

ENGINEERING DEPARTMENT

MASTER'S DEGREE IN ARTIFICIAL INTELLIGENCE AND DATA ENGINEERING



UNIVERSITA DEGLI STUDI DI PISA

ACADEMIC YEAR 2022-2023

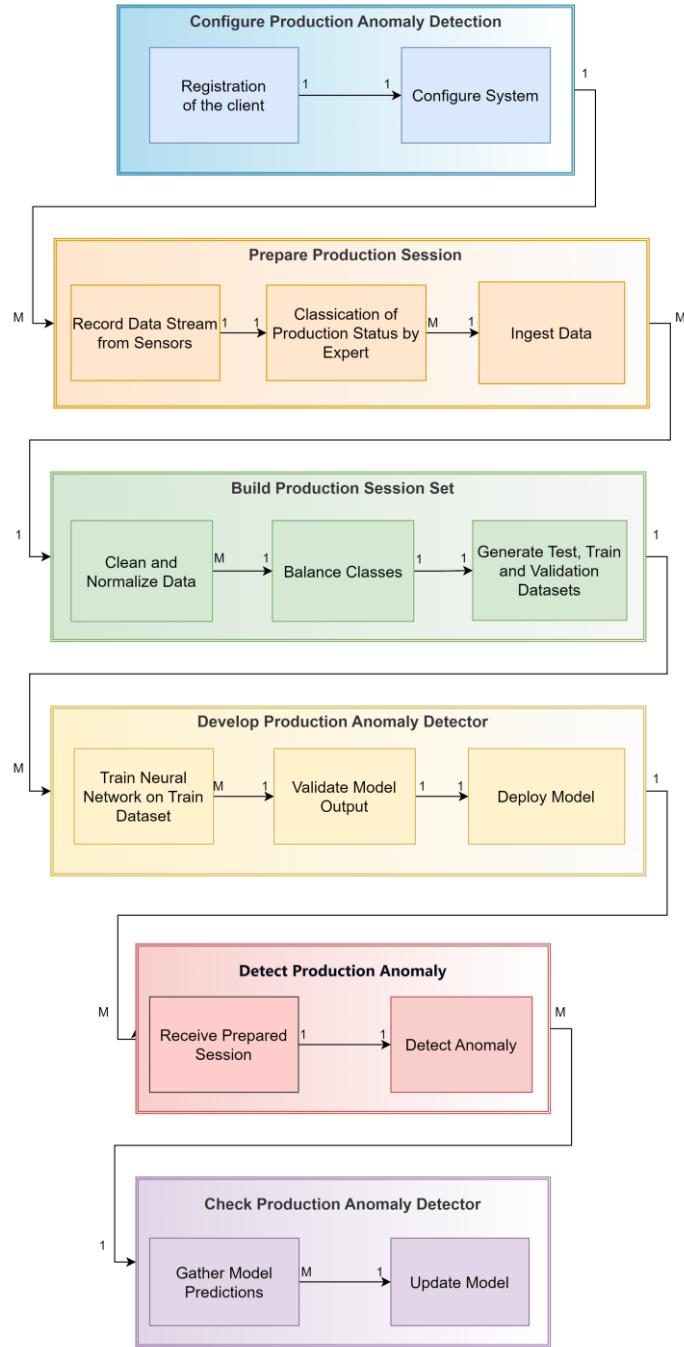
PROCESS MINING AND INTELLIGENCE

ANOMALY DETECTION IN MANUFACTURING

*Fabio Cognata, Giovanni Paolini, Giulio Fischietti, Luca Di Giacomo*



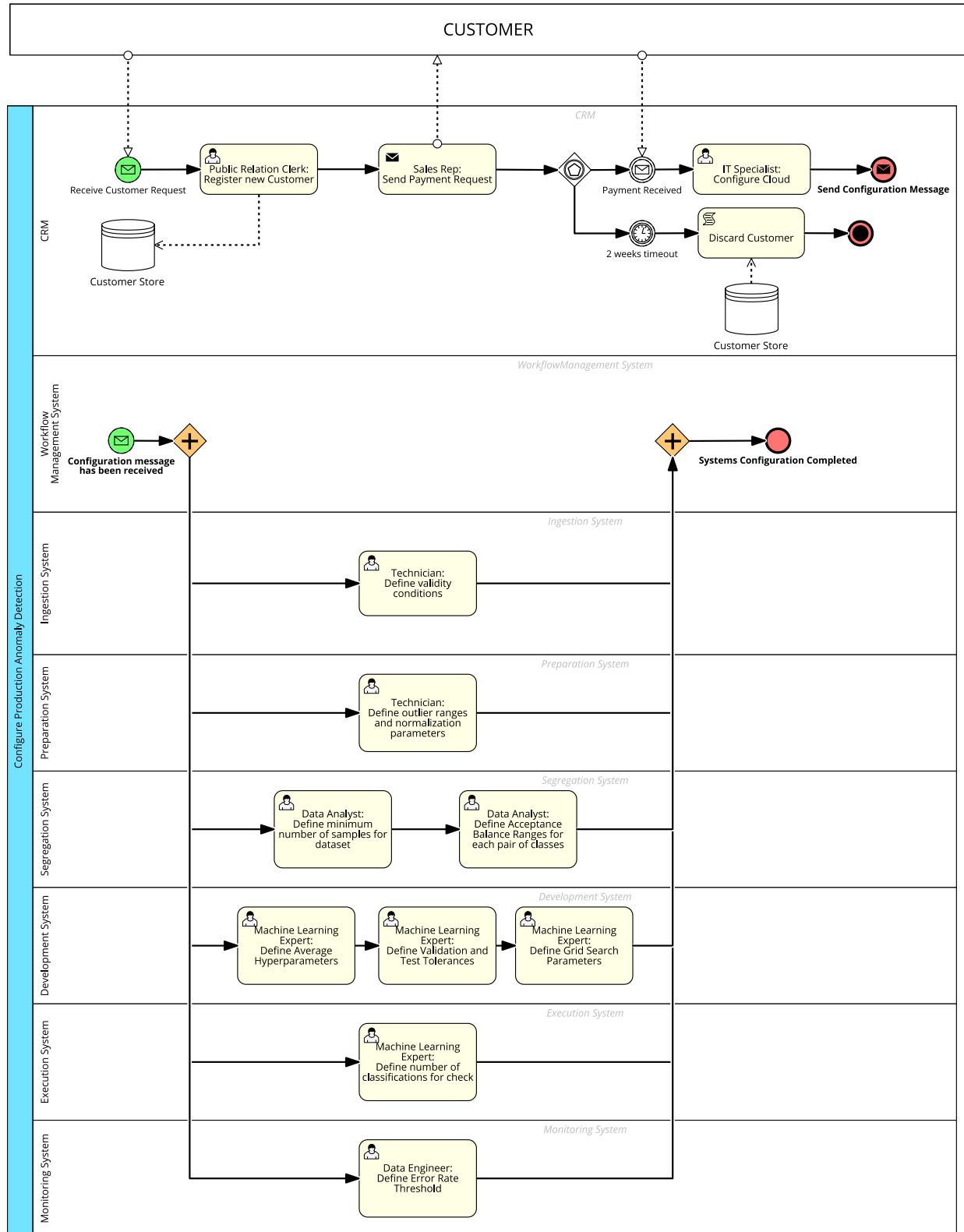
## PROCESS LANDSCAPE (All)



