Analysis and prediction of purchasing compliance using process mining

Vincent F. A. Meyer zu Wickern, Mounisha Juluru, Anshu Roy, Thi-Thu-Hang Nguyen, and Viet-Hung Vu

Faculty of Communication and Environment, Rhine-Waal University of Applied Sciences, Friedrich-Heinrich-Allee 25, 47475 Kamp-Lintfort, Germany

Abstract. The annual Business Process Intelligence (BPI) challenge is an opportunity to find and apply new techniques for process mining analyses on provided datasets, which in this year 2019 is an event log of purchase-to-pay events of a multinational paints and coatings company. The anonymized process owner is interested in gaining information on the compliance of their process through the discovery of a collection of process models, the analysis of throughput times in their invoicing process and the separation of deviating purchase order items. This paper is focused on the detection of incompliant purchase order items to support the process owner's compliance overview. Compliance of purchasing may be violated through value differences between the company's order items, their goods receipts, and their invoices. Further, the defined process flow may be violated. Multiple process models are discovered belonging to different purchase order item attributes to understand the process landscape of the anonymized company. Additionally, a method to determine incompliant cases is developed. Predictive process mining is used to identify influences of purchase order item attributes, derived additional features and process flow to the probability of breaching compliance. With the consideration of these influences, more incompliant process executions may be detected or prevented.

Keywords: Process Mining · Three-Way Matching · Purchasing Compliance · Predictive Process Monitoring.

1 Introduction

The Business Process Intelligence (BPI) Challenge 2019 is a challenge in the field of process mining related to the International Conference on Process Mining (ICPM) conference which is to be held in Aachen, Germany [5]. The BPI Challenge 2019 requests all participants to analyze an event log of a purchase-to-pay process provided by an anonymized process owner. The process owner is especially interested in insights into their compliance. The BPI Challenge is split into student and non-student submissions, and this paper strives to analyze the given dataset on compliance as a student submission. For this, the major influencing factors on purchasing compliance and an opportunity to improve it were sought after. Predictive process mining was used to identify attributes of the purchase order (PO) items, the additional derived features, and the process flow, which increase the probability of breaching compliance.

The following tasks should be answered to support the process owner in the procurement compliance improvement as defined in the challenge description [5]:

- Find one or more process models that could accurately describe the processes in the event log.
- Find the throughput time of the invoice process and suggestions to mitigate the bottlenecks.
- Find unusual activities and documents in the event log.

These questions are expected to be solved with the help of process mining. Process mining is a research discipline that aims to discover, monitor, and improve real processes by extracting knowledge from event logs that are available in organizations' information systems [3]. For this paper, multiple machine learning techniques were used to support the analysis of the company's purchasing compliance and to potentially further leverage machine learning techniques in process analyses. The main goal of the paper is to find correlating attributes in the given event log that may provide insights into the occurrence of incompliant cases. Predictive process mining was therefore used to classify PO items based on their compliance and find predictors for incompliant PO items both with the complete knowledge of the event log and with the restriction of knowing only information up to a given event time in a PO item process.

As the quality of the prediction may be influenced by a reasonable preprocessing of the dataset, the dataset was first explored for exceptional and potentially unusable data points in section 2. An understanding of the company's processes may additionally be beneficial for a classifier's success, so the processes as seen in the data were visualized. Trace clustering techniques were used in section 3 to identify different process areas and thereby enhance the simplicity and precision of the models. A calculation technique for the second request of the company regarding throughput time is provided in section 4, followed by the creation of additional dimensions in section 5, in which deviating PO items were identified. The key focus in the detection and analysis of deviating PO items was laid upon incompliant PO items and PO items with rework. The classifiers to predict incompliant behavior are further described in section 5.

2 Dataset exploration and preprocessing

The provided anonymized data consists of a list of PO item events submitted in 2018, which were collected from a coating-and-paints company in the Netherlands [5]. From the dataset, each PO item (or case) can be uniquely identified using the *_case_concept_name_* column representing the concatenated PO number and the line item number [5]. There can be multiple events (or activities) from one PO item, each of which has its own event name (*_event_concept_name_* column) and timestamp (*_event_time_timestamp_* column). In general, the data is split into attributes that are valid for an entire case, i.e., a PO item, and attributes that are valid only for single events, i.e., the activities that were performed. The PO items could be pre-classified into four main categories as defined by the BPI challenge:

- 3-way match, invoice before GR
- 3-way match, invoice after GR
- 2-way match
- Consignment

Not all of the presented events could be used for the analyses in this paper, because they may not be comparable. The lack of comparability may be due to the following reasons, which are taken into account in this chapter:

1. Events did not occur in the extraction timeframe
2. Cases are not complete

A special focus of the analyses in this pager is set on the compliance of cases, which is not provided in the BPI challenge dataset. Therefore the specifications of compliance used in this paper are explained in this chapter.

2.1 Timeframe exclusion

The BPI challenge description defines that there are "over 1.5 million events for purchase orders submitted in 2018" in the dataset [5]. Therefore the validity of PO items created before or after this timeframe could not be ensured, and the dataset needed to be filtered on the timeframe for further processing of the data. As the data was submitted on 28.01.2019, this was the final date that was accepted in the preprocessing of the event log. The earliest point that was accepted in the data was the beginning of 01.01.2018.

2.2 Compliance of cases

As the process owner has compliance questions, compliance is critical in the analyses created in this paper. To verify whether a PO item is compliant, the necessary activities have have been reviewed, and the monetary values set in the events had to be checked. For this, first, a compliance check was made verifying all possible compliance criteria. In the second step, the non-compliant PO items, which have not been completed yet, were separated from full cases that are non-compliant due to a mismatch in 3-way matching or 2-way matching. The word "incompliant" does not necessarily indicate any mistakes made in the process, because also ongoing and therefore incomplete cases have to be regarded as incompliant, e.g., a PO item without payment cannot be seen as compliant. Therefore, the split into complete and incomplete cases is necessary.

Verification of compliance
To review the compliance of the PO items, the monetary values, and counts of PO items, goods receipts (GRs), and invoice receipts were checked. Both the

counts and monetary values should be equal for the PO items to be compliant. However, the case item category has to be considered because not for all case item categories, invoices and GRs can be expected. Since the invoicing process for consignment order items is not conducted in the observed system, no invoices or clearing of these invoices can be expected. For items of the category *2-way match*, no GRs are required, therefore also PO items without GR have to be regarded as compliant. The count of GRs and invoice receipts were reduced if cancellations were recorded. Because the purchase-to-pay process requires a payment to meet the customer's obligation, the *Clear Invoice* activity was seen as necessary for compliance. Consequently, the rules presented in the following were applied to categorize incompliant PO items in relation to their item category:

All, except *2-way match*: Number of GRs > 0

All, except *Consignment*:

- Number of *Clear Invoice* events > 0
- Number of invoice receipts > 0

Consignment: Cumulated value of the GRs divided by the number of GRs = PO item value

3-way match, invoice after GR, *3-way match, invoice before GR*:

- PO item value = Cumulated value of the invoice receipts divided by the number of invoice receipts
- Cumulated value of the GRs = Cumulated value of the invoice receipts
- Number of GRs = Number of invoice receipts

Incomplete PO Items

A subset of the incompliant PO items are the incomplete cases. These are not incompliant due to any wrong entries or behavior, but only because they have not yet been finished. Inclusion of these items in defining criteria that lead to incompliant cases may, therefore, lead to incorrect findings, and the items have to be excluded for such an investigation. Incomplete PO items can be recognized by their end activities, which is assumed to be *Clear invoice* for all cases except *Consignment* PO items, which do not involve an invoicing process. The recording of a GR is assumed to be the end activity for *Consignment* PO items. Furthermore, more than one GR and invoice receipt are possible in the purchasing process, which may also lead to multiple clearings of invoices. In *3-way match: invoice after goods receipt* order items, invoices have to follow goods receipts, so incomplete orders are those, in which fewer invoice receipts were received than goods receipt, reduced by cancellations. For *3-way match: invoice before goods receipt* order items, the order of GRs and invoice receipts is insignificant, so both a lower count of goods receipts or invoice receipts compared to the respective other implies an incomplete PO item.

Figure 1 presents an overview of the four main clusters with the number of compliant and complete PO items within them. It is evident that all incomplete

PO items are also incompliant, but that complete PO items are split into compliant and incompliant ones. Especially for *2-way match* PO items, there are many incomplete cases in the dataset and the number of these even surpasses the count of complete cases. These cases lack payments and can therefore not be seen as complete. Considering that *Limit* PO items, which these *2-way match* items are, are typically open for a long time, this observation reflects their estimated distribution for a one-year extraction limit, in which many PO items may be assumed to still be open. Similarly, *3-way match: Invoice after GR* PO items facilitate the receipt of multiple deliveries with GR-based invoicing and therefore were assumed to have a comparatively longer throughput time than *3-way-match: Invoice before GR*. This is also visible in the ratio of cases that were not completed within the extraction time frame.

