Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 220 (2023) 446–455

Procedia Computer Science

www.elsevier.com/locate/procedia

The 14th International Conference on Ambient Systems, Networks and Technologies (ANT)
March 15-17, 2023, Leuven, Belgium

Business rule extraction using decision tree machine learning techniques: A case study into smart returnable transport items

Rob Bemthuis*, Wei Wang, Maria-Eugenia Iacob, Paul Havinga

University of Twente, Drienerlolaan 5, 7522 NB Enschede, the Netherlands

Abstract

Decision support systems are becoming increasingly sophisticated (e.g., being machine learning-based), attempting to automate decisions as much as possible. However, it remains challenging to extract meaningful value from large quantities of data while also maintaining transparency in seeking justification for the choices made. Instead of creating methods for increasing the interpretability of black box models, one way forward is to design models that are inherently interpretable in the first place. Rule-based methods can automate decisions with great transparency and accuracy, helping to ensure compliance with regulations and adherence to organizational guidelines. In this paper, we propose an approach that uses a decision tree machine learning classification technique for extracting business rules from IoT-generated data to predict the asset status of Smart Returnable Transport Items (SRTIs). We report on an industrial case study that uses two years of historical data, obtained from an SRTI provider in the Netherlands, to predict the status of smart pallets. We compare the performance with the results obtained by using a support-vector machine (SVM) technique. Our experiments show that our solution is both accurate and flexible in terms of business rule elicitation. The obtained decision trees are human-interpretable, can easily be combined with other decision-making techniques, and provide a prediction accuracy marginally higher than an SVM technique.

Keywords: business rules; machine learning; decision tree; logistics; smart pallets; case study

1. Introduction

Recent advancements in sensor networks, cyber-physical systems, and the ubiquity of Internet-of-Things (IoT) have significantly increased the collection of data [24]. Today's information systems continuously monitor the physical environment and produce large quantities of data that are a great source for deriving information to support decision-making [2, 41]. However, although context-aware devices and wireless communication provide advanced services [29], there are still many types of unpredictable disruptions that affect our daily lives [9, 43] and business activities

* Corresponding author. Tel.: +31 534897009.
E-mail address: r.h.bemthuis@utwente.nl

10.1016/j.procs.2023.03.057

(such as those in supply chains) that cannot be avoided. One of the causes is that extracting meaningful information from data is a complex and challenging task, requiring innovative techniques and algorithms to analyze and understand such data [46]. Yet, although the sophistication of data-driven artificial intelligence (AI) approaches has recently increased to such an extent that human intervention is minimized [6], we also need to understand what the underlying models are doing, instead of treating computational machines as merely black box models [16, 38]. Organizations have increased their need for understanding how decisions supplied by data-driven AI systems are made, especially when these may affect human lives [23, 39] or when there is little margin for errors due to the criticality of the decisions. However, this remains challenging, as the underlying rules of AI models (e.g., typical neural networks) are not always transparent and human-interpretable [13], making it difficult to determine when the AI model can be trusted [45].

Developing methods for interpreting these black box models and alleviating some of these problems must be beneficial, but is likely to perpetuate bad practice and may cause great harm to society [38]. As argued further by Rudin [38], instead of developing such methods, one way forward is to design models that are inherently interpretable in the first place. To this end, rule-based methods can automate decision-making with great transparency and accuracy, while also ensuring adherence to organizational rules [1, 21, 27, 31, 33]. Such rules encapsulate unambiguous statements that capture (certain parts of) organizational/business policies [44]. An example of such a rule is “if the weather temperature is too high, then inform the customer and reschedule following an alternative route”.

More broadly, business rules are an expression of business logic, which involves determining a course of action [47]. In this research, we are primarily concerned with the business logic close to the point at which the actual data is being collected. Business logic related to the physical proximity of the location where such actions occur has the advantage that business objects involved in such actions can anticipate quickly and take preventive measures to ensure the resilience of the system [9]. Business objects, such as smart pallets, can monitor, detect, and respond to disruptive behaviors close to the locations where these occur [8]. For example, an intelligent pallet can signal on the spot through sounds, colors, or vibrations, and report disruptions to customers. In this work, we mainly consider declarative business logic expressed in business rules, meaning that for such rules the primary goal is to specify the condition on which a decision depends and how it should be expressed, and not *per se* when this decision is to be made (e.g., the sequence).

Although modern information systems are governed by ubiquitous rules that encode the policies and procedures of operations [4, 42], there are still key challenges related to the application of business rules close to where the “action” happens. First, creating, maintaining, and deploying a coherent large number of rules can be daunting and inefficient [20]. Traditionally, such rules are obtained through expert discussions and examining relevant documentation [42], which can be time-consuming and yield outdated results. Many rules are not formally documented, as tacit knowledge still plays a prominent role in many organizations, and can thus be “hidden” business logic. Moreover, even when rules are properly stored and managed in a business rule management system [36], they may not always be followed in practice due to e.g., technological advances, lack of proper IT integration, inconsistencies, and disruptive and unexpected behaviors. The presence of workarounds and legacy systems may play important roles. Thus, without rigor, it is easy to fail to apply and extract rules, especially ones that are rarely invoked. Second, business rules fulfill the need to realize specific business outcomes and are not an inherent goal. Achieving an optimal configuration of business rules to achieve a certain outcome can be difficult or even impossible (due to unpredictable emergent behaviors) [10, 11].