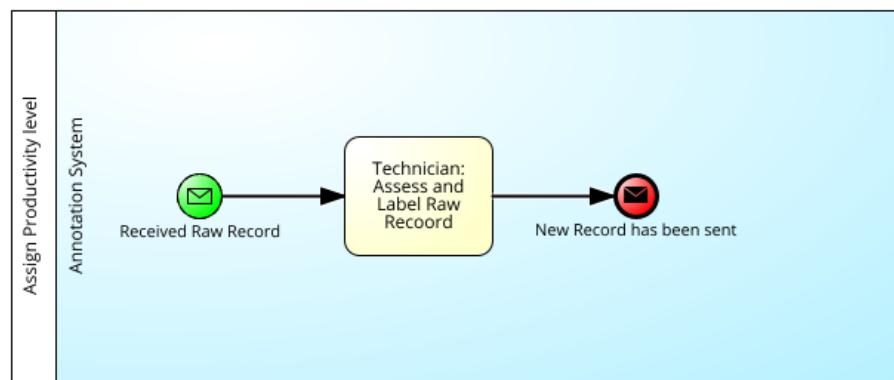
# BPMN Diagrams

## BPMN Diagram for the Configuration Process

(All)



BPMN Diagram for Assign productivity Label

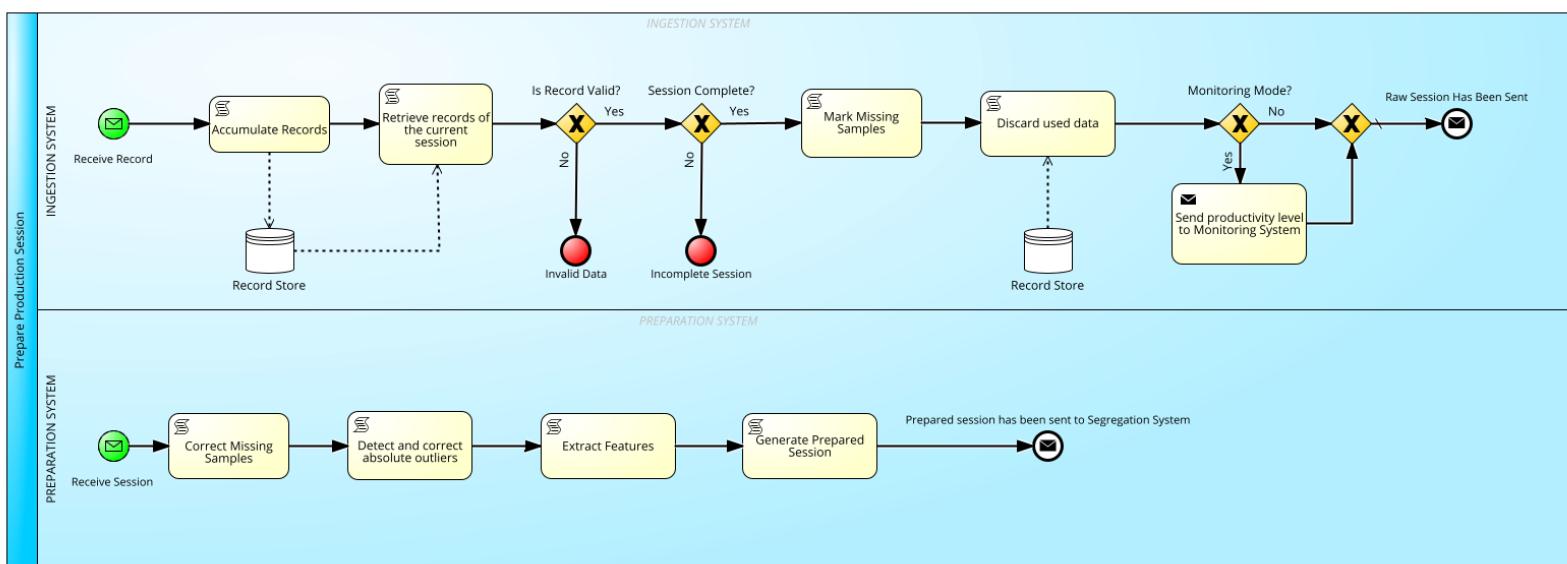


## BPMN Diagram for the Prepare Production Session

(Giovanni)

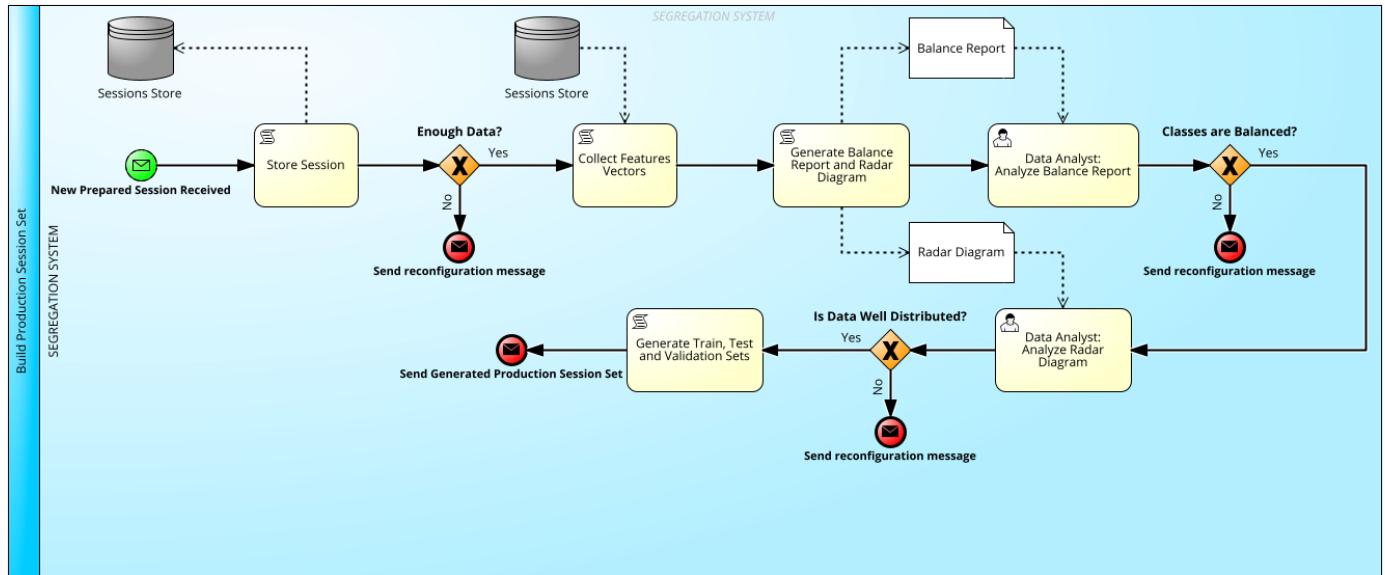
The records received are of the following types:

- Environmental**: Temperature, Humidity, number of scrolls produced
- Annotation System**: Productivity level (Green, Yellow, Red)
- (HRM System) Human Resources**: Clerk Experience (1-30)
- (ERP System) Materials**: Paper quality (Low, Medium, High), machinery lifetime



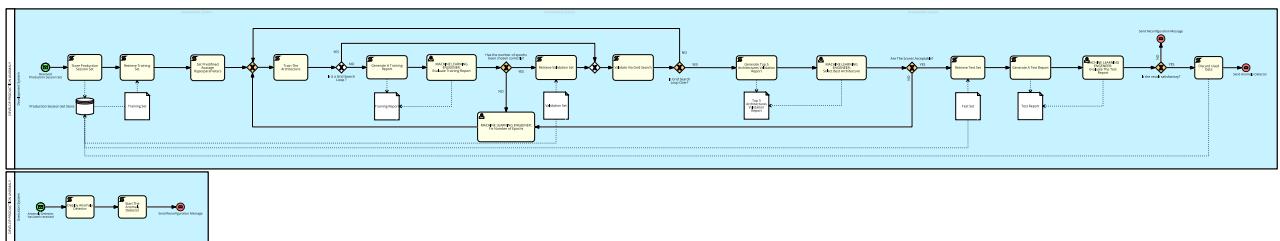
## BPMN Diagram for Build Production Session Set

(Luca)



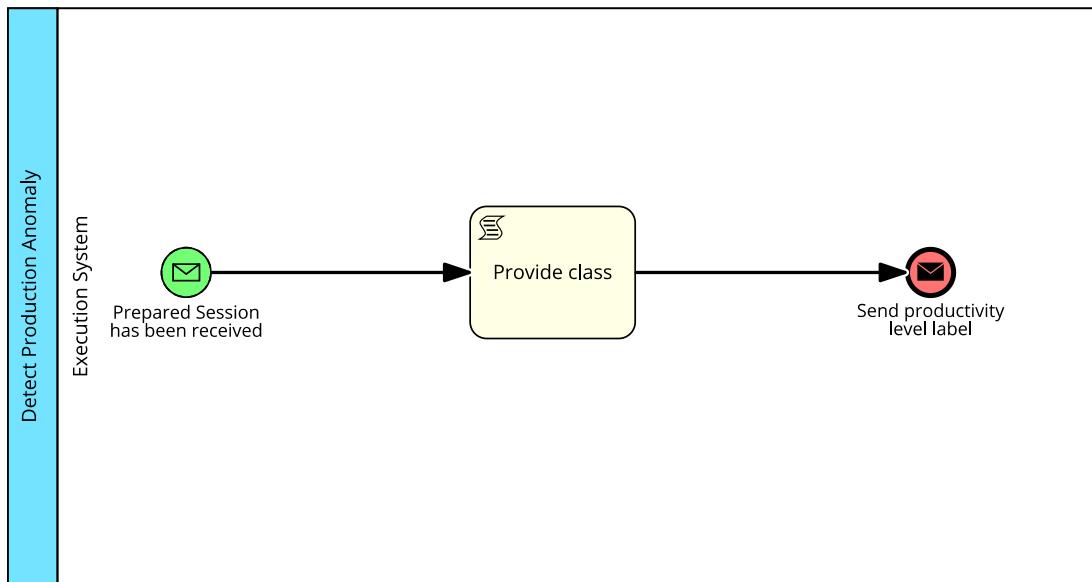
## BPMN Diagram for Develop Production Anomaly

(Fabio)



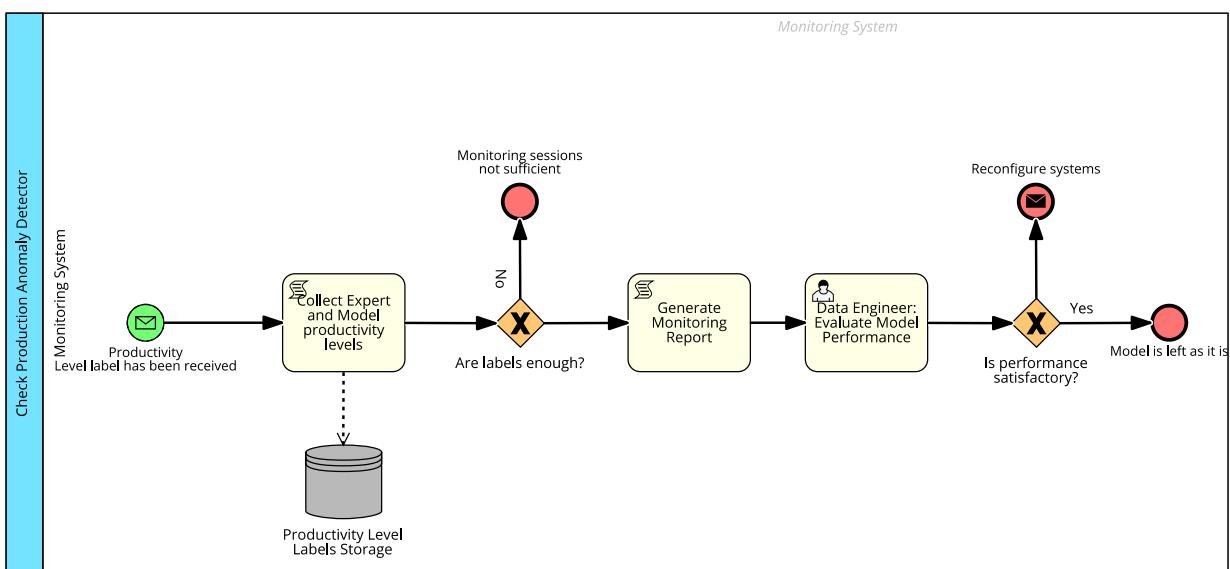
### BPMN Diagram for Detect Production Anomaly

(Giulio)



### BPMN Diagram for Check Production Anomaly Detector

(Giulio)