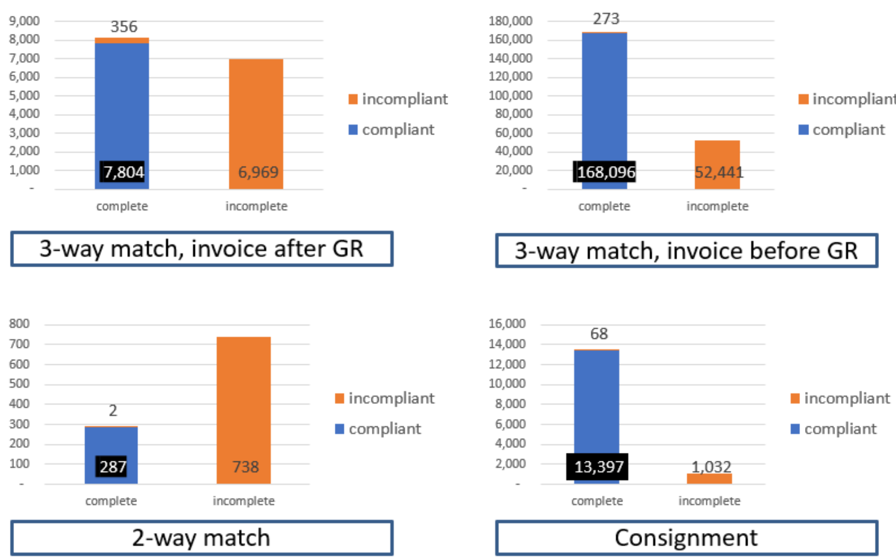

Fig. 1. The four main clusters by compliance and completeness

3 Process models generation

The first main request of the company in focus is the description of its process models that can reflect the current situation of purchasing process execution in the company. Therefore, the analysis in this chapter is directed on the discovery of the as-is process models. Activities, which are not regularly considered part of a standard purchase-to-pay process, such as *Change quantity* or *Change price*, may be part of these models if they are conducted in the majority of cases. This may enable the company to compare these models to the normative models that they have set up. However, to avoid the consideration of a high level of noise and deviations, only compliant cases as defined in chapter 2 were used for determining the process models.

Each generated model should satisfy the four competing success indicators of *fitness*, *simplicity*, *generalization*, and *precision* as described in [4]. Generalization and precision are indicated by the new behavior that is possible with the process model and cannot be observed in the event log. While generalization is positively influenced by enabled additional behavior, the precision indicator is designed to restrict the process model to the observed behavior. As such, in this paper, a high but not maximized precision is aimed for, especially restricting precision when designing a model for a smaller log. Fitness refers to the ability to replay the behavior of the event log in the process model. Fitness of the models in this paper was ensured by allowing for as many activities and connections as necessary to replay the event log. Fitness was measured with alignment-based replay implemented in the *Multi-Perspective Process Explorer* ProM plugin [2] in this paper. This plugin was also used for measuring model precision. Including more connections and activities in the process models may improve model fitness but may damage the simplicity, which, however, can be improved with a reasonable splitting of the dataset to logical sub-groups and with filtering infrequent activities and connections. Therefore, in the following, the grouping of the dataset into reasonable sub-groups is described.

3.1 Split into case groups with similar process flows

The goal for the creation of the models was to have a small number of process models that together can describe the traces encompassed by them. In the first split, the dataset was grouped into four main groups that were already defined in the BPI challenge. In figure 1 in section 2, the counts of these groups can be seen together with their classification on compliance. The process models are based on the filtered data as described in chapter 2 and only compliant cases are considered. The notion of incompliance does not necessarily indicate a wrong behavior in a case, but may also be due to a case that was not conducted until the end and in which for example a payment may be missing.

In the next step, the event log was analyzed for a pattern in the event count and event description. In the event count, it is noticeable that seven activities with the prefix *SRM:* were all executed in 1,119 cases. This led to the assumption that an SRM (Supplier Relationship Management) system was used for a subset of the PO items. In a further analysis of the PO items with SRM activities, it was found that all occur for the document type *EC Purchase Order* and that they were only conducted for *3-way match, invoice before GR*- and *3-way match, invoice after GR*-cases. Since the inclusion of an SRM system indicates other process flows, it was decided to group the event log further into cases including SRM activities and cases without SRM activities.

To analyze, whether other case attributes would enhance the clarity of case process flows for these clusters, the numbers of distinct values for the different case attributes were reviewed. In table 1, these numbers are listed. The first two attributes *_case_Purch__Doc__Category_name_* and *_case_Source_* both have a value of 1, which means that another split on these attributes is not possible. The columns *_case_GR_Based_Inv__Verif__*, *_case_Goods_Receipt_* and *_case_Item_Category_* are all used to build the first level grouping as seen above and would not split the data further so that these attributes could further be neglected. *_case_concept_name_*, *_case_Purchasing_Document_* and *_case_Item_* are identifiers of the POs and their items, so these should also be neglected as split criteria for a broad grouping of purchases. Other attributes, which allow for many distinct values are *_case_Spend_area_text_*, *_case_Sub_spend_area_text_*, *_case_Name_* and *_case_Vendor_* ranging between 21 and 1,552 values. To ensure simplicity in understanding the collection of models, these columns were further neglected due to the high number of process models that would have been generated when splitting by these columns. The columns left to be analyzed were *_case_Company_*, *_case_Document_Type_*, *_case_Spend_classification_text_* and *_case_Item_Type_*.

column name	count of distinct values
_case_Purch__Doc__Category_name_	1
_case_Source_	1
_case_GR_Based_Inv__Verif__	2
_case_Goods_Receipt_	2
_case_Company_	3
_case_Document_Type_	3
_case_Item_Category_	4
_case_Spend_classification_text_	4
_case_Item_Type_	6
_case_Spend_area_text_	21
_case_Sub_spend_area_text_	133
_case_Item_	318
_case_Name_	1487
_case_Vendor_	1552
_case_Purchasing_Document_	55994
_case_concept_name_	189584

Table 1. Count of distinct values per case column

The dissimilarity within each generated group was analyzed to verify whether a split by these columns would add more precise process models than with a broader dataset. If the sum of the within-sum-of-squares (WSS) was significantly lower in the split groups than in the overall dataset, a split by these groups would be reasonable. A simple version of trace profiling was used to measure the dissimilarity of cases within a cluster. For each case, the number of occurrences for each activity was counted, and to measure the dissimilarity between two cases, the Euclidian distance between two activity count vectors was calculated. To measure the WSS of one cluster, the dissimilarity of each case vector to the center of this cluster was added to one sum. For the comparison to the dissimilarity of only four groups by item category, all WSS values for one split criterion, e.g., document type or company, were added. In figure 2, the WSS values for all split criteria are displayed.

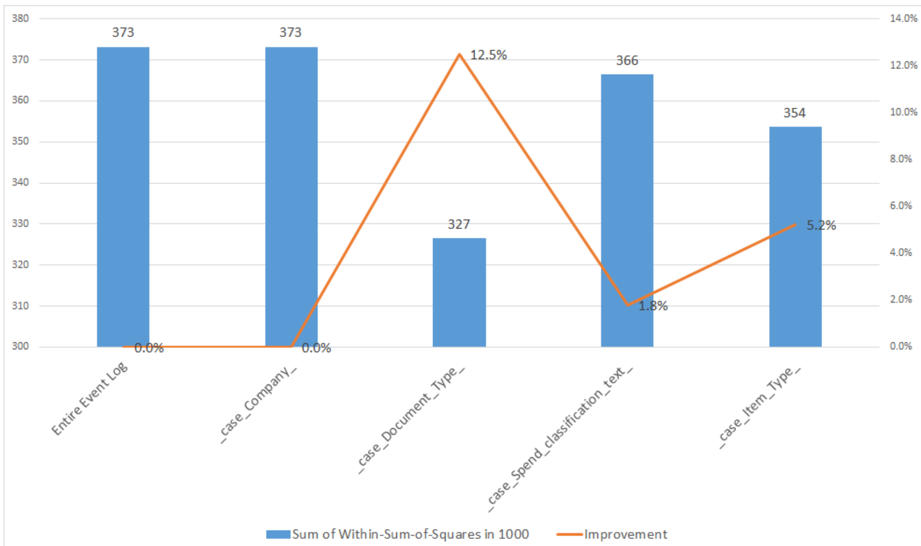

Fig. 2. Comparison of the inner cluster dissimilarity by case criteria

With figure 2, it becomes evident that the case attributes *_case_Document_Type_* with 12.5% improvement and *_case_Item_Type* with 5.2% improvement achieve the highest differences in WSS values. The difference induced by the split by *_case_Document_Type_* matches the above-explained difference of process flows through the inclusion of an SRM system in EC Purchase Orders. Based on these split criteria, the process models were created. When the process models were generated, it was found that the datasets for framework orders and standard POs for *Service* items in the category *3-way match, invoice after GR* yielded the same process models, hence these were grouped. Since a difference between framework orders and standard POs then did not lead to splitting, the split of document types was changed to a split into PO items with SRM activities and PO items without SRM activities. Similarly, the groups of Subcontracting and Third-Party item type PO items for standard POs in the category *3-way match, invoice after GR* were grouped, because they also yielded the same process models. The new groups are displayed in table 2. The cluster IDs for these groups are used for reference in the following.