Drawing upon these two strands of research, this study attempts to utilize IoT data and performance evaluations governed by existing information systems to extract (indicators of) business rules *a posteriori*. In this paper, we propose a machine learning approach for extracting business rules from IoT-generated data, using two years of historical data to predict the asset status of Smart Returnable Transport Items (SRTIs). As a machine learning model, we use decision tree algorithms, which are suitable for producing business rules. Such extracted business rules can assist SRTI providers and users in their decision-making processes. Following the CRoss Industry Standard Process for Data Mining (CRISP-DM) methodology [49], we discuss a real-life logistics case study of an SRTI provider in the Netherlands that utilizes condition-based monitoring data generated by smart pallets. Additionally, we compare the extracted decision tree models against results obtained using a support-vector machine (SVM) technique to address the suitability of the obtained models. Our approach can be used in other domains to extract business rules from IoT data, instead of traditional approaches that define business rules *a priori* and then evaluate the performance of those rules based on the data.

The remainder of this paper is organized as follows. Section 2 discusses related work on extracting business rules and using predictive features for SRTIs. Section 3 presents the case study, guided by CRISP-DM. Next, Section 4 discusses the key findings. Finally, Section 5 draws conclusions and suggests directions for future research.

2. Related work

Although many data mining techniques have been proposed to extract business rules, there remain several gaps in the existing research [30]. One challenge is likely to be the many factors that influence the decision about whether a business rule should be explicitly modeled or not [48]. Nevertheless, due to the ever-increasing data availability and computational power, novel machine learning techniques can create added value for extracting business rules [20, 30]. However, few studies have yet explored business rule extraction approaches using machine learning techniques. Techniques such as neural networks [3] and SVM [32] provide accurate classifications, but the black box nature of these methods prevents users from understanding the precise contribution of specific business rules, thus limiting their applicability. Obtaining interpretable business rules from such approaches tends to be both complex and difficult [7].

Existing methods also use decision trees for extracting business rules, but typically yield a classification accuracy worse than, e.g., neural networks and SVM techniques. However, this does allow the easy creation of business rules [7, 37]. Bazhenova and Weske [7] proposed a four-step approach to derive decisions from process models by using decision trees. Rozinat and van der Aalst [37] present an algorithm to detect data attributes that affect the routing of a case where there appears to be a choice of route. Both works [7, 37] show that decision rule extraction can benefit from the inclusion of knowledge of a process model. However, those works assume the availability of an existing process model and decision points, and knowledge of the data dependencies between (predefined) decision points.

Furthermore, several articles address the extraction of business logic through process mining techniques (e.g., [12, 14, 18, 37]). Although the process mining paradigm has enabled a shift in the fact-driven discovery of processes, several challenges arise from the use of such event log-based techniques. First, process mining often misses significant details in a process because event logs, if available at all, do not always provide sufficient visibility of all possible actions. Second, a process mining project often requires existing knowledge of a process, which is counterintuitive in the sense that getting value from a technique that is supposed to identify discovered processes relies on existing process knowledge. Additionally, research focuses mainly on the control-flow perspective, while attributes of decisions that drive the control-flow are often overlooked [19]. According to De Smedt et al. [19] it is crucial to understand how data attributes evolve and contribute to decision-making. Khemiri et al. [28] also call for more research on the broader data mining topic of discovering relevant attributes in relation to business rules.

Smart business objects in the logistics domain provide a wealth of possible predictive data features that can be scrutinized to help develop business rules. However, although many smart devices provide real-time monitoring of business objects in terms of position or state (e.g., temperature, weight, humidity, shock, etc.) [15, 22, 26, 35], limited research has been done on extending the concept of physical monitoring with ambient intelligence and distributed and autonomous control [29]. Considering this identified literature gap and the need to capture interpretable business rules from large amounts of data, especially by using machine learning, we contribute by presenting a machine learning approach that uses decision trees to discover business rules of SRTIs. Furthermore, as little is known about the specific features influencing the SRTIs, we aim to advance the state of the art in the field of smart logistics. Our work generalizes the work of Barachi et al. [5], as we validate the benefits and feasibility of SRTIs. More specifically, we show an implementation of business rules within an IT infrastructure for SRTIs, as we build further on the use case proposed by Piest et al. [34]. Finally, we offer practical insights based on a case study of an SRTI provider in the Netherlands.

3. Case study

Smart objects with logic closely related to where the “action” happens have been adopted by many organizations to improve the effectiveness of business processes. To better understand the importance of logistics objects that embed sensors and business logic, a case study concerning the shipment of perishable goods (i.e., products likely to spoil, decay, or become unsafe for consumption or use) by SRTIs is adopted. We structure the case study according to the CRISP-DM methodology, which is the de-facto standard and an industry-independent methodology for applying data mining projects [40]. For this, we follow the guidelines as presented in [17, 49].