## Use cases

### Cost Table (Normalized)

Actor	Source	Cost (Hourly)	Normalized Cost
<b>Data Analyst</b>	<a href="#">Stipendio per Data analyst (Italia)</a>	€ 30,28	2,73
<b>IT Specialist</b>	<a href="#">Stipendio per IT specialist (Italia)</a>	€ 31,18	2,82
<b>Data Engineer</b>	<a href="#">Stipendio per Ingegnere (Italia)</a>	€ 28,91	2,61
<b>Technician</b>	<a href="#">Stipendio per Tecnico (Italia)</a>	€ 16,03	1,45
<b>Public Relations Clerk</b>	<a href="#">Stipendio per Pubbliche relazioni (Italia)</a>	€ 11,07	1
<b>Machine Learning Engineer</b>	<a href="#">Stipendio per Machine Learning e Data Scientist (Italia)</a>	€ 27,86	2,52

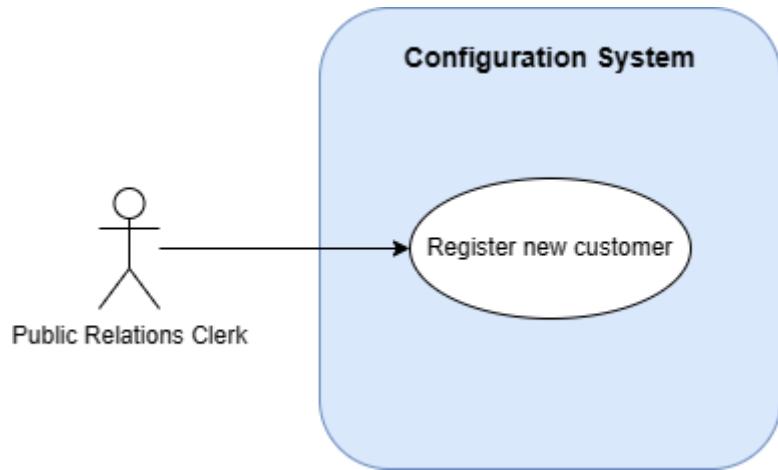
The cost for a step is calculated with the following formula:

$$Step Cost = (CE \times EP) \times AC$$

Where:

- CE is the Cognitive Effort
- EP is the Execution Probability
- AC is the hourly Actor Cost normalized

CRM: Register New Customer  
(Giovanni)



Register New Customer: Mockup Interfaces

The image displays two user interface mockups. The top one is titled 'Client Registration Form' and contains fields for 'Company Name' (text input), 'Annual Paper Volume' (spin box with value 0), and 'Nº of Paper Variety' (spin box with value 0). A 'Register' button is at the bottom right. The bottom one is titled 'Received Registration Requests' and shows a table of registered companies with columns for 'Approve' (checkbox), 'Company' (name), 'Volume' (value), and 'Variety' (value). The table includes rows for 'Paper&co', 'Cartiere Burgo', 'Cartiere Miliani', and 'Industria Cartaria Pieretti'. A 'Confirm' button is at the bottom right.

	Approve	Company	Volume	Variety
▶	<input checked="" type="checkbox"/>	Paper&co	100k	3
	<input type="checkbox"/>	Cartiere Burgo	50k	2
	<input checked="" type="checkbox"/>	Cartiere Miliani	250k	2
	<input checked="" type="checkbox"/>	Industria Cartaria Pieretti	150k	6
*	<input type="checkbox"/>			

Register new customer

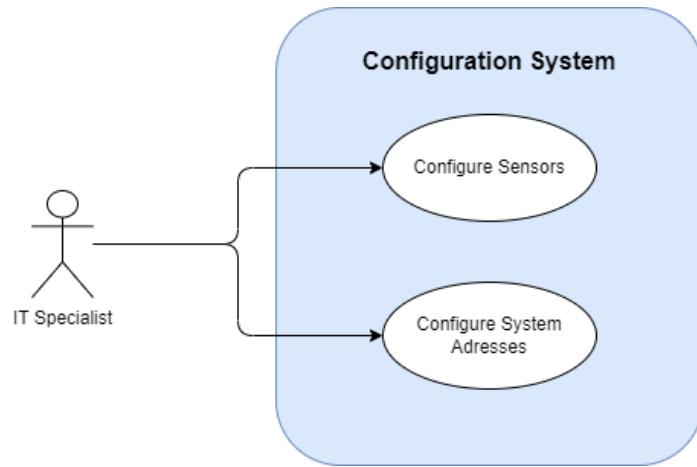
1. PR Clerk opens received registration requests
2. IF: The request is valid (80%)
  - 2.1. PR Clerk inserts the registration in the Database

Register new customer: Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	PR Clerk	Open Registration Requests	Remember (1)	1	1
2 - IF		Registration Request is valid			
2.1	PR Clerk	Accept and Insert Registration	Understand (2)	0,8	1,6
					<b>Total Cost</b>
					2,6

## CRM: IT Specialist Configure Cloud

(Giovanni)



### Configure sensors

1. IT Specialist identifies the used sensors
2. IF: Sensor type handler is not already present in the cloud system (20%)
  - 2.1. IT Specialists create handling script for the new sensor
3. IT Specialist configures the cloud connection to the sensors
4. IT Specialist creates a DB table model for the required data

### Configure System Addresses

1. IT Specialist identifies the needed systems for the configuration
2. IF: Required system doesn't have a static address in the network (30%)
  - 2.1. IT specialist assigns an address to the system
3. IT Specialist creates a lookup table with the address of each system
4. IT Specialist initialize a static address to the lookup table to be forwarded to each system

## Configure Cloud: Cost Table

### Configure Sensors

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	IT Specialist	Identify Sensors	Understand (2)	1	5,64
2 - IF		Handler not present			

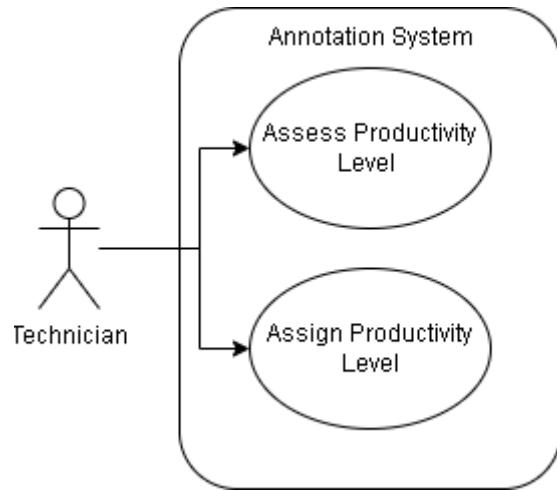
2.1	IT Specialist	Add new Handler	Remember (1) Paste the correct settings from the sensor's manufacturer config	0,2	0,56
3	IT Specialist	Configure Cloud Connection	Remember (1) Configure the connection to the cloud changing an already existing archetype	1	2,82
4	IT Specialist	Create mock DB table model	Remember (1) Create a mock DB table by adding eventual new data columns to an already existing archetype	1	2,82
				<b>Total Cost</b>	11,84