cluster ID	name	count of PO items
01_01_01	3-w. match, inv. after GR (with SRM; Item Type: Service)	318
01_01_02	3-w. match, inv. after GR (with SRM, Item Type: Standard)	96
01_02_01	3-w. match, inv. after GR (without SRM; Item Type: Service)	676
01_02_02	3-w. match, inv. after GR (without SRM, Item Type: Standard)	6446
01_02_03	3-w. match, inv. after GR (without SRM, Item Type: Subcontracting and Third-Party)	268
02_01	3-w. match, inv. before GR (with SRM)	705
02_02_01	3-w. match, inv. before GR (without SRM, Item Type: Standard)	161613
02_02_02	3-w. match, inv. before GR (without SRM, Item Type: Subcontracting)	1611
02_02_03	3-w. match, inv. before GR (without SRM, Item Type: Third-Party)	4167
03	2-way match	287
04	Consignment	13397

Table 2. Case groups by item category, SRM inclusion and item type

3.2 Description of the process models

With the split event log, all process models could be generated. To be able to compare the process models, the same filter and mining parameters were set for all of them. With a heuristics filtering of a minimum event observation as well as start event and end event observation of 10% for each group, the simplicity of the generated models was enhanced. After this, the inductive miner was applied with a noise threshold of 20%, reducing the rarely observed edges in the graphs and enhancing simplicity. As the models are intended to be understood by the process owner, BPMN as a standard for business process modeling was used for the visualization of the models [1]. The fitness and precision of the initially created Petri nets and the simplicity in the form of the number of nodes and edges of the converted BPMN models for the case clusters are presented in table 3. The predominantly high values in precision and thereby low values in generalization are accepted due to the simplicity of the models, which implies a lower degree of overfitting to the event data. Weighted averages of 93% for fitness and 99% for precision could be achieved. All process models are attached in the appendix of this paper.

In the following, the main activities and building blocks of the purchasing process are described without specifying the activity sequence, since the sequence and repetition of activities may be different for every process model.

In the process models, a PO item is created either with an SRM system or another not further specified system. The creation of the PO item may be combined with other activities for the request and approval of this PO item, e.g., *Create Purchase Requisition Item* or *SRM: Awaiting Approval*, but this approval is said to not be mandatory in the challenge description [5]. In the progress of a PO item then a form of goods or service receipt is recorded with the activities *Record Goods Receipt* and for PO items of the item type *Service* additionally the activities *Record Service Entry Sheet*. The vendor creates an invoice, expressed in the activity *Vendor creates invoice* and this invoice receipt is recorded with

cluster ID	number of cases	fitness	precision	number of nodes	number of edges
01_01_01	318	87%	78%	18	22
01_01_02	96	96%	100%	15	15
01_02_01	676	91%	62%	13	17
01_02_02	6,446	94%	100%	10	11
01_02_03	268	94%	100%	8	8
02_01	705	95%	100%	15	15
02_02_01	161,613	93%	100%	10	11
02_02_02	1,611	90%	88%	15	19
02_02_03	4,167	92%	91%	12	13
03	287	97%	70%	10	11
04	13,397	94%	100%	6	6
Sum / Weighted Avg.	**189584**	**93%**	**99%**	**10**	**11**

Table 3. Process model quality indicators

the event *Record Invoice Receipt*. Invoices are cleared with the activity *Clear Invoice*. If a payment block was prior set for the invoice or PO item, e.g., due to a missing GR for the invoice, this payment block has to be removed before the clearing of the invoice with the activity *Remove payment block*. Additional to the described activities, there may be more activities relevant to only some cases.

The main difference of the process models can be recognized in the following areas, which are described in the following.

- Inclusion of the SRM system
- Loops of GRs, invoice creations, invoice receipts and invoice clearings
- The sequence of GRs and invoice receipts
- Integration of payment block removals
- Other special activities

Inclusion of the SRM system

An SRM is included in the clusters 01_01_01, 01_01_02, and 02_01 and supports the creation of the PO item at the start of the purchasing process. The SRM block always starts with the *SRM: Created* activity and additionally a PO item is created, which is mandatory in clusters 01_01_01 and 02_01 and only optional in group 01_02_01. After the document creation, approval is awaited in all clusters. Afterward, there is a change transmittance, which is followed by the *SRM: Complete* activity and a *SRM: Document Completed* event in all case groups. The two activities *SRM: Ordered* and *SRM: In transfer to Execution System* are conducted subsequently; however, in the different process models, different direct sequences were discovered. While cluster 01_01_01 identifies *SRM: Ordered* and *SRM: In Transfer to Execution System* as parallelly running activities, the groups 01_01_02 and 02_01 discover a direct succession from *SRM: In Transfer to Execution System* to *SRM: Ordered*.

Loops of GRs, invoice receipts and invoice clearings

In the analysis of the activities of GRs, invoice receipts, and invoice clearings the possibilities for different frequencies of these activities per PO item could be observed in the various case groups. The following possible process flows can be distinguished:

1. One GR, one invoice receipt, one invoice clearing
2. Multiple GR, multiple invoice receipts, multiple invoice clearings
3. No GR, one invoice receipt
4. Multiple GR, no invoice receipt

The first group with one GR, one invoice receipt, and one invoice clearing could be found for all item types except the *Service* item type. The possibility for multiple GRs, multiple invoice receipts, and multiple payments is observable in process models of the item type *Service* represented by case groups 01_01_01 and 01_02_01. The occurrence of multiple GRs and invoices, therefore, appears predominantly for service items, while standard, subcontracting and third-party items at least in the vast majority of cases do not show this behavior.

An invoicing without GR is conducted for PO items of the item category *two-way matching* according to the process model generated for them in case group 03. This matches the definition of two-way matching in the BPI challenge 2019 description [5].

For consignment PO items, one GR record and no invoice receipt can be created according to its process model for case group 04. This also matches the specifications of the BPI challenge 2019, because the invoicing process for consignment orders are conducted in another system that is not included in the event log [5].

Sequence of GRs and invoice receipts

A difference in the sequence of the GRs and invoice receipts for PO items can be seen in the process models discovered from the event log. One possibility observed in the process models is that a GR is succeeded by one or multiple invoice receipts, which is true for the majority of the process models and PO items. If there are loops of multiple combinations of GRs and invoice receipts, in which for the loop iterations the invoice receipt(s) succeed the receipt of the goods, the process models are also counted to be in the GR-invoice receipt succession group. The following case groups follow this pattern: 01_01_02, 01_02_02, 01_02_02, 01_02_03, 02_01, 02_02_01, 02_02_03, 03 and 04.

Another behavior observed for the case groups 01_01_01 and 02_02_02 is a parallel execution of GR and invoice receipt, in which the order of these is not critical. Although a GR and invoice receipt parallelization describes the specification of the item category *3-way match, invoice before GR*, the majority of cases follows the GR-invoice receipt succession.

Inclusion of payment block removals

For three case groups, the activity *Remove payment block* can be seen in the process models. This activity is primarily described as used for releasing the payment after a mismatch between invoice receipt and GR is solved by an incoming GR [5]. This behavior would be assumed in the item category *3-way match, invoice before GR* due to the possible parallel execution of *Record Invoice Receipt* and *Record Goods Receipt*. Two of the case groups are from this item category, i.e., the case groups 02_02_01 and 02_02_03 for the item types *Standard* and *Third-Party*. Additionally, the activity can be found in the case group 01_02_02, which is valid for the PO items of category *3-way match, invoice after GR*, and type *Standard*.

Other special activities

There are activities which are only present in process models for some case groups. The reasons for including or excluding these are not given and provide a lead for further investigation.

A purchase requisition item is only included beyond the noise threshold in case cluster 04. In the BPI challenge 2019 description, it is said that approval workflows are not in focus. However, it may be valuable for the company to check whether the purchase requisition items are created according to the rules that have been set up for them.

The activity *Change approval for Purchase Order* can be found for the case group 03. In cluster 03 for 2-way-matching, the approvals for the PO may even be changed multiple times and at least one time. It is not evident, what the meaning of these approval changes are and why they occur; however, this may also be a valuable area to research for the company in focus.

Prices changes and quantity changes are part of the process model for case group 02_02_02. Here it is a choice in the process model to change the price after the PO item creation. Quantity changes are even regarded as a mandatory activity. With this finding, an opportunity for investigation is released asking why so many price changes and quantity changes occur that they are included despite the applied noise threshold.

3.3 Validation of case group and process model quality

While the generated process models lead to satisfactory model criteria in fitness, precision, and simplicity, the respective process models do not always represent the expected behavior and possibilities. As the case group 02_02_01 covers the most cases with 161,613 process models, this group is reviewed more closely.

In the process model presented in figure 3, a recording of a GR always precedes a recording of an invoice receipt. This does not reflect the behavior for PO items of item category *3-way matching, invoice before goods receipt*, for which "invoices can be entered before the goods are receipt" [5].

To review, whether an infrequent, but correct behavior was filtered by noise filtering, the trace clustering ProM plugin ActiTraC was used, which allows

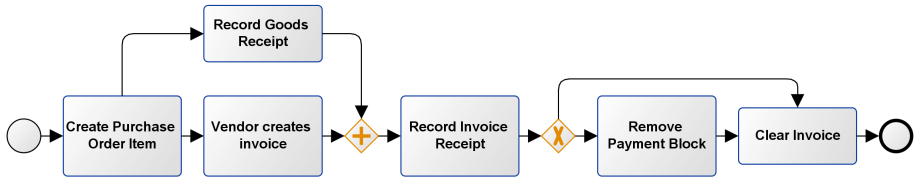

Fig. 3. Process model for case group 02_02_01

grouping traces into groups of similar behavior. With this technique, two large groups of traces could be identified, and the third group with additional behavior seen in the log was created. All process models with their number of PO items are shown in figure 4.