3.1. Business understanding

The case study revolves around an SRTI provider in the Netherlands that provides an intelligent returnable asset pooling platform for the logistics market and is introduced in [34]. We investigate the possibility of extending the platform with a machine learning-based approach to predict the SRTI status based on business rules, and thereby help the SRTI provider to enable new services to its users and customers. The aim is to propose a suitable machine learning technique to extract business rules from the data generated by smart pallets. The main stakeholders include SRTI service providers, compliance organizations (e.g., governmental bodies), transport companies and personnel (e.g., truck drivers), IT system developers, and end consumers. The anticipated benefits from implementing this approach as an integral feature of the platform include:

- Enabling new services based on real-time monitoring, detection, analysis, and prediction of emergent behaviors;
- Fewer lost, depreciated, or damaged perishable goods during transportation and storage;
- Improved real-time quality control and multimodal planning in terms of resilience, efficiency, and sustainability;
- Faster anticipation of disruptions, by implementing dynamic planning capabilities at the point of action;
- Increased safety for logistics goods handlers and compliance with regulations;
- Lower SRTI maintenance costs, better SRTI lifecycle estimates, and higher SRTI utilization;
- Increased transparency in interacting with human decision-makers;
- More suitable for real-life, large-scale implementations, while also enabling distributed services that incorporate intelligence (e.g., decision-making capabilities) close to the point of proximity of where the data is collected.

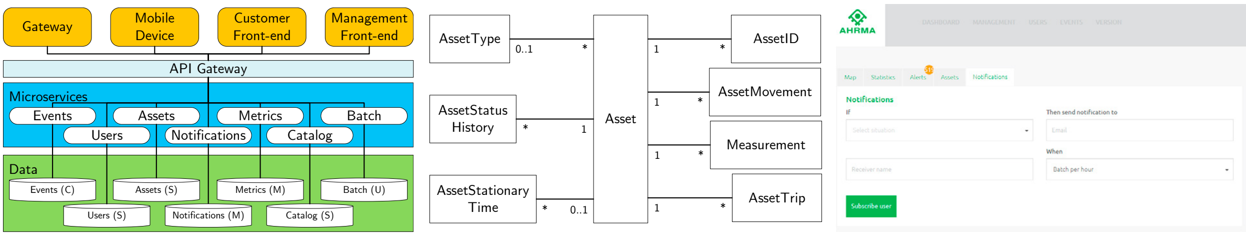
3.2. Data understanding

This subsection provides an overview of the entire SRTI dataset, as well as some general analysis that helps us achieve a brief understanding of the case study. The data provided by the SRTI provider concerns a selected subset within a pool of pallets and spans two years' data (April 2016 - April 2018). The data comprise approximately 300GB, covering more than 6000 assets, from various sources. The dataset consists mainly of readings from various types of sensors installed on each pallet. These readings include temperature from thermal sensors, vibration from shock sensors, and Geo-location coordinates from GPS. Compared to rich-info-based sensors such as cameras or microphones, these sensors provide discrete and scalar readings each time they are triggered. Other data sources include manual recordings, such as MAC address, pallet ID, battery status, and status of the pallet.

Regarding the SRTI status, two types of state are recorded: availability and maintenance. Availability is denoted as: AVAILABLE (i.e., asset is not assigned and can be deployed to a customer), ALLOCATED (i.e., asset is allocated to customer), IN-TRANSIT (i.e., asset has not been seen by a gateway for > 3 hours), and LOST (i.e., asset has not been seen by a gateway for > 30 days). Maintenance describes whether a pallet is DAMAGED, REPAIRED, or NEW. For this, ground truth is recorded through a manual inspection by a customer or maintenance engineer.

Various types of data points, when collected in real-time, are propagated to back-end servers and eventually stored in several databases (see Figure 1a). Various types of data are aggregated and stored in various tables via one or more index entries. For example, temperature, signal, and battery sensor data are stored in the same table, and these data records are sampled per device periodically, so that the index entry is 'device-id + timestamp'. However, the table related to the Geo-location records of devices only stores data records when SRTIs are moved, and therefore has fewer records (with an index entry of 'device-id'). Figure 1b contains a snippet of the assets' relational database structure. More database and data details are not included in this paper for either reasons of space or confidentiality.

Alerts can be generated by means of specified business rules that trigger an action (e.g., a notification). Some alert flags are: 0 = at least one shock event has occurred; 1 = low temperature limit reached; 2 = high temperature limit reached; 3 = the battery capacity is below a threshold; 4 = weight limit reached; 5 = temperature file alert (if the number of the temperature file entries has reached a threshold); and 6 = weight file alert (if the number of the weight file entries has reached a threshold). Each alert flag has parameters that need to be specified. Figure 1c shows how business rules can be configured, and we refer the reader to the work of Piest et al. [34] for an example of how complex business rules can be configured.