### Configure System Addresses

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	IT Specialist	Identify Systems	Understand (2)	1	5,64
2 - IF		Static Address not present			
2.1	IT Specialist	Add new static address	Remember (1) Copy and paste of Addresses	0,3	0,84
3	IT Specialist	Create Lookup Table	Remember (1) Copy and paste of Addresses	1	2,82
4	IT Specialist	Assign Lookup table Address	Remember (1) Copy and paste of Addresses	1	2,82
				<b>Total Cost</b>	12,12

## Assign productivity Label

(Fabio)

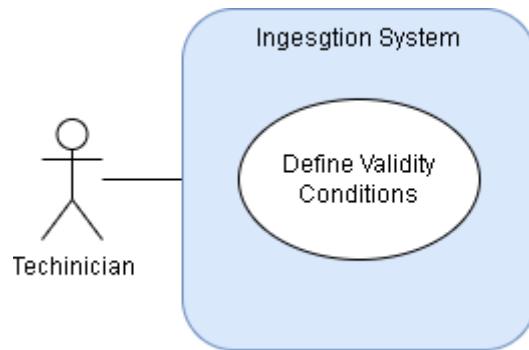


Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	Technician	Open Annotation Interface	Remember (1)	1	1,45
2	Technician	Asses Productivity Level	Analyze (4) Multicriteria: Given all the information such as Humidity, Temperature, Number of Produced Rolls, Machinery lifetimes, paper quality and clerk experience Evaluate the level of performance	1	4,35
3	Technician	Assign Productivity Level	Remember (1)	1	1,45
			<b>Total Cost</b>	7,25	

## Configuration Process: Technician Define Validity Conditions and Define Outlier Ranges

(Giovanni)



### Define Validity Conditions

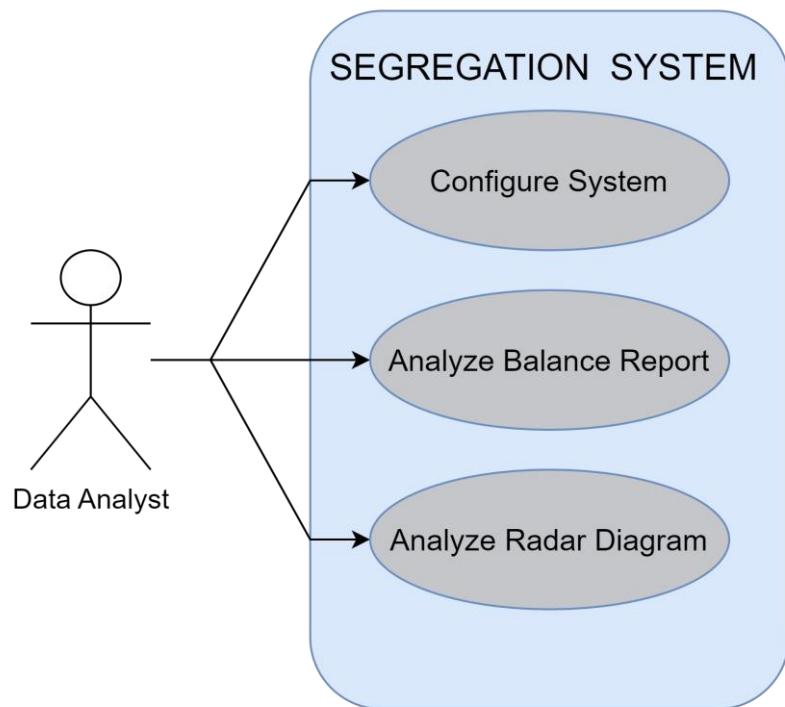
1. Expert technician opens a configuration request
2. Expert technician defines validity ranges for the specified sensors to detect outliers
3. Expert technician defines the average values for missing samples
4. Expert technician defines the average values for outlier correction

Sub-Task	Actor	Step	Cognitive Effort [1-4]	Occurrence	Step Cost
1	Expert Technician	Open Configuration Request	Remember (1) Click the request	1	1,45
2	Expert Technician	Defines validity ranges	Remember (1) Set the ranges according to the specifics from the machinery's manual	1	1,45
3	Expert Technician	Defines AVG values for missing samples	Apply (3) 1)Temperature default value 20° 2)humidity default value 45% 3)Number of rolls produced/min 25	1	4,35

4	Expert Technician	Defines AVG values for outlier correction	Apply (3) 1) If temperature is outside the range (15°,35°) approximate to nearest value in range 2) If humidity is outside the range (0,100) approximate to nearest value in range 3) If the n° of produced rolls/min is outside the range (10,30) approximate to nearest value in range	1	4,35
				<b>Total Cost</b>	11,60

## Segregation System Tasks

(Luca)



