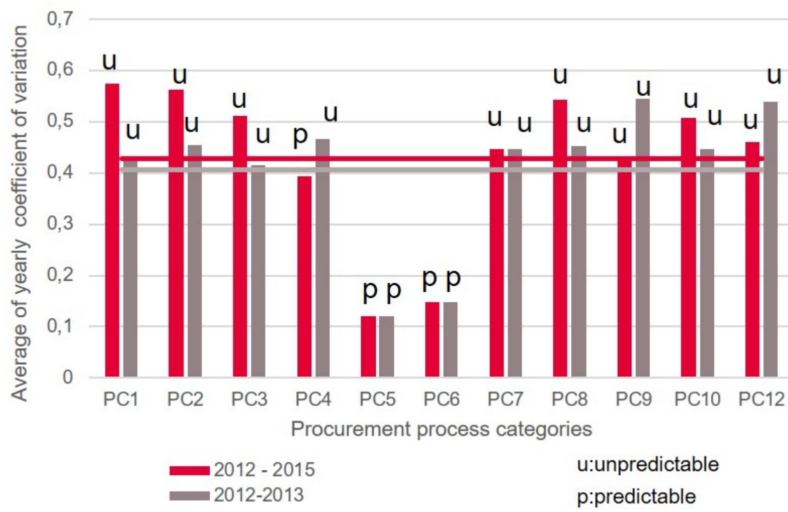

Figure 4: Demand predictability evaluation for PPC; results of competency questions 01 and 03.

High volatility can be induced in an early phase, where the associated sales product already has an unpredictable demand. The uncertainty might also appear when the demand for sales products is redistributed to the underlying supply chain processes.

The aggregations of high volatility forecast could lead to a predictable procurement process category, as trends do not necessarily emerge on the sales product level. A *PPC* can be, for example, specifically associated with a specific technology. The aggregate demand for this *PPC* could be predictable because of the underlying technological trend.

Competency question 3 illustrates the possibility of restraining the scope to a specific period. This functionality compares how the predictability status evolves depending on the time scope adopted. This is illustrated by *PPC* 4, in which demand predictability status is evaluated as predictable on a full-time range but unpredictable in 2012-2013.

The inherited knowledge can always be interpreted in multiple ways. However, the convenience of using Semantic Web technologies is that all steps necessary to generate the knowledge can be well defined within a common framework. A human or machine can consider all underlying information about such knowledge to support a decision. Depending on the relevant knowledge, a specific use case might require evaluating a particular time scope or multiple times scope.

Apart from evaluating the supply chain process, the results are similar for competency question 02. This illustrates how the semantic framework can aggregate the demand with multiple views.

5.2 Discussion and Implications

The created semantic framework, integrated data silos, and derived knowledge graph have been successfully used to perform the necessary aggregations and operations. As such, Semantic Web technologies have been successfully implemented to calculate the desired measure for the classification of processes based on the predictability of their underlying demand. Consequently, we assess that predictability of underlying demand can be quantified for resources, for which relations to sales products can be accurately described within the semantic framework. Predictability and volatility are relevant concepts for many indicators in supply chain management (Castilho, Lang, Peterson, and Volovoi 2015).

One can use this approach not only for demand but for indicators such as resources, costs, prices, and lead time. Semantics allow a flexible framework to simplify the generalization of the approach. Further development would use the same query structures to perform predicability evaluation on multiple indicators and supply chain levels.

Furthermore, the Semantic Web gives us key possibilities to build contextualization on a data level. This means better data sources, inferred knowledge, and data transformation visibility. One can see how and from which data information is constructed. One can perform multiple demand predictability evaluations using different metrics or criteria for evaluation and give a high-level status that considers all conducted evaluations with the same scope. One can also incrementally add new concepts in the ontology to improve the quantity and quality of inserted knowledge. Relevant external data can be described in the framework and accessed by SPARQL end-points. Semantic Web reasoning abilities to verify and discover facts could be used to understand volatility and investigate new forecasting models that would make unpredictable indicators predictable.