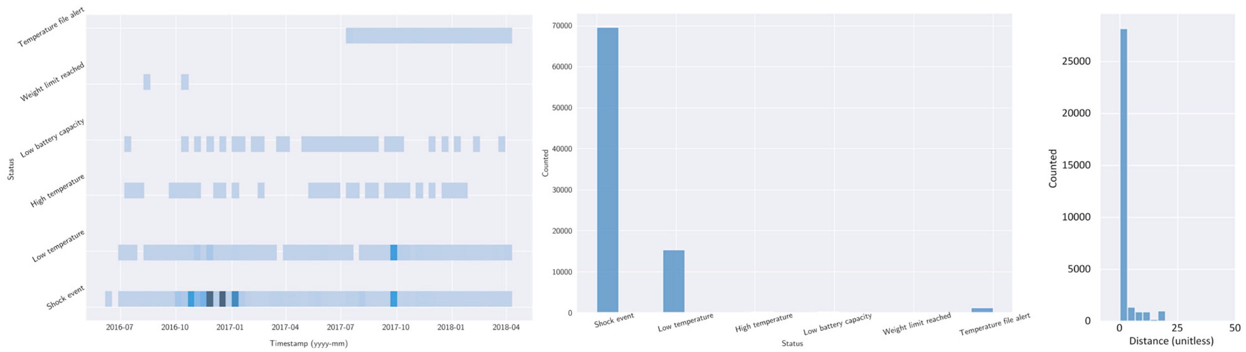
(a) Microservices architecture (C = Cassandra, S = MySQL, M = MongoDB, U = Unknown).

(b) Assets relational database structure.

(c) Business rule configuration form.

Fig. 1: High-level system overview.

After this initial data collection and case synthesis, we explored and verified the data quality of the datasets. Figures 2 and 3 show initial overviews of (aggregated) asset data. Some preliminary insights can be obtained from these overviews. Figures 2a and 2b show that there are many shocks observed and that the number of times the low temperature limit is reached is higher than the high temperature counterpart. Figure 2c shows how often certain movement distances occurred. Remarkably, many registrations took place without any physical movement of the pallet, possibly suggesting that external factors affected the pallet's condition. Figure 3a gives an impression of how often the status of SRTIs changes. In the overview presented in Figure 3b, we observe a densely populated area of around 20 degrees Celsius, which is in line with expectations. Furthermore, clusters across various temperature ranges can be observed, which may reflect the climates in which the smart pallets are being used. Figure 3c shows the physical distance between sensor measurement locations, in which we see a few exceptional behaviors. Please note that although our initial observations have been verified with the SRTI provider, we have not disclosed additional insights due to confidentiality considerations, data limitations, data visualization limitations, or space restrictions.



(a) Asset status over time.

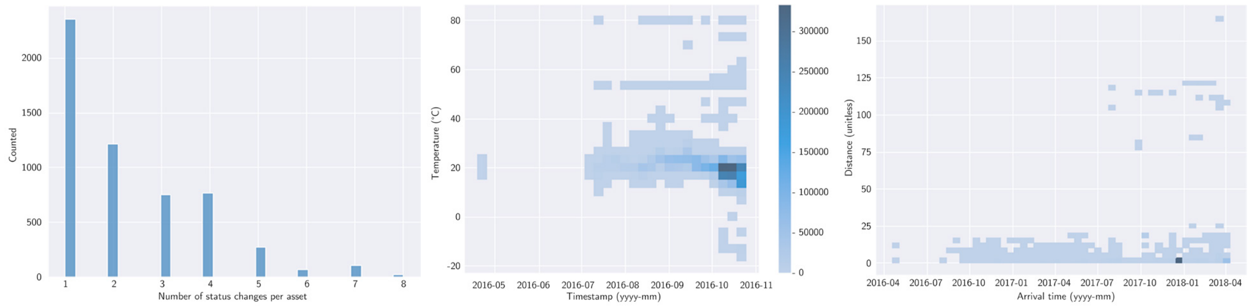
(b) Asset status distribution.

(c) Movement distances.

Fig. 2: Descriptive statistics about the datasets (part 1).

3.3. Data preparation

For data preparation, we loaded, filtered, constructed, and integrated data. Initially, we attempted to use a Distributed File System (DFS), which generally offers good access times and network efficiency for big data, and the ability to exchange data across multiple networks and across multiple platforms. A common approach would be to store the data with the index 'asset-id', but a disadvantage of that is that we cannot guarantee that each entry can be accessed in strict order during the training phase, unless rigid training rules are written. Instead of using DFS that loads the entire dataset, we adjusted the pipeline to first filter out (likely to be irrelevant) data, in order to make the dataset smaller.



(a) Asset status change counts (e.g., > 2000 assets changed their status once).

(b) Temperature over time.

(c) Distance over time.

Fig. 3: Descriptive statistics about the datasets (part 2).