### Configure System

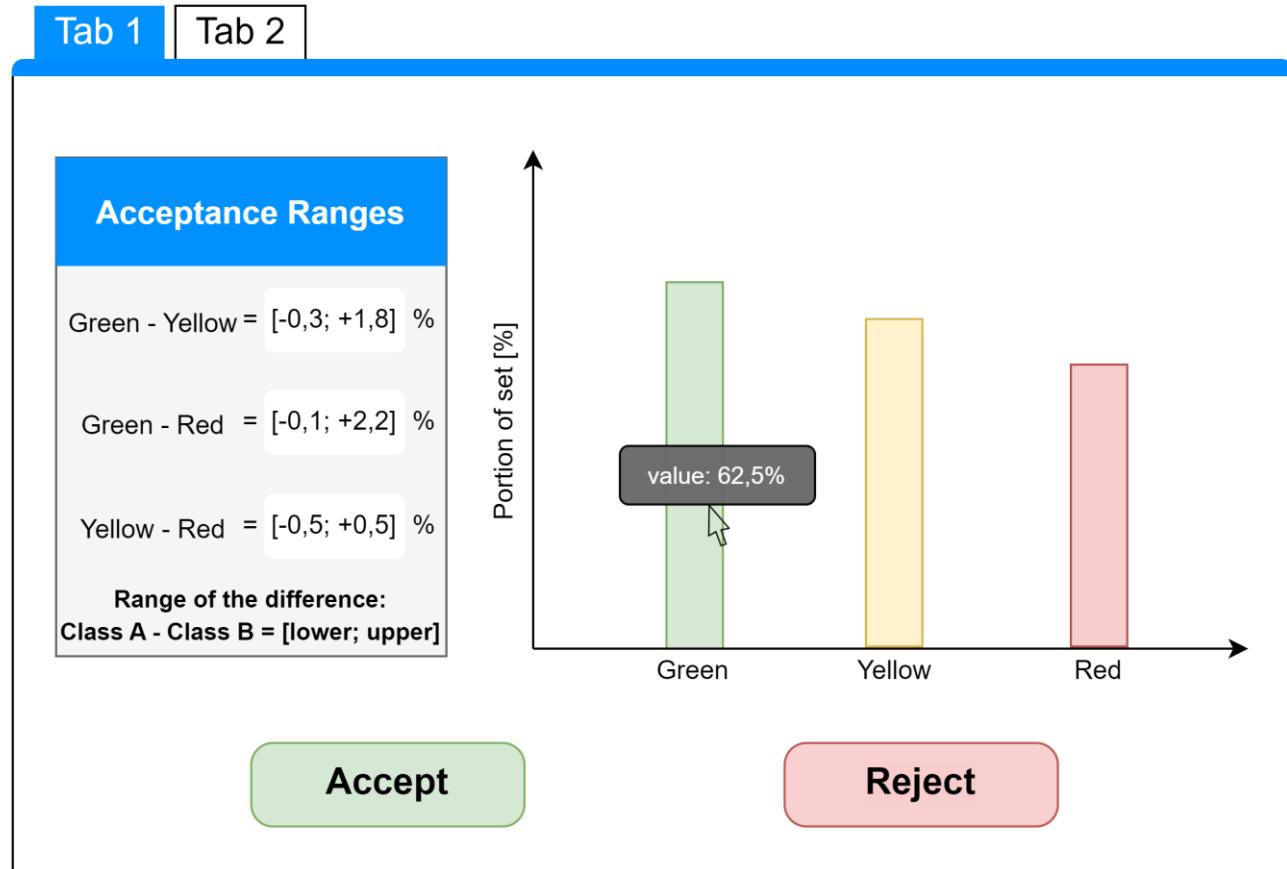
#### Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	Data Analyst	Open Configuration Form	Remember (1)	1	2,73
	System	Displays Configuration Form			
2	Data Analyst	Inserts minimum number of samples for datasets construction	Apply (3) 1)Check the number of features 2)Calculate the lower bound based on the number of features and classes 3)Evaluate eventual adjustments	1	8,19

<b>3 - FOR</b>		For each class pair			
3.1	Data Analyst	Evaluate Acceptance Ranges for the pair	Analyze (4) The ranges must be decided without a clustering defining thresholds for similar cases, basing on the type of distribution of the classes which depends on multi-criterial factors such as the expertise of the employees working with the machinery, the Ambiental factors or Lifetime of the machinery.	3	32,76
				<b>Total Cost</b>	€ 43,68

# Analyze Balance Report

## Balance Report Mockup



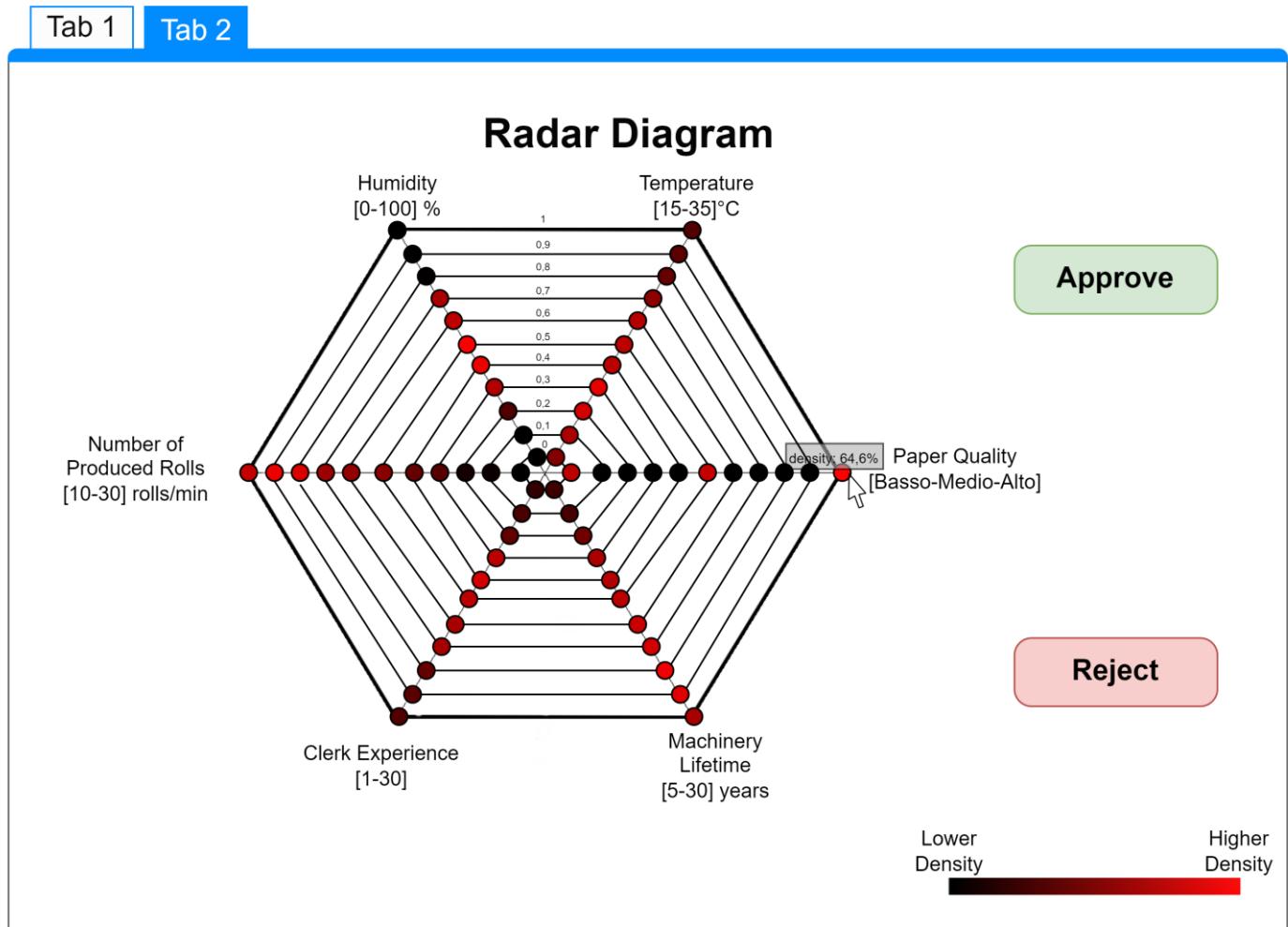
## Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	Data Analyst	Open Balance Report	Remember (1)	1	2,73
	System	Display Balance Report			
2 - FOR		For each class pair in the dataset			
2.1	Data Analyst	Compare class pair balance with configured acceptance ranges	Understand (2) Just verifying classes are in acceptance range	3	16,38