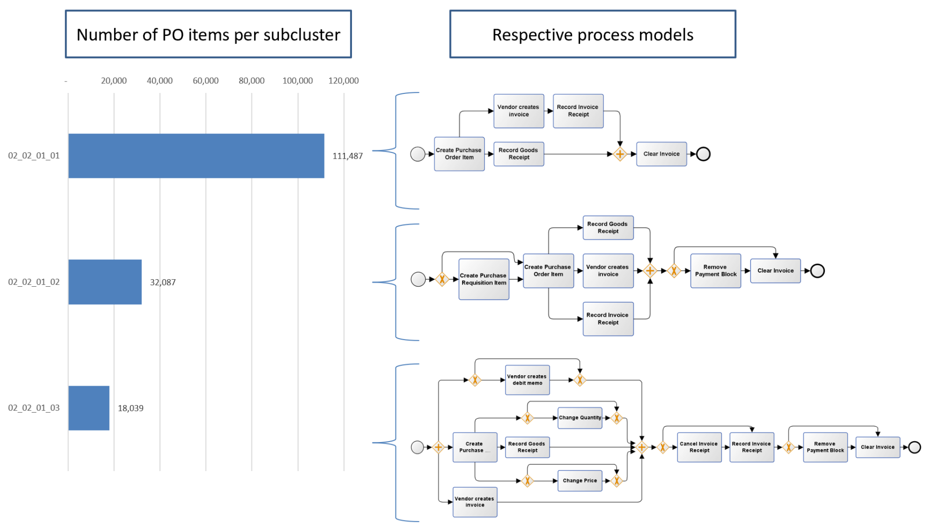

Fig. 4. Subcluster models for case group 02_02_01

In the first two process models, no order is defined for the sequence of the activities *Record Goods Receipt* and *Record Invoice Receipt*. In the third process model, which is applicable to ca. 11% of the PO items, an indirect succession from a GR to an invoice receipt is defined. It can be derived that for the majority of PO items, a GR is followed by an invoice receipt and the cases, in which an invoice receipt is followed by a GR, are filtered in the noise threshold for the combination of all cases of the group 02_02_01 or for the subcluster 02_02_01_3.

Due to the possibility of finding subclusters by ActiTraC, it was inspected whether there is a possibility to split the event log further while still taking into account the case attributes with more distinct values, namely the columns _case_Spend_area_text_, _case_Sub_spend_area_text_, _case_Name_, and _case_Vendor_. The asynchronous correlation of these columns to the found clustering was calculated with the uncertainty coefficient, and the values are presented in table 4.

column	uncertainty coefficient
_case_Spend_area_text_	0.0208
_case_Sub_spend_area_text_	0.0505
_case_Name_	0.1736
_case_Vendor_	0.1780

Table 4. Uncertainty coefficients of case attributes with many distinct values

Although the values depicted in table 4 indicate a small level of correlation, this meets the expectations since there are many values for both the dependent and the independent variable. A maximum uncertainty coefficient of 0.1780 did not reason a further split of any these columns to decide upon the process models. Therefore the presented groups are assumed to be natural presentations of user and PO item process behavior of the same category or linked to an attribute, which may not be provided in the dataset.

4 Throughput Time Analysis

The performance of an organization can be measured using time-related indicators like the throughput time of a process. The average time that a PO item spends in the process is termed throughput time. For this challenge, in this section the focus is to answer the challenge question "What is the throughput of the invoicing process, i.e., the time between GR, invoice receipt, and payment (clear invoice)? " [5].

As per the as-is processes and the models explained in section 3, for one PO item, there can be multiple GR messages and multiple invoice receipt messages. To identify which of them belong together within one PO item and further calculate the time taken for the execution of the events, the sequence of the below steps is executed.

1. A consecutive number of rows is assigned using ROWNUMBER() in SQL by sorting on the _eventID__ in ascending order OVER partitioning by _case_concept_name_ and _event_concept_name_. This newly identified column is termed as *iteration_number*.
2. The time between applicable activities *Record Goods Receipt*/*Record Invoice Receipt* to *Record Invoice Receipt*/*Clear Invoice* that belong to the same iteration_number and same PO item is calculated and the average value of them over every _case_concept_name_ is taken as the time between GR, invoice receipt, and payment.

The above-mentioned technique bears the risk that not all the above-defined activities are present in every case. Further, the existence of the activities *Cancel Invoice Receipt* and *Cancel Goods Receipt* can result in negative values which lead to inaccuracy in the throughput time of the invoicing process. The entire data is filtered by the rules below, to avoid this inaccuracy.

1. All complete cases, as explained in section 2 are considered.
2. Cases that contain at least one of the activities *Cancel Invoice Receipt* and *Cancel Goods Receipt* are not considered.
3. Cases in which the invoice is cleared before the invoice receipt are not considered.
4. Cases belonging to _case_item_category_ *3-way match, invoice after GR*, for which invoice was created before the receipt of the goods are not considered. Finally, the time difference between the GR and the invoice receipt as well as the time difference between the invoice receipt and the invoice clearing are calculated.
5. For cases which belong to _case_item_category_ *3-way match, invoice before GR*, the invoices can be recorded before the receipt of goods. Hence, the entire data of this category is divided into the following subcategories:
   - PO items, for which invoices were recorded before good receipts: The time difference between the invoice receipt and the GR as well as the time difference between the GR and the invoice clearing are calculated.
   - PO items for which invoices were recorded after GRs: The time difference between the GR and the invoice receipt as well as the time difference between the invoice receipt and the invoice clearing are calculated.
6. Cases belonging to _case_item_category_ *2-way match*: Since there is no separate GR message, the time difference between the invoice receipt and the invoice clearing is calculated.
7. Consignment cases were not considered since no invoice receipts were recorded in the system.

The implemented results of the above-mentioned steps, along with case attributes and event attributes were imported into *celonis*. The throughput time of the invoicing process for different values of _case_item_category_ was analyzed by grouping on _case_item_type_ and _case_Document_Type_ using the column chart component in *celonis* as shown in figure 5 and figure 6.

The interpretation and further analysis of the throughput times are not followed due to the following reasons:

1. There could be a delay in the company receiving the invoice receipt and the user recording the same invoice receipt in the system, but these cannot be known from the data.
2. The payment terms are not specified. However, the time between the invoice receipt and the payment depends on these. If the payment terms were known, the adherence to these could be calculated with the throughput times as described above.

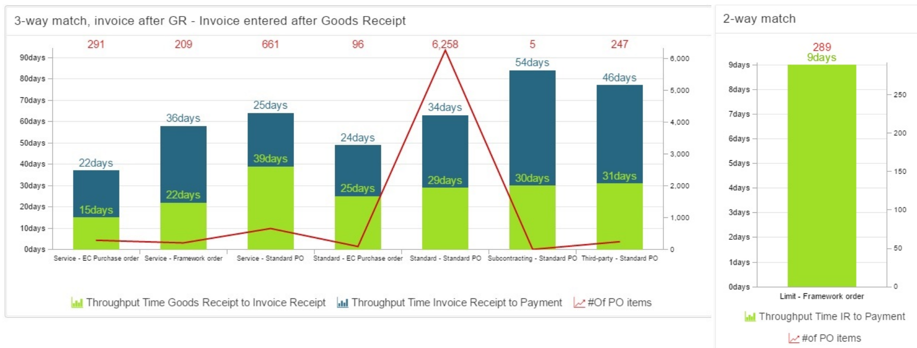

Fig. 5. Throughput times for *3-way match - invoice after GR* and 2-way match PO items

Fig. 6. Throughput times for *3-way match - invoice before GR*

5 Analysis of process deviating purchase order items

In this chapter, the given event log is analyzed on the compliance and the factors for incompliant and thereby deviating PO items. For this, the first additional dimensions are proposed which subsequently can be used in classifiers to detect incompliant cases based on their attributes and events. One of the additional dimensions aims to identify rework activities, which are presented in subsection 5.1. Fostered by the additional dimensions, the compliance classifiers are then described, and findings in their interpretation are presented.

5.1 Additional dimensions

Additional dimensions were created to analyze other parameters influencing the cases to turn out compliant or incompliant. The focus was to introduce other features, deduced from the event log to evaluate their contribution towards the case compliance.

Since a typical case involved several events associated with it, additional dimensions were further calculated on both case level and event level and could thereby be used for either post-mortem based predictions or pre-mortem based predictions, as explained in subsection 5.3. The approach was to gain more in-depth insight into the case compliance using all the information available at the case level. On the event level, the same features were fed to the classifiers that were available until the timestamp of a particular event occurrence.

Number of handovers: On the case level, the number of users excluding "batch_user" and "NONE" who were involved in a particular case defined the sum of handovers of work per case. At the event level, the dimension was calculated to understand the sequence in which each user was handed over the work against each event in a case.