As mentioned earlier, many pallets are not uniformly sensed or operated, are seldom moved, and are only present in the status of AVAILABLE, which is expected to provide little insight. Furthermore, of the total assets, only 282 ($\approx 4.7\%$) had ever been damaged, which may bias our dataset (as there are many "good" assets). The SRTI provider informed us that many data records for pallets that had left the factory cannot be made available to us (for confidentiality reasons) and that many pallets stayed at the factory (e.g., for manufacturing or testing purposes). Therefore, we decided to filter and maintain a selection of representative assets only. Assets that had never been broken or never moved are consequently removed from the data, causing the number of SRTIs considered to shrink to 997. Further research could evaluate whether this might result in biases in the generated models.

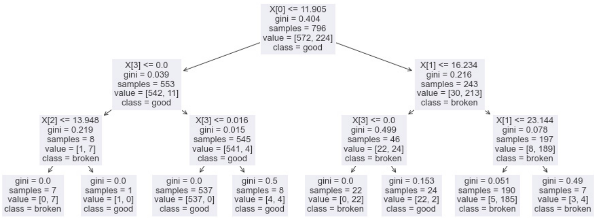
Next, we combined all the data tables into one big and uniform dataset. Data for any other assets (i.e., beyond the 997 SRTIs) are discarded. Out of all the available data, we chose just temperature and location. Some fields are discarded as they can hardly be connected to the status of pallets, such as 'mac-address' and 'vendor'. Some data fields might also be interesting to examine but are not used for practical reasons. For example, shock sensor readings contained many distorted, invalid and unrealistic entries, making their validity questionable. After aligning with the SRTI provider, we decided to proceed with data on the status, temperature, and movements of the individual SRTIs. Further work could investigate the inclusion of more data, such as weight, stationary time, and shock duration.

Apart from field selection, one other important aspect of entry concatenation is the key selection. As we investigate the smart pallets over a long time range, it seems intuitive to select the 'asset-id + timestamp' union as the key. However, the granularity of this 'timestamp' is a major challenge, as data records are recorded at the millisecond level, making the joining of tables difficult. Therefore, we downsampled the timestamp to the day level by merging data points on the same day. We can justify this choice as the data only last two years and day-level observation seems acceptable, because the information about the day time of the events is not expected to influence the judgment about the status of a pallet significantly compared to e.g., the temperature. However, since multiple temperature measurements can be made in one day and we do not want to lose important data, we used three data points for each day's temperature: minimum, maximum, and average. Regarding location, we approximated the traversed distance per day by dividing the total distance by the duration of days. Since we only had the start and end points recorded in a movement data table, we calculated the (unitless) Euclidean distance between the longitude and latitude coordinates. After concatenating and restructuring the data, we have a dataset indexed by the 'asset-id / day' union, in which each entry has three temperature values, distance, and a 'status' label.

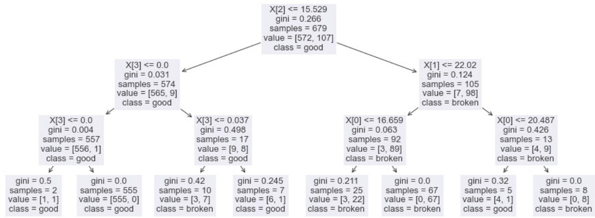
3.4. Modeling

After consulting the main stakeholders, we ensured a good fit between the envisioned results, available input data, and the available techniques, which is important [19]. We used the family of decision tree models, which are one of the most common machine learning algorithms for producing business rules [20]. One benefit of using decision trees is their simplicity and interpretability. Also, it does not require data to be scaled to a uniform range.

As a modeling technique, we treat the data as a time series in which each sample is an 'asset-id' with one label and all the corresponding 'days' are the measurements within each time series. Now we can use the historical data to predict whether a pallet has been damaged. The next step can be to predict if (and possibly when) a pallet will be damaged in the future. As cross-validation, we split the data into training and test sets (with the number of folds being 5) and we implemented the models using the Python Scikit-learn machine learning library. To benchmark our decision tree models, we also implemented an SVM algorithm, which yielded an accuracy of 95%. SVM models are widely used, require only a few parameters, and are known to yield good performance for solving classification problems. We build two decision tree models, both visualized in Figure 4. The first model uses the entire sample set, while for the second model we discarded non-moving pallets.



(a) Decision tree model (accuracy: 0.975)



(b) Decision tree model, with discarded non-moved smart pallets (accuracy: 0.976)

Fig. 4: Decision tree models (X[0] = minimum temperature; X[1] = maximum temperature; X[2] = average temperature; X[3] = distance).

3.5. Evaluation

The prediction accuracy of the decision tree shown in Figures 4a and 4b are 97.5% and 97.6% respectively, which is higher than that obtained by using the SVM technique. Remarkably, both decision tree models use all the features, which may indicate the importance of effective feature selection. The decision trees can easily be translated into business rules. For example, considering Figure 4a we may say: “if the minimum temperature is above 11.905 and the maximum temperature is higher than 16.234, then the pallet is considered to be damaged”. Nevertheless, some simplifications to the models can be made, because some branches do not differentiate between the various classes.

3.6. Deployment

The deployment of the machine learning-based approach to make improvements within the organization is not within the scope of this article. However, we refer the reader to the system architectures of Iacob et al. [25] and Piest et al. [34] to get an idea of how to integrate the approach in an organizational context. The decision trees are not computationally demanding and can be translated to business rules, which is promising for real-time deployment. Nevertheless, further research should evaluate the run-time performance of the proposed approach for generating decision trees, especially when considering large datasets.