2.1 - IF (20%)		IF classes pair balance is not acceptable			
2.1.1	Data Analyst	Press the “Reject”	Remember (1)	0,6	1,64
	System	Send the reconfiguration message			
2.2 - ELSE (80%)		No Action Required			
3	Data Analyst	Press the “Accept” button	Remember (1)	1	2,73
	System	Displays the radar diagram			
				<b>Total Cost</b>	€ 23,48

# Analyze Radar Diagram

## Radar Diagram Mockup



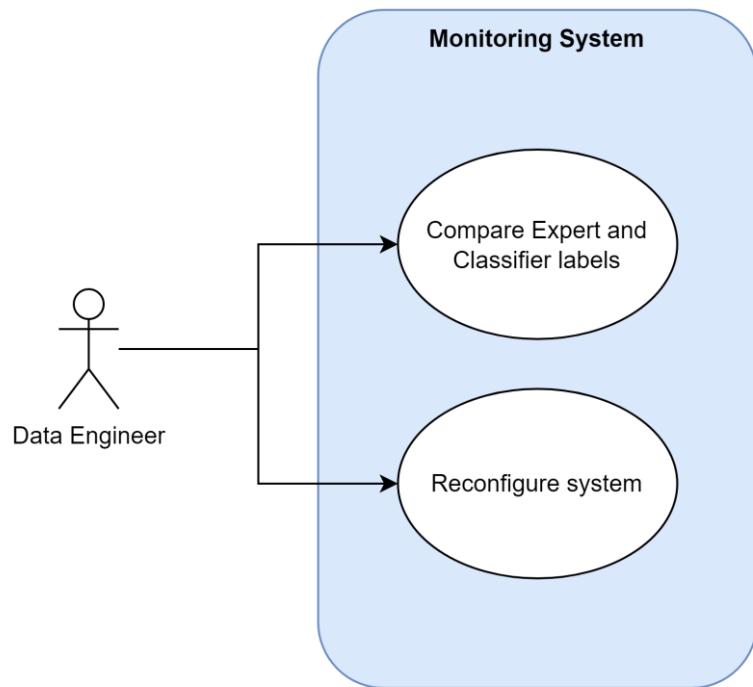
## Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
	System	Displays Radar Diagram			
1 - FOR		For each feature in radar diagram			
1.1	Data Analyst	Evaluate feature distribution	Analyze (4) If the distribution follows the expectations the requirements are satisfied.	3	32,76

			Data distribution depends on factors such as Ambiental factors for humidity and temperature, expertise and machinery lifetime for the number of rolls.		
<b>2 – IF (20%)</b>		If training requirements are not satisfied			
2.1	Data Analyst	Click the “Reject” Button	Remember (1)	0,2	0,55
2.2	System	Send reconfiguration message to workflow management system			
<b>3 – ELSE (80%)</b>					
3.1	Data Analyst	Click the “Accept” Button	Remember (1)	0,8	2,18
3.2	System	Start Generating Production Session Set			
				<b>Total Cost</b>	€ 35,49

# Check Production Anomaly Detector

(Giulio)



## Mockup

Expert and Model Predictions		
Anomaly Detector Report		
Expert Label Class	Detected Class	Result
1	2	✗
2	2	✓
3	2	✗
1	1	✓
2	2	✓

Threshold: 0.2  
Current Error Rate: 0.4  
Error rate is higher than threshold!  
Reconfigure System

## Cost table

Evaluate Model Performance

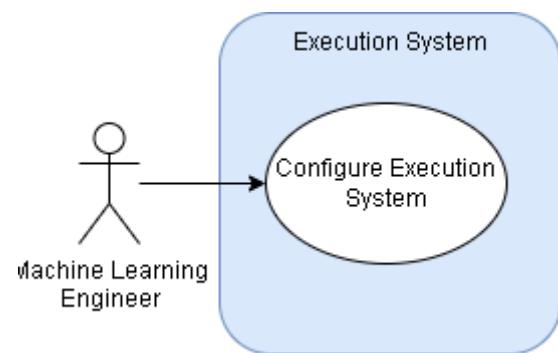
Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Execution Probability	Step Cost
1	Data Engineer	Access Monitoring Application Interface	Remember (1) Click	1	2,61

2	Data Engineer	Open Report	Remember (1) Click	1	2,61
	System	Show Report			
3	Data Engineer	Check if error rate is below threshold	Understand (2)	1	5,22
IF[90%]		If error rate is below threshold			
3.1	Data Engineer	No action is required		0.9	0,00
ELSE [10%]					
3.2	Data Engineer	Click "Reconfigure System" Button	Remember (1)	0.1	0,26
	System	Sends reconfiguration message to management system		0	0
				<b>Total Cost</b>	10,70

#### Configure Monitoring System

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	Data Engineer	Open Configuration Form	Remember (1)	1	2,61
	System	Show Configuration Form			
2	Data Engineer	Define Error Threshold	Remember (1)	1	2,61

	<b>Total Cost</b>	5,22
--	-------------------	------

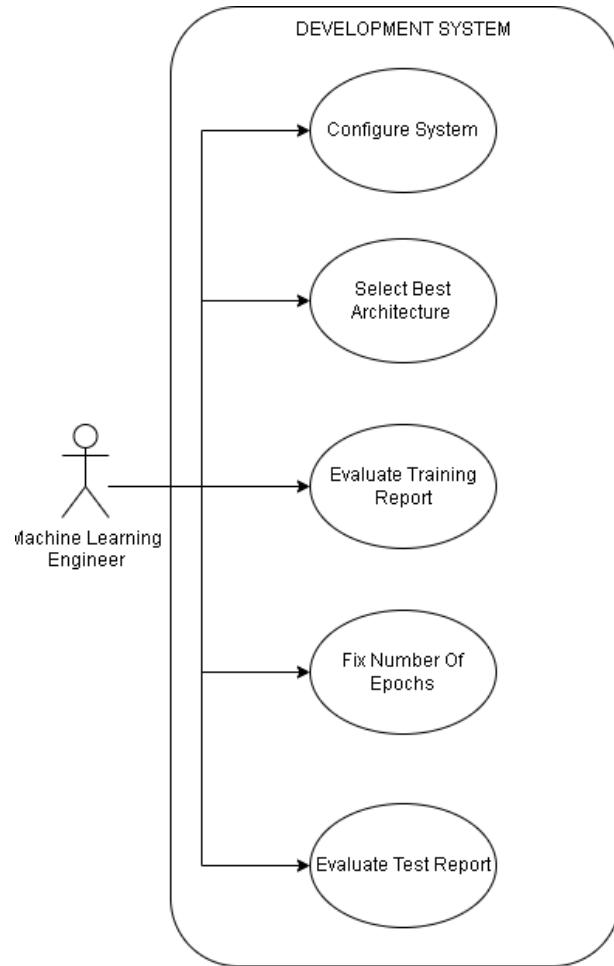


### Execution System

Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	Machine Learning Engineer	Open Configuration Form	Remember (1)	1	2,52
	System	Show Configuration Form			
2	Machine Learning Engineer	Inserts number of classifications for check	Remember (1)  A fixed percentage of the entire dataset, which is a value calculated before	1	2,52
				<b>Total Cost</b>	€ 5,04