The evaluation of the supply chain processes could be greatly improved. First, the coefficient of variation is a sufficient metric for this use case, but other metrics and methodologies might be necessary for other works. Furthermore, the demand predictability status presented in Figure 4 can be misleading as we split *PPC* into only two categories: predictable or unpredictable. One *PPC* could have a metric value slightly below the average of the group, another slightly up. The result would be two different predictability statuses. This demonstrates why contextualization of the output of an evaluation is so important. In the end-view, integration of medians should be considered, as it allows separation into quantiles, dividing the population with the desired granularity. For example, quartiles could be determined to classify processes with four different levels of demand predictability. The first level would be the more predictable, and the fourth level would be less predictable. SPARQL provides the possibility of ordering the solution sequence of a query but does not provide an efficient way to save the rank of a solution in the sequence. This is needed to find medians, as their rank defines them within an ordered set of values. Finding medians would also allow removing outliers, as quartiles are used in the formula used in the pre-defined to identify outliers in a population.

6 CONCLUSION

In this work, we correctly computed a metric to evaluate the demand predictability of supply chain processes using an ontology-driven approach and SPARQL queries. Furthermore, we use this metric to perform a first predictability evaluation which can be further developed with the extension of the framework.

One can use this approach for multiple indicators across end-to-end supply chains. All stakeholders can interact using a common framework to enhance cross-organization cooperation. This work was performed using a small data population and did not investigate topics such as query optimization. As a next step, we propose studying the scalability of our solution. Other solutions are possible while keeping an OBDA approach. Semantic Web Technologies can integrate and describe the results of metrics calculations from other tools like Python. Other possible steps are to improve the evaluation by using medians and adding other relevant concepts and data to the framework.

REFERENCES

Abolghasemi, Mahdi and Gerlach, Richard and Tarr, Garth and Beh, Eric 2019. "Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion". https://arxiv.org/abs/1909.13084.

Abowd, G. D., A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles. 1999. "Towards a Better Understanding of Context and Context-Awareness". In *Handheld and Ubiquitous Computing*, edited by H.-W. Gellersen, 304–307. Berlin, Heidelberg: Springer Berlin Heidelberg.

Ameri, F., and L. Patil. 2012. "Digital manufacturing market: a semantic web-based framework for agile supply chain deployment". *Journal of Intelligent Manufacturing* 23(5):1817–1832.

Bishr, Y. 1998. "Overcoming the semantic and other barriers to GIS interoperability". *International journal of geographical information science* 12(4):299–314.

Castilho, J. A., T. E. Lang, D. K. Peterson, and V. Volovoi. 2015. "Quantifying Variability Impacts upon Supply Chain Performance". In *Proceedings of the Winter Simulation Conference*, edited by L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 1892–1903. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chaudhuri, S., and U. Dayal. 1997. "An overview of data warehousing and OLAP technology". *ACM Sigmod record* 26(1):65–74.

Cyganiak, Richard and Reynolds, Dave and Tennison, Jeni 2014. "The RDF data cube vocabulary". https://www.w3.org/TR/vocab-data-cube/.

De Giacomo, G., D. Lembo, M. Lenzerini, A. Poggi, and R. Rosati. 2018. "Using Ontologies for Semantic Data Integration". In *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*, edited by S. Flesca, S. Greco, E. Masciari, and D. Saccà, 187–202. Cham: Springer International Publishing.

Ehm, H., N. Ramzy, P. Moder, C. Summerer, S. Fetz, and C. Neau. 2019. "ECSEL Digital Reference: A Semantic Web for Semiconductor Manufacturing and Supply Chains Containing Semiconductors". In *Proceedings of the Winter Simulation Conference*, edited by N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 2409–2418. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Huang, M.-G., P.-L. Chang, and Y.-C. Chou. 2008. "Demand forecasting and smoothing capacity planning for products with high random demand volatility". *International Journal of Production Research* 46(12):3223–3239.