4. Discussion

Previous research has shown that decision trees, obtained through machine learning techniques, are effective in extracting business rules from big data. Although smart business objects provide a wealth of data to scrutinize, machine learning has received limited attention in the field of smart logistics. Practitioners typically apply predefined rules as an effective course of action for business objects. However, as well as tending to be time-consuming and repetitive, there is often no performance benefit, as it is beset by validation concerns (e.g., due to exceptions).

Our work contributes to the existing literature by using machine learning to predict the status of a smart pallet. Our decision tree models provide human-interpretable rules based on data and can be used to bootstrap a decision support system or to adapt to changing behaviors found in real data. Our approach is illustrative of a logistics case study that consists of a large dataset, and intended to encourage further thought about extracting business rules from smart objects by applying logic closely related to where the “action” is in order to improve the effectiveness of business processes.

Future work may examine the depth of the decision trees in more detail and compare the obtained results with manually obtained decision logic. Combined with the deployment of new and/or updated business rules (e.g., via some learning mechanism as conceptualized in [10]), our approach may also pave the way for more automated decision-making in which tasks previously performed by humans are taken over by smart business objects, allowing the human user to focus on other aspects of processes that may require more creativity. Thus, this work may be considered as an initial step forward in moving from physical monitoring towards autonomous decision-making, as also called for by Koot et al. [29].

5. Conclusions and future work

This research has proposed a machine learning approach for extracting business rules from IoT data to predict the asset status of SRTIs. In this real-life logistics case study we adhere to the CRISP-DM methodology and use historical data about smart pallets. Our results indicate that decision trees and SVM techniques can predict smart asset status with high accuracy, and that the obtained decision trees are interpretable for humans.

Even though we present our case study in a logistics domain, our approach is generalizable and brings novelty to the business rule field. Our approach can be applied to other domains that deploy machine learning techniques on IoT data with the purpose of extracting the rules, rather than traditional approaches that evaluate emerging business performance based on *a priori* defined rules. The latter generally requires exhaustive, inefficient, and intrusive methods, while not even assuring that all effectively executed rules are captured, and that a certain performance is achieved (due to e.g., emergent and disruptive behaviors). In contrast, inductive reasoning by machine learning models usually requires a large amount of data to discern meaningful patterns but requires less intrusive means, making it particularly suitable for large-scale implementations. Nevertheless, despite its exploratory nature (mainly due to data limitations) this study offers insights into rule-based logic extraction from SRTIs' generated data.

However, this research has led to many questions that call for further investigation. Future work could address the automation of data preparation steps. Further studies may address some of the drawbacks associated with decision trees, including examining how to include more complex data structures, pruning to avoid overfitting, and the handling of outdated data. Furthermore, as typical supply chains involve many stakeholders, data-sharing intentions, and large volumes of data, decision tree calculations can quickly grow cumbersome. Therefore, it would be valuable to explore ways of ensuring a proper balance between computational time, complexity of decision trees, and data governance. Further studies on the role of the human-in-the-loop and on how the decision tree algorithm may incorporate implicit knowledge would also be worthwhile to investigate. We also suggest using additional and more complex datasets, such as open data sources (e.g., weather, road traffic, etc.) and establishing integrations with enterprise systems (e.g., ERP, TMS, WMS, etc.). Likewise, logistics expansion opportunities should be explored, such as adding multimodal transport, and knowledge about the goods transported and shipment providers. Additionally, it would be valuable to dive more deeply into predictive and prescriptive asset maintenance. We suggest exploring the expected SRTI lifetime and predictive asset maintenance plans. Finally, research on compliance checking and handling exceptions is needed, from which deviations, new rule designs (e.g., rule maintenance), and best practices may be derived.

Acknowledgements

This work was supported by the DataRel project, which has received funding from the Dutch Research Council (NWO) and Commit2Data under Grant 629.009.015. This work was also supported by OP Oost, project CountDown. The authors would like to thank all project partners.

References

[1] Abdollahi, B., Nasraoui, O., 2018. Transparency in fair machine learning: The case of explainable recommender systems, in: Human and Machine Learning: Visible, Explainable, Trustworthy and Transparent. Springer, pp. 21–35. doi:10.1007/978-3-319-90403-0_2.

[2] Ahmad, A.F., Sayeed, M.S., Tan, C.P., Tan, K.G., Bari, M.A., Hossain, F., 2021. A review on IoT with big data analytics, in: 2021 9th International Conference on Information and Communication Technology (ICoICT), IEEE. pp. 160–164. doi:10.1109/ICoICT52021.2021.9527503.

[3] Baesens, B., Setiono, R., Mues, C., Vanthienen, J., 2003. Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science 49, 312–329. doi:10.1287/mnsc.49.3.312.12739.

[4] Bajec, M., Krisper, M., 2005. A methodology and tool support for managing business rules in organisations. Information Systems 30, 423–443. doi:10.1016/j.is.2004.05.003.