Count of rework activities: Any event with _event_concept_name_ that indicated rework was identified and counted. These activities generally include "change," "cancel," or "delete" keywords in their activity names. A Boolean indicator "is_rework" was calculated against each event and later summed up by _case_concept_name_ to measure the final count of rework as the case level dimension and until a particular event for the event level dimension. Rework activities are explained in detail in subsection 5.2.

Segregation of duty (SOD): The dimensions *sod_create_poi_and_gr* and *sod_create_poi_and_ir* would be true against each case if the same user was responsible for creating the purchase order item with the user who recorded goods receipt or invoice receipt. On the event level, this dimension would inform whether the segregation of duty was achieved or not at each event occurrence.

Retrospective PO items: To understand the deviations from the process flow, particular records, in which the receipt of the goods was logged before the creation of purchase order item or in which the invoice was recorded before logging the receipt of the goods, were marked as retrospective purchase order items by a boolean. The background of this measure was that retrospective PO items may indicate that the purchasing process was not followed and the order may have been placed without creating a PO item in the system.

Resource Workload: This dimension provided the information about the work performed by each user in the past two days and seven days to understand the workload of the resource at the time of occurrence of the event. To use these values, they were aggregated on the case level with their average and their maximum values. Since it is possible that the quality of some activities is affected differently by higher workload than for others, this dimension was drawn in relation to the activity names.

Create Order net value: The total worth of a PO was calculated and appended in the case level attributes.

Throughput time: The throughput time is defined as the time from creating a PO item to the last event in the particular case. This time was further calculated based on the timestamp of occurrence of each event in days.

Material count: It was noticed that some POs contain items belonging to different groups of _sub_spend_area_text_. Therefore, this dimension was added to count the different materials in one purchase order.

Missing material: This dimension indicates in a Boolean value, whether a_sub_spend_area_text_ was given in the data or not. The assumed relation to compliance was drawn, because missing sub spend areas may indicate missing master data and thereby a higher chance of wrong entries.

Process cluster: As explained in section 3, all cases with their events were grouped into clusters, and this metric was also used as an additional dimension or feature for the decision tree learning.

Number of orders placed on the same day to the same vendor: This feature measures the count of orders on one day to the same vendor and applies this to the creation days of the PO items. The reason that this indicator was drawn is that confusion and potentially wrong allocations may arise between different orders that were placed on the same day for the same vendor.

5.2 Rework analysis

In this section, the rework steps were analyzed using *celonis* to identify the number of rework activities and the number of PO items with rework activities. Initially, a variable *reworkActivities* was defined with the set of defined activities as listed below:

Change Price, *Change Quantity*, *Cancel Invoice Receipt*, *Cancel Goods Receipt*, *Delete Purchase Order Item*, *Change Approval for Purchase Order*, *Change Currency*, *Change payment term*, *Change Final Invoice Indicator*, *Change Delivery Indicator*, *Change Rejection Indicator*, *Change Storage Location*, *Cancel Subsequent Invoice*, *Change Rejection Indicator*

The total distribution of the case and event data explained in section 2 was considered and a component filter on the defined set of *reworkActivities* was implemented on the column chart to visualize the number of PO items and the number of events in which rework was performed. The results are visualized in figure 7. The rework activities *Change Quantity*, *Change Price* and *Delete Purchase Order Item* were observed to occur most often in the total distribution.

5.3 Compliance classification on case level

The identification of the origins of incompliant cases was aimed to be solved with predictive techniques of machine learning. Using the relevant data to learn which features provide information about the compliance of a PO item, a classifier was built that could categorize a set of features to belong to either an incompliant or a compliant PO item. Using a *post-mortem* dataset with the complete knowledge about complete PO items, it was verified whether such a classification is possible. In a productive environment, however, during the execution of a purchasing process, not all information about a case is known, and only the information up to a specific event can be taken into account. Therefore, a *pre-mortem* dataset,

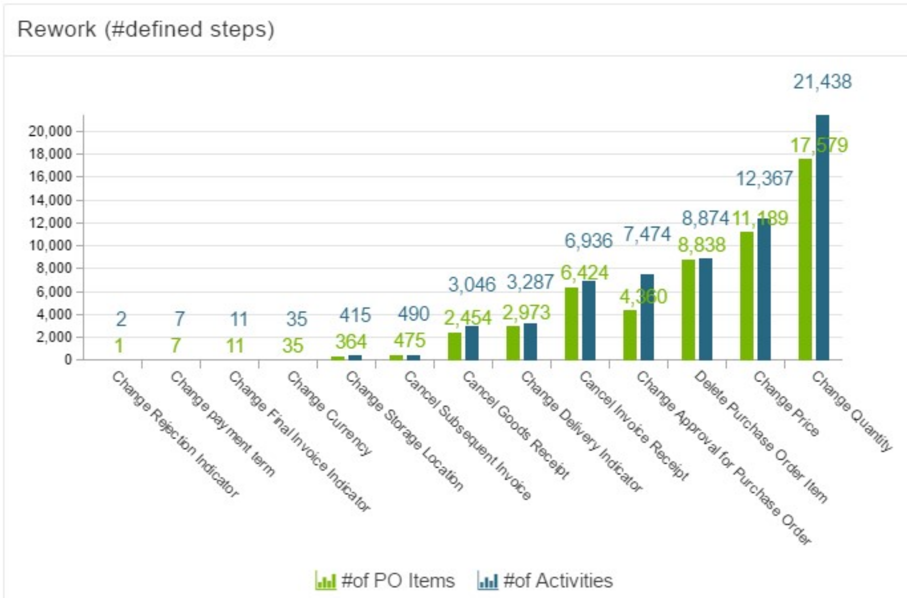

Fig. 7. Number of rework events and cases per rework activity name

was additionally used to indicate whether a particular combination of features during process execution has a higher chance to lead to incompliant behavior in a case.

The primary goals of the implemented classifier encompassed the understandability to learn about the influencing factors on compliance and the ability to find a large share of the incompliant cases with a low amount of false predictions. A decision tree algorithm was the classification technique of choice due to the inherent possibility of visualization and the manageability of simplicity. However, with an increasing number of depth and leaves, decision trees grow more complex, so the achieved prediction correctness was balanced out with the complexity of the decision tree. Analyzing an imbalanced data set with 99.6% compliant PO items, accuracy as the measure of correct predictions was not an insightful model quality measurement. 99.6% of accuracy could be reached by only guessing all PO items to be compliant. Therefore, recall as the ratio of correctly predicted incompliant cases of all incompliant cases and precision as the ratio of the false positives from all positive predictions, i.e., the proportion of actually incompliant cases from all predicted incompliant cases, were used as the key measures of the model. As the harmonic mean between recall and precision, the F1-score was additionally used as a classifier quality indicator.

In the preprocessing of the decision tree classifier, all relevant data was imported into Python and the categorical variables were binarized. The deduced values of the GRs and invoice receipts, as well as their receipt quantities, were included in neither case level nor event level prediction. The exclusion was due to the reason that factors leading to these values were sought after, and the values could nevertheless be calculated directly. In the following, all used data columns are presented.

Categorical attributes:

- _case_Document_Type_
- _case_Item_Category_
- _case_Spend_classification_text_
- _case_Item_Type_
- _case_Sub_spend_area_text_
- Process cluster

Numerical attributes:

- Presence of events per activity name
- Average resource workload per activity type
- Maximum resource workload per activity type
- Number of handovers
- Number of rework events
- Number of sub-spend areas in the superordinate PO
- Missing material flag
- SOD: PO created and GR recorded by the same user
- SOD: PO created and invoice recorded by the same user
- Create Order net value
- Flag to indicate whether it is a retrospective PO item
- Total throughput time in days

The data was fitted to a balanced decision tree classifier, which was created by splitting the leaves by the best possible *gini* information gain optimization. With four layers of depth in the decision tree, satisfactory results in recall could be achieved at an acceptable level of complexity. Applying 5-fold cross-validation to the dataset and classifier, a mean recall of 88.40% with a standard deviation of 3.34 percent points was measured. For the same dataset and classifier, the precision amounted to 5.06% with a standard deviation of 1.61 percent points. A mean weighted F1-score of 96.11% with a standard deviation of 1.24 percent points was measured with the same parameters. These numbers indicate that a high majority of incompliant PO items are detected at the expense of low precision. Therefore, applying this decision tree's learnings could mean that almost all incompliant cases could be found without knowing the monetary values of goods and invoice receipts. However, only ca. one in every twenty predicted incompliant PO items would actually be incompliant. Since a general suspicion of all PO item with a more in-depth review of all PO items would result in the precision of 0.37%, the precision could still be improved with the findings. In this paper, the concentration was set towards achieving a high recall, but another approach for another direction may be to maximize precision and thereby find incompliant PO items with as little effort as possible.

The four-layer decision tree that was generated is presented in figure 12 and 13 in the appendix. It describes the combination of features that indicate whether a PO item has a higher chance to be compliant or incompliant. Despite the relatively high simplicity of the decision tree, it is still difficult to understand, why particular splits were taken and which features influence the decision tree in which way. Therefore, the inspected features were further analyzed on their importance on the classification. In figure 8, the importance values for the encoded features of the decision tree with the importance of more than 0.09% are listed. The term *encoded* here refers to the binarization for the categorical variables.