Landolfi, G., A. Bami, G. Izzo, E. Montini, A. Bettoni, M. Vujasinovic, A. Gugliotta, A. L. Soares, and H. Diogo Silva. 2018. "An Ontology Based Semantic Data Model Supporting A Maas Digital Platform". In *Proceedings of the 2018 International Conference on Intelligent Systems (IS)*. September 25$^{\text{th}}$-27$^{\text{th}}$, Funchal - Madeira, Portugal, 896-904.

Lohmann, S., S. Negru, and D. Bold. 2014. "The ProtégéVOWL Plugin: Ontology Visualization for Everyone". In *Proceedings of ESWC 2014 Satellite Events*, edited by V. Presutti, E. Blomqvist, R. Troncy, H. Sack, I. Papadakis, and A. Tordai, Volume 8798 of *LNCS*, 395–400. Cham: Springer International Publishing.

Loucopoulos, P., and V. Kavakli. 1999. "Enterprise Knowledge Management and Conceptual Modelling". In *Conceptual Modeling: Current Issues and Future Directions*, edited by G. Goos, J. Hartmanis, J. van Leeuwen, P. P. Chen, J. Akoka, H. Kangassalu, and B. Thalheim, 123–143. Berlin, Heidelberg: Springer Berlin Heidelberg.

Musen, M. A. 2015, jun. "The Protégé Project: A Look Back and a Look Forward". *AI Matters* 1(4):412.

Packowski, J. 2013. *LEAN supply chain planning: the new supply chain management paradigm for process industries to master today's VUCA World*. CRC Press.

Pal, K. 2017. "A semantic web service architecture for supply chain management". *Procedia Computer Science* 109:999–1004.

Quamar, A., F. Özcan, D. Miller, R. J. Moore, R. Niehus, and J. Kreulen. 2020, aug. "Conversational BI: An Ontology-Driven Conversation System for Business Intelligence Applications". *Proc. VLDB Endow.* 13(12):33693381.

Yavari, A., P. P. Jayaraman, D. Georgakopoulos, and S. Nepale. 2017. "ConTaaS: An Approach to Internet-Scale Contextualisation for Developing Efficient Internet of Things Applications". In *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS)*. January 4$^{\text{th}}$-7$^{\text{th}}$, Hilton Waikoloa Village, Hawaii, USA, 1-9.

Ziegler, P., and K. R. Dittrich. 2007. "Data Integration — Problems, Approaches, and Perspectives". In *Conceptual Modelling in Information Systems Engineering*, edited by J. Krogstie, A. L. Opdahl, and S. Brinkkemper, 39–58. Berlin, Heidelberg: Springer Berlin Heidelberg.

AUTHOR BIOGRAPHIES

NOUR RAMZY is a Master of Science in *Information Technology* from the University of Stuttgart. She wrote her Master Thesis in collaboration with Infineon Technologies in the Supply Chain Innovation department on the topic Using Semantic Web technologies to improve Infineons Open Online Sales and Marketing Platform. Now, she works in the same department at Infineon Technologies as a Ph.D. student from the Leibniz University Hannover with the research focus on semantic data integration for supply chain management Her email is address is Nour.Ramzy@infineon.com.

PHILIPP ULRICH is a PhD candidate at Infineon Technologies AG working in the field of Semantic Web and Artificial Intelligence. He has a Master of Science in information systems from Technical University of Munich. His interests are end-to-end supply chains, transparency in semiconductor supply chains, Semantic Web and AI. His email address is philipp.ulrich@infineon.com.

LANCELOT MAIRESSE is a Bachelor of science in *Production and Automatisation* from the Munich University of Applied Sciences and pursue now a Master in the same field. He is currently working at Infineon in the Supply Chain department and wrote his Bachelor Thesis on the subject *Demand Predictability for Supply Chain Processes*. His email address is mairesse.lancelot@infineon.com.

HANS EHM is the Lead Principal Supply Chain heading the Corporate Supply Chain Engineering Innovation department at Infineon Technologies. He has a Master of Science in *Mechanical Engineering* from Oregon State University. His interest is on Supply Chain Innovation with focus on capacity and demand planning, deep learning, artificial intelligence, simulation, and block chain applications for traceability. His email address is hans.ehm@infineon.com.