[5] Barachi, M.E., Salman, S., Mathew, S., 2022. A sensor-embedded smart carton for the real-time monitoring of perishable foods' lifetime, in: 2022 7th International Conference on Smart and Sustainable Technologies (SpliTech), IEEE. pp. 1–8. doi:10.23919/SpliTech55088.2022.9854304.

[6] Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58, 82–115. doi:10.1016/j.inffus.2019.12.012.

[7] Bazhenova, E., Weske, M., 2016. Deriving decision models from process models by enhanced decision mining, in: Business Process Management Workshops, Springer. pp. 444–457. doi:10.1007/978-3-319-42887-1_36.

[8] Bemthuis, R., 2019. Business logic for resilient supply chain logistics, in: 2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW), IEEE. pp. 190–195. doi:10.1109/EDOCW.2019.00041.

[9] Bemthuis, R., Iacob, M.E., Havinga, P., 2020a. A design of the resilient enterprise: A reference architecture for emergent behaviors control. Sensors 20, 6672. doi:10.3390/s20226672.

[10] Bemthuis, R., Mes, M., Iacob, M.E., Havinga, P., 2020b. Using agent-based simulation for emergent behavior detection in cyber-physical systems, in: 2020 Winter Simulation Conference (WSC), IEEE. pp. 230–241. doi:10.1109/WSC48552.2020.9383956.

[11] Bemthuis, R.H., Koot, M., Mes, M.R.K., Bukhsh, F.A., Iacob, M.E., Meratnia, N., 2019. An agent-based process mining architecture for emergent behavior analysis, in: 2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW), IEEE. pp. 54–64. doi:10.1109/EDOCW.2019.00022.

[12] Bemthuis, R.H., Lazarova-Molnar, S., in press. Discovering agent models using process mining: Initial approach and a case study, in: 2022 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom).

[13] Bolt, A., de Leoni, M., van der Aalst, W.M.P., 2018. Process variant comparison: Using event logs to detect differences in behavior and business rules. Information Systems 74, 53–66. doi:10.1016/j.is.2017.12.006.

[14] Campos, J., Richetti, P., Baião, F.A., Santoro, F.M., 2018. Discovering business rules in knowledge-intensive processes through decision mining: An experimental study, in: Business Process Management Workshops, Springer. pp. 556–567. doi:10.1007/978-3-319-74030-0_44.

[15] Capello, F., Toja, M., Trapani, N., 2016. A real-time monitoring service based on industrial internet of things to manage agrifood logistics, in: 6th International Conference on Information Systems, Logistics and Supply Chain, pp. 1–8.

[16] Castelvecchi, D., 2016. Can we open the black box of AI? Nature News 538, 20–23. doi:10.1038/538020a.

[17] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., 2000. CRISP-DM 1.0: Step-by-step data mining guide.

[18] Crerie, R., Baião, F.A., Santoro, F.M., 2009. Discovering business rules through process mining, in: Enterprise, Business-Process and Information Systems Modeling. Springer, pp. 136–148. doi:10.1007/978-3-319-74030-0_44.

[19] De Smedt, J., vanden Broucke, S.K.L.M., Obregon, J., Kim, A., Jung, J.Y., Vanthienen, J., 2017. Decision mining in a broader context: An overview of the current landscape and future directions, in: Business Process Management Workshops, Springer. pp. 197–207. doi:10.1007/978-3-319-58457-7_15.

[20] Duling, D.R., 2017. Use machine learning to discover your rules. Paper SAS579-2017.

[21] Freitas, A.A., 2014. Comprehensible classification models: A position paper. SIGKDD Explorations Newsletter 15, 1–10. doi:10.1145/2594473.2594475.

[22] Golpîra, H., Khan, S.A.R., Safaeipour, S., 2021. A review of logistics internet-of-things: Current trends and scope for future research. Journal of Industrial Information Integration 22, 100194. doi:10.1016/j.jii.2020.100194.

[23] Goodman, B., Flaxman, S., 2017. European Union regulations on algorithmic decision-making and a "right to explanation". AI magazine 38, 50–57. doi:10.1609/aimag.v38i3.2741.

[24] Hariri, R.H., Fredericks, E.M., Bowers, K.M., 2019. Uncertainty in big data analytics: Survey, opportunities, and challenges. Journal of Big Data 6, 1–16. doi:doi.org/10.1186/s40537-019-0206-3.

[25] Iacob, M.E., Charismadiptya, G., van Sinderen, M., Piest, J.P.S., 2019. An architecture for situation-aware smart logistics, in: 2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW), IEEE. pp. 108–117. doi:10.1109/EDOCW.2019.00030.

[26] Janiesch, C., Koschmider, A., Mecella, M., Weber, B., Burattin, A., Di Ciccio, C., Gal, A., Kannengiesser, U., Mannhardt, F., Mendling, J., Oberweis, A., Reichert, M., Rinderle-Ma, S., Song, W., Su, J., Torres, V., Weidlich, M., Weske, M., Zhang, L., 2017. The internet-of-things meets business process management: Mutual benefits and challenges. CoRR, pp. 1–9.