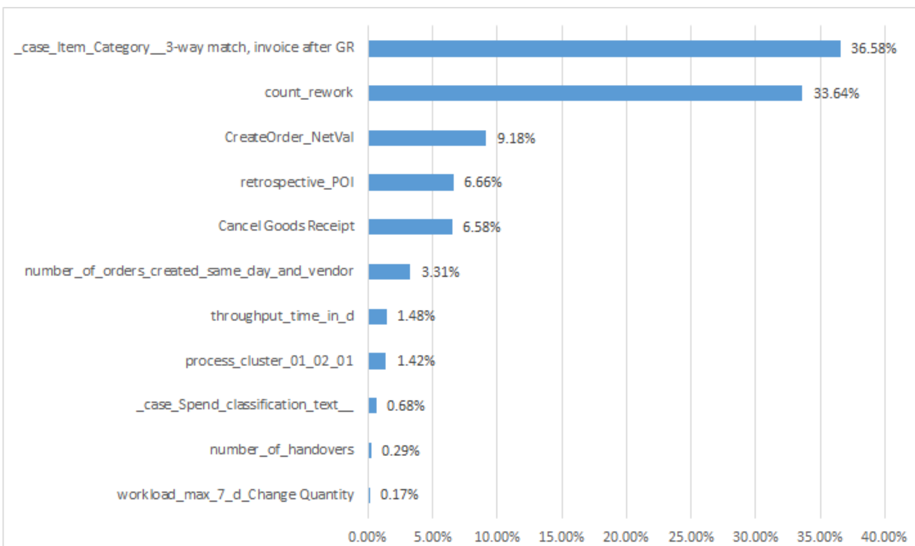

Fig. 8. Feature importance values for case level decision tree

In this step, the contributing features to compliance were known, and the decision tree indicated how a particular feature contributes to the compliance prediction. However, the question was left open, whether certain feature values increase or decrease the chance of compliance. Hence, for the attributes that the decision tree algorithm indicated as relevant, the distribution of compliant and incompliant cases was calculated. The change of proportion of incompliant PO items of all PO items filtered on a specific value range of the feature was calculated to see if the value range increases or decreases the proportion of incompliant cases. The ratios were compared to the ratio of incompliant cases within the total distribution amounting to ca. 0.37%. For categorical variables, all possible categories were taken as inputs, and for the numerical values, the ranges proposed by the decision tree were used. As an example, 20 highest values of difference to the incompliance ratio in the total population are presented in table 5.

column	value	imcompl. ratio	change to total incompl.
_case_Sub_spend_area_text_	Road Packed	40.69%	+10,978%
Cancel Goods Receipt	>0.5	11.82%	+3,119%
_case_Sub_spend_area_text_	Marketing Support Services	10.00%	+2,622%
_case_Sub_spend_area_text_	Sea	9.09%	+2,375%
_case_Sub_spend_area_text_	Government payments	9.09%	+2,375%
_case_Sub_spend_area_text_	Third Party Labor	6.25%	+1,601%
_case_Sub_spend_area_text_	Polyurethane Resins	6.06%	+1,550%
_case_Item_Category_	3-way match, invoice after GR	4.36%	+1,088%
_case_Sub_spend_area_text_	Packaging	4.18%	+1,039%
_case_Sub_spend_area_text_	Other Logistics Services	4.11%	+1,019%
_case_Sub_spend_area_text_	Digital Marketing	3.85%	+947%
_case_Sub_spend_area_text_	HR Services	3.39%	+823%
_case_Sub_spend_area_text_	Commercial Printing	2.79%	+660%
_case_Spend_classification_text_	OTHER	2.02%	+450%
_case_Sub_spend_area_text_	Waxes	1.73%	+370%
_case_Sub_spend_area_text_	Customers	1.68%	+358%
retrospective_POI	>0.5	1.33%	+262%
count_rework	>0.5	1.27%	+245%
CreateOrder_NetVal	>12729.5	1.23%	+234%
_case_Sub_spend_area_text_	Design	1.20%	+228%

Table 5. Feature values with the 20 highest incompliance ratios

In these 20 value distributions, some strong influences of feature values on compliance are already observable. Cancellations of goods receipt strongly increase the chance of impliance as well as when a PO item is entered retrospectively and with a high net value. If rework activities are registered, the chance for incompliance also more than doubles. Multiple sub-spend areas show a higher than average ratio of incompliant cases, so a focused review of these sub-spend areas may support the company in reducing incompliant cases.

Analyzing only these 20 highest feature values intervals or values with their incompliance may already provide the company with a basis for a guided improvement of their incompliance. The mentioned sub-spend areas pose a higher risk for incompliance and their processes can be observed more closely or a stricter approval process could be introduced for the related goods. Cancellations of goods receipts not only create additional effort, but also have a stronger tendency to turn out incompliant, so a filter of these PO items from a purchase manager could reveal the incompliant cases, before they are completed or clear open questions. The fact that higher net order PO items are affected with a higher impliance ratio even increases the issue of incompliance for these

PO items, so a more frequent review of these PO items may be used to reduce incompliance.

5.4 Compliance classification on event level

Although the decision tree on case level may enhance the understanding of the factors leading to incompliant PO items, it would have only limited applicability in a running system. This is because it uses all information that is only known after a completed purchasing process. Therefore, a further decision tree was implemented on the event level using only a pre-mortem dataset with only the information that is known up to the timestamp of a given event. A classification into compliant and incompliant cases should then be given for this particular case at the time of the given event. As the indicators recall and precision are expected to depend on the number of events in the case that is classified, not every event is expected to provide enough information for accurate predictions. The goal of the development of the classifier is the verification of the underlying concept, and the classifier may benefit from further information fed to it. The primary use cases aimed for with this concept is that warnings could be given to process managers when the probability of incompliance of a given case surpasses a defined threshold. The potential independent features comprise of the case level attributes known after the PO item creation, the presence for each activity as well as all additionally created dimensions. All of this information is restricted to the data that was recorded before or at the given timestamp and event-ID within the case. In the designed concept, not only the current information of a specific event is taken into account, but also the history up to this event. The categorical features were the same as for the case level detection, but the numerical features changed and are listed below:

- Missing material flag
- Retrospective PO item flag
- Count of rework activities
- CreateOrder_NetVal (PO item worth)
- Number of sub-spend areas in the superordinate PO
- SOD: PO created and GR recorded by the same user
- SOD: PO created and invoice recorded by the same user
- Total throughput time in days
- Task load past two days / seven days

Building, similarly to the case level decision tree, a four-layer model, similar features were used by the model to predict compliance. The most influential attributes were the *_case_item_category_*, the process cluster and the net value of the PO item. While these attributes are known from the case level and already known during PO item creation, also features on event level, like the throughput time in days, the task load and the presence of activities such as *Record Goods Receipt*, *Cancel Goods Receipt* and *Change Delivery Indicator* are used.

Applying 5-fold cross validation on the classifier, a recall of 79.47% with 4.16 percent points standard deviation could be achieved.The precision amounted to

8.50% with a standard deviation of 1.86 percent points, and the mean weighted F1-score was 96.05% with a standard deviation of 1.20%. These values indicate that the risk of breaching compliance of a PO item can be predicted with such a decision tree. These findings support the gained understanding from the case level decision tree that some attribute values are related with a higher risk of incompliance. The created decision tree is shown in figures 14 and 15 in the appendix.

The pre-mortem predictive incompliance detection could be used in the process owner's indication to warn the respective employees of the high risk of incompliance in a given PO item. Such a warning system could reduce the incompliance by raising awareness and concentration in the right moments. As a more severe consequence, the purchase managers could be also directed to the high risk PO items for a general review.

6 Discussion and future work

In this paper, the purchase-to-pay process could be described by multiple process models designed for different process areas given by a group of PO item attributes. Additionally, multiple calculated indicators could be created that may increase the understanding of the process and may support in the advisory of its quality. Among these measures were the compliance of the process and the throughput times between GRs, invoices, and payments. Consequently, the compliance influencing factors could be determined with the type of their influence and the strength of their impact. The predictive process mining techniques applied for this also were used on a pre-mortem dataset to verify the possibility of a compliance warning method in a running system.

The analysis of the influences was not exhaustive, and another research into the complementary forces between attributes could be a field of study in the future. However, the simplification of the decision tree findings may be applied similarly to other process mining analyses in the future and thereby constitute one good practice of interpretation. While also the process model creation may be abstracted to a general methodology, the creation of additional dimensions and calculations between the given variables and events is a unique result to this project. In this, it is specific to a purchasing process and even the given data columns. Therefore, the findings of this paper divide into the followed analysis steps and methodology as a potential good practice for the community of process mining scientists and practitioners and into the dimensions and findings that affect the particular company and purchasing analysts.

The designed classifiers to predict compliance can be used in the process owner's organization to reduce incompliance in a guided approach by determining the influential factors beforehand. With the pre-mortem decision tree, a continuous monitoring of the compliance and a shift of factors may be observed. A warning system or an introduced focused review system of PO items with higher incompliance risk may be introduced on basis of the implemented classifiers. However, if additional factors potentially influencing the compliance can be assumed in the future, these would first need to be implemented as additional dimensions.

7 Conclusion

Structured along with the BPI challenge requests, the contributions of this paper can be grouped into three parts.

First, a group of process models was created, representing the as-is conduct of purchasing within the process owner's organization. An analysis was conducted into the case attributes that may be used to split the data and which features show the highest within sum-of-squares similarity.

In the set-up of a throughput time calculation method, the rules for obtaining throughput times between GRs, invoice receipts, and payments even with multiple iterations of these documents were developed. The throughput times between invoice receipt and payment mainly depends on the payments terms within the organization. Furthermore, the time between GR and invoice receipt is both dependent on the supplier's invoice delivery as well as the organization's invoice recording speed. As data for both payment terms as well as invoice arrival time were not available, no further analysis was performed into the throughput times.

In a third step, the purchasing compliance of the organization was analyzed by creating additional dimensions combined with the given dataset to feed a decision tree classifier. With this classifier and subsequent analyses, the influencing factors to purchasing compliance could be identified. The concept of a compliance warning decision tree in a running system was explored using pre-mortem perspectives of the data. By obtaining results, with which more incompliant cases could be detected early in the process, this concept was found to be valid to reason further research and company implementations.

While therefore the findings of this paper may support the process owner in their process understanding and compliance transparency, the applied methodology and the proved concept into predictive compliance detection with process mining may lay the ground for new scientific advancements in this direction.

References

1. Business process model & notation™ (bpmn™), https://www.omg.org/bpmn/
2. Hompes, B.F., Buijs, J., van der Aalst, W.M., Dixit, P.M., Buurman, J.: Discovering deviating cases and process variants using trace clustering. In: Proceedings of the 27th Benelux Conference on Artificial Intelligence (BNAIC), November. pp. 5–6 (2015)
3. van der Aalst, W.: Process mining: discovery, conformance and enhancement of business processes, vol. 2. Springer (2011)
4. van der Aalst, W.: Data science in action. In: Process Mining. Springer (2016)
5. van Dongen, B.F.: Bpi challenge 2019. https://doi.org/10.4121/UUID:D06AFF4B-79F0-45E6-8EC8-E19730C248F1