[27] Jia, L., Alizadeh, R., Hao, J., Wang, G., Allen, J.K., Mistree, F., 2020. A rule-based method for automated surrogate model selection. Advanced Engineering Informatics 45, 101123. doi:10.1016/j.aei.2020.101123.

[28] Khemiri, A., Hamri, M.E.A., Frydman, C., Pinaton, J., 2018. Improving business process in semiconductor manufacturing by discovering business rules, in: 2018 Winter Simulation Conference (WSC), IEEE. pp. 3441–3448. doi:10.1109/WSC.2018.8632509.

[29] Koot, M., Mes, M.R.K., Iacob, M.E., 2021. A systematic literature review of supply chain decision making supported by the internet of things and big data analytics. Computers & Industrial Engineering 154, 107076. doi:10.1016/j.cie.2020.107076.

[30] Leewis, S., Smit, K., Zoet, M., 2020. Putting decision mining into context: A literature study, in: Digital Business Transformation, Springer. pp. 31–46. doi:10.1007/978-3-030-47355-6_3.

[31] Liu, H., Gegov, A., Cocea, M., 2015. Rule based systems for big data: A machine learning approach. volume 13. Springer. doi:10.1007/978-3-319-23696-4.

[32] Lovell, B.C., Walder, C.J., 2006. Support vector machines for business applications, in: Business Applications and Computational Intelligence. Idea Group Inc., pp. 267–290. doi:10.4018/978-1-59140-702-7.ch014.

[33] Lutu, P.E.N., Engelbrecht, A.P., 2010. A decision rule-based method for feature selection in predictive data mining. Expert Systems with Applications 37, 602–609. doi:10.1016/j.eswa.2009.06.031.

[34] Piest, J.P.S., Bemthuis, R.H., Charismadiptya, G., 2020. Demonstrating the architecture for situation-aware logistics using smart returnable assets, in: 2020 IEEE 24th International Enterprise Distributed Object Computing Workshop (EDOCW), IEEE. pp. 86–90. doi:10.1109/EDOCW49879.2020.00024.

[35] Plebani, P., Marrella, A., Mecella, M., Mizmizi, M., Pernici, B., 2017. Multi-party business process resilience by-design: A data-centric perspective, in: Advanced Information Systems Engineering, Springer. pp. 110–124. doi:10.1007/978-3-319-59536-8_8.

[36] Ross, R.G., 2003. Principles of the business rule approach. Addison-Wesley Professional.

[37] Rozinat, A., van der Aalst, W.M.P., 2006. Decision mining in ProM, in: Business Process Management, Springer. pp. 420–425. doi:10.1007/11841760_33.

[38] Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence 1, 206–215. doi:10.1038/s42256-019-0048-x.

[39] Sasaki, Y., 2022. A survey on IoT big data analytic systems: Current and future. IEEE Internet of Things Journal 9, 1024–1036. doi:10.1109/JIOT.2021.3131724.

[40] Schröer, C., Kruse, F., Gómez, J.M., 2021. A systematic literature review on applying CRISP-DM process model. Procedia Computer Science 181, 526–534. doi:10.1016/j.procs.2021.01.199.

[41] Singh, H., 2021. Big data, industry 4.0 and cyber-physical systems integration: A smart industry context. Materials Today: Proceedings 46, 157–162. doi:10.1016/j.matpr.2020.07.170.

[42] Steinke, G., Nickolette, C., 2003. Business rules as the basis of an organization's information systems. Industrial Management & Data Systems 103, 52–63. doi:10.1108/02635570310456904.

[43] Sullivan, Y., Wamba, S., 2022. Artificial intelligence, firm resilience to supply chain disruptions, and firm performance, in: Proceedings of the 55th Hawaii International Conference on System Sciences, pp. 5913–5922. doi:10125/80059.

[44] Taveter, K., Wagner, G., 2001. Agent-oriented enterprise modeling based on business rules, in: Conceptual Modeling — ER 2001, Springer. pp. 527–540. doi:10.1007/3-540-45581-7_39.

[45] Vorm, E.S., 2018. Assessing demand for transparency in intelligent systems using machine learning, in: 2018 Innovations in Intelligent Systems and Applications (INISTA), IEEE. pp. 1–7. doi:10.1109/INISTA.2018.8466328.

[46] Wang, J., Yang, Y., Wang, T., Sherratt, R.S., Zhang, J., 2020. Big data service architecture: A survey. Journal of Internet Technology 21, 393–405.

[47] Wang, M., Wang, H., 2006. From process logic to business logic—a cognitive approach to business process management. Information & Management 43, 179–193. doi:10.1016/j.im.2005.06.001.

[48] Wang, W., Indulska, M., Sadiq, S., 2018. Guidelines for business rule modeling decisions. Journal of Computer Information Systems 58, 363–373. doi:10.1080/08874417.2017.1285683.

[49] Wirth, R., Hipp, J., 2000. CRISP-DM: Towards a standard process model for data mining, in: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, pp. 29–39.