8 Appendix

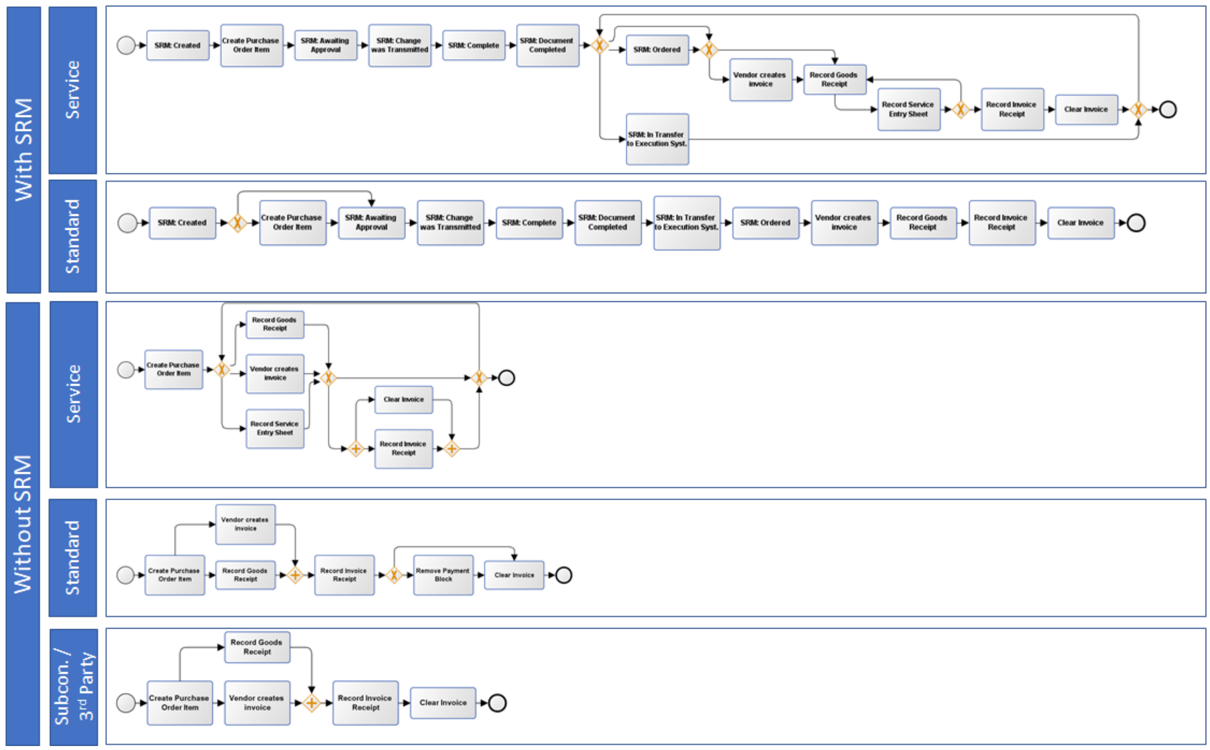

Fig. 9. BPMN models for *3-way match: Invoice after GR* order items

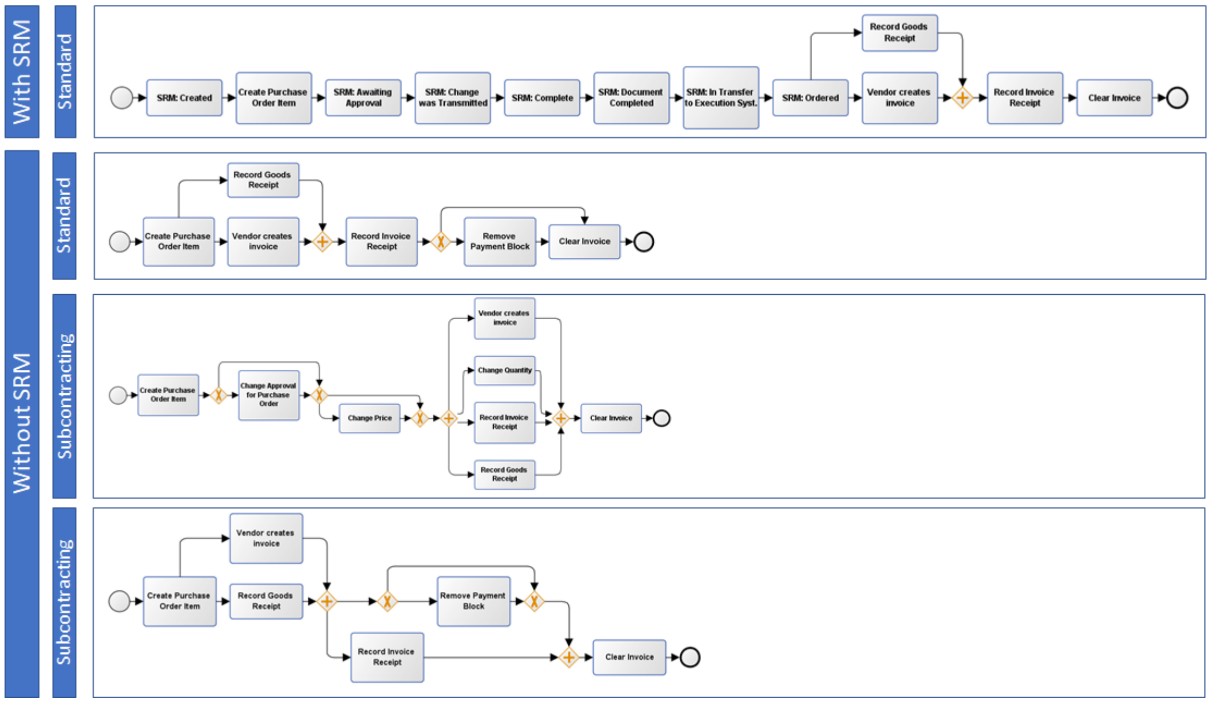

Fig. 10. BPMN models for *3-way match: Invoice before GR* order items

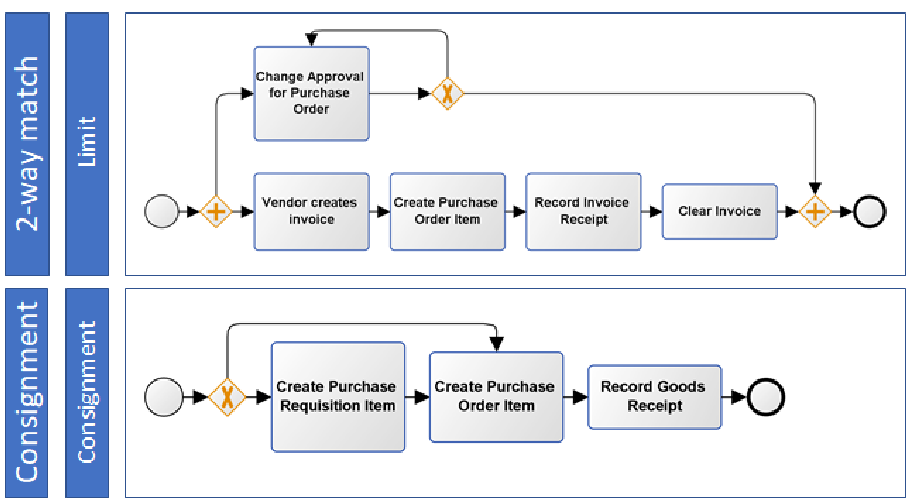

Fig. 11. BPMN models for *2-way match* and *Consignment* order items

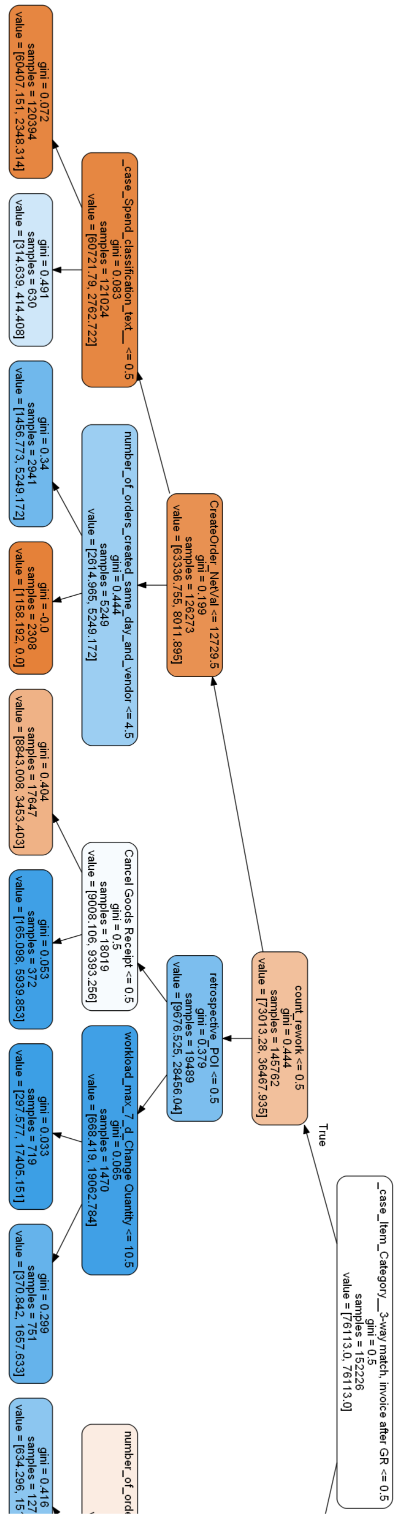

Fig. 12. Four layer case level decision tree [Left part]

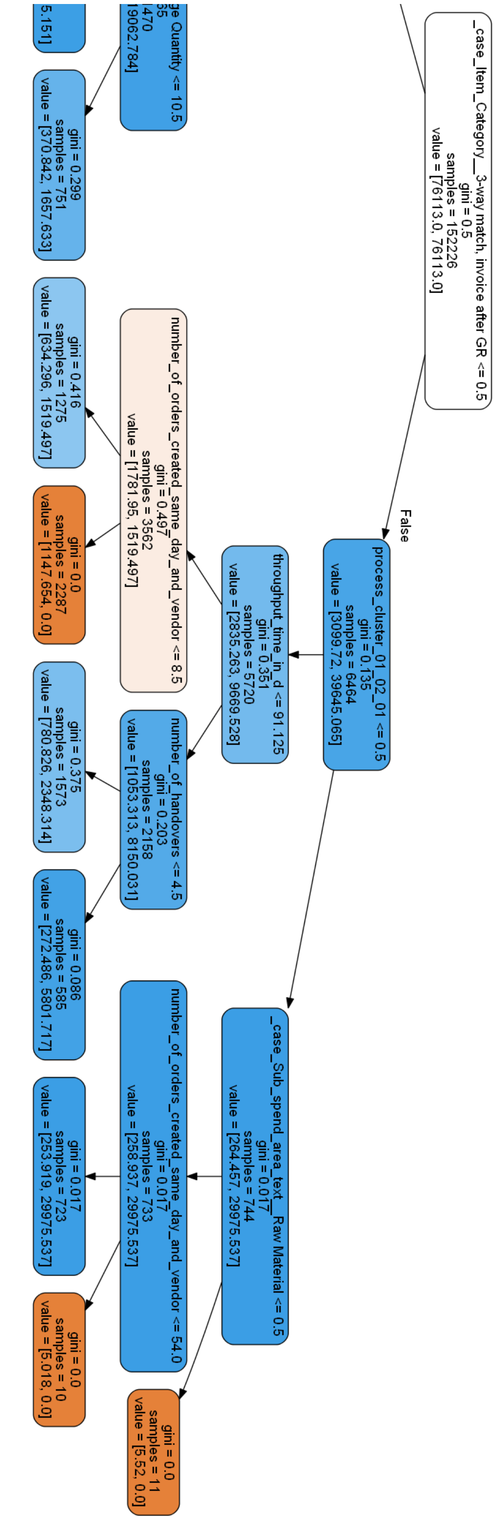

Fig. 13. Four layer case level decision tree [Right part]

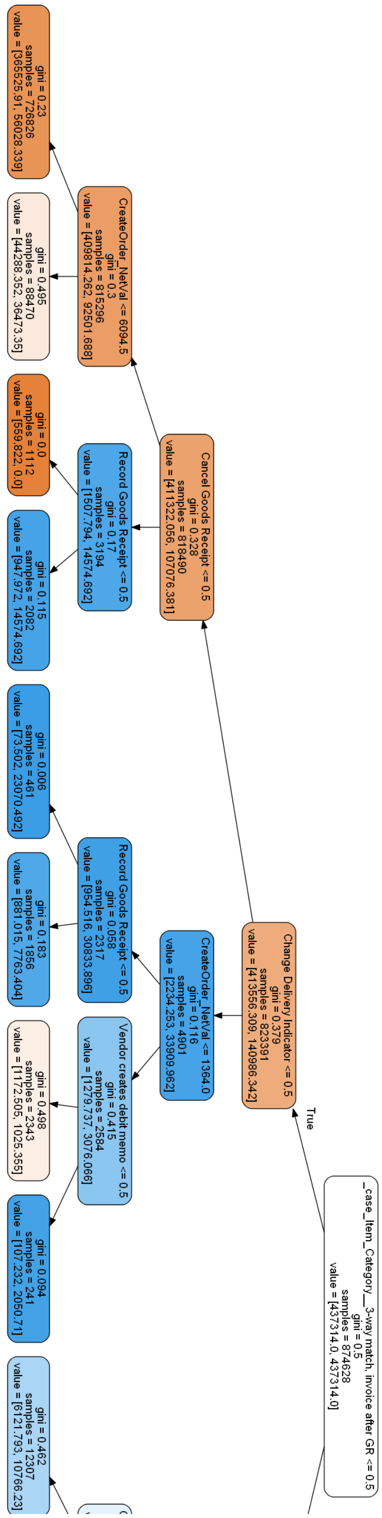

Fig. 14. Four layer event level decision tree [Left part]

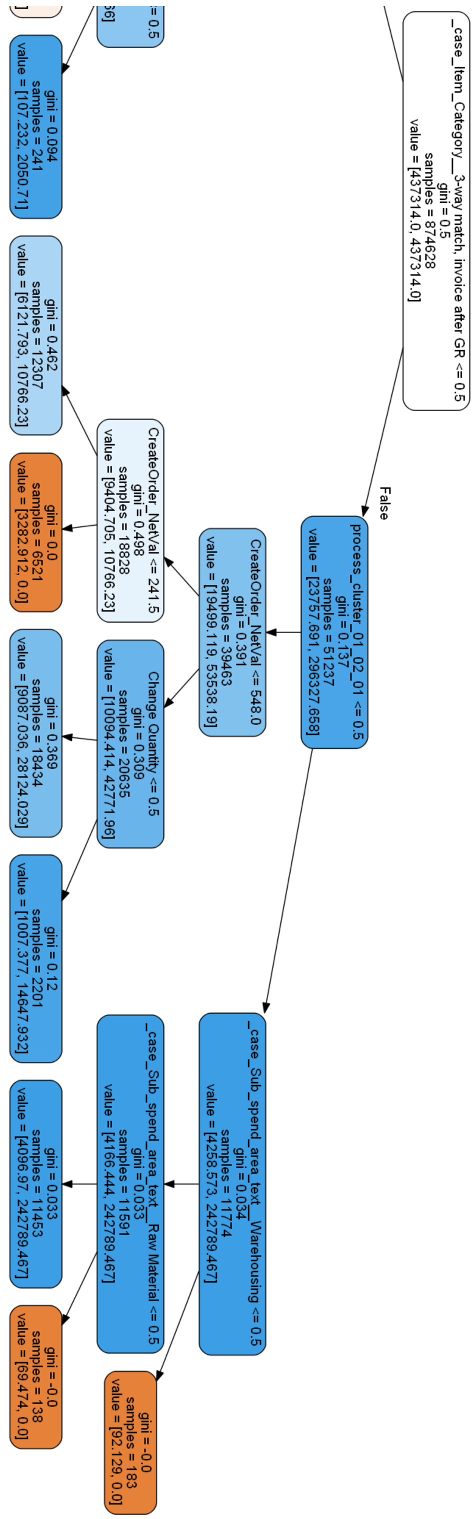

Fig. 15. Four layer event level decision tree [Right part]