## DEVELOP PRODUCTION ANOMALY TASKS

(Fabio)



### CONFIGURE DEVELOPMENT SYSTEM

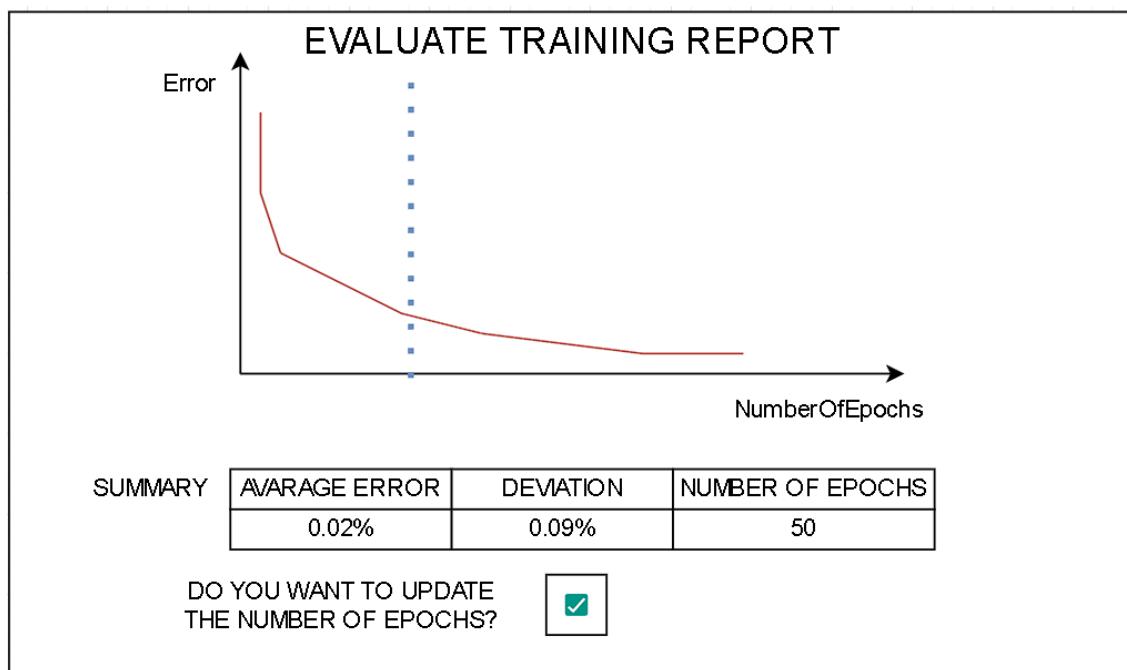
#### Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1-4]	Occurrence	Step Cost
1	Machine Learning Engineer	Open Configuration Interface	Remember (1)	1	2,52
2	Machine Learning Engineer	Define Default Hyperparameters	Remember (1)	1	2,52

3	Machine Learning Engineer	Define Validation and Test Tolerances	Remember (1) Copy and Paste of the values given by the systems requirements	1	2,52
4	Machine Learning Engineer	Define Grid Search Parameters	Apply (3) 1) See number of samples for session 2) Calculate the number of neurons needed 3) choose the number of architectures to evaluate	1	7.56
					<b>Total Cost</b> 15,12

#### EVALUATE TRAINING REPORT

#### Evaluate Training Report Mockup



#### Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1-4]	Occurrence	Step Cost
1	Machine Learning Engineer	Open Training Report	Remember (1)	1	2,52
2	Machine Learning Engineer	Evaluate Training Report	Understand (2) Check if the Deviation and The Error are good enough	1	5,04
3 IF [30%]	IF the number of epochs is not acceptable				
3.1	Machine Learning Engineer	Click Fix Number of Epochs Button	Remember (1)	0.3	0,75
3.2	System	System Opens Fix Number of Epochs Interface			
				<b>Total Cost</b>	8,31

FIX NUMBER OF EPOCHS

## Mockup

### FIX THE NUMBER OF EPOCHS

DO YOU WANT TO UPDATE THE NUMBER OF EPOCHS?

NEW NUMBER OF EPOCHS

## Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1-4]	Occurrence	Step Cost

1	System	System Opens Fix Number of Epochs Interface			
2	Machine Learning Engineer	Fix Number of Epochs	Apply (3) (a) if the loss is flat for at least the half of the iterations, (a.1) then reduce by one third the number of iterations to manage overfitting; (b) if the loss is not flat at the end of the iterations, (b.1) then enlarge by one third the number of iterations; (c) otherwise, the no. of iterations is good.	1	7,56
			<b>Total Cost</b>	7,56	

SELECT BEST ARCHITECTURE

### Select Best Classifier Mockup

## SELECT BEST ARCHITECTURE

TOLERANCE : 0.5 %

Top 5 Architectures Validation Report

ARCHITECTURE ID	VALIDATION ERROR	TRAINING ERROR	NUMBER OF NEURONS
1	0.02%	0.01%	50
2	0.04%	0.02%	60
3	0.08%	0.04%	60
4	0.7%	0.6%	55
5	0.4%	0.3%	70

SELECT 1°  
SELECT 2°  
SELECT 3°  
SELECT 4°  
SELECT 5°

SELECTED 1° ARCHITECTURE

DO YOU WANT TO UPDATE  
THE NUMBER OF EPOCHS?

### Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1-4]	Occurrence	Step Cost
1	Machine Learning Engineer	Open top 5 architectures validation report	Remember (1)	1	2,52
2	Machine Learning Engineer	Evaluate Top 5 Architectures Validation Report	Apply (3) a)Take the architecture with the lowest validation error.  b)Check if the training error is acceptable compared to the tolerance , it will be a good candidate.	1	7,56

			c)Choose another architecture that has a validation error very similar to the first one  d)Check if the training error of the last one is acceptable by comparing it with the tolerance.  e) IF the validation error is not much bigger than the previous one and fewer hyperparameters are needed THEN choose the latter, otherwise choose the former		
<b>3</b>	Machine Learning Engineer	Select Best Classifier	Remember (1)	1	2,52
4 IF [10%]	IF the number of epochs is not acceptable				
<b>4.1</b>	Machine Learning Engineer	Click Fix Number of Epochs Button	Remember (1)	0.1	0,25

4.2	System	System Opens Fix Number of Epochs Interface			
				<b>Total Cost</b>	12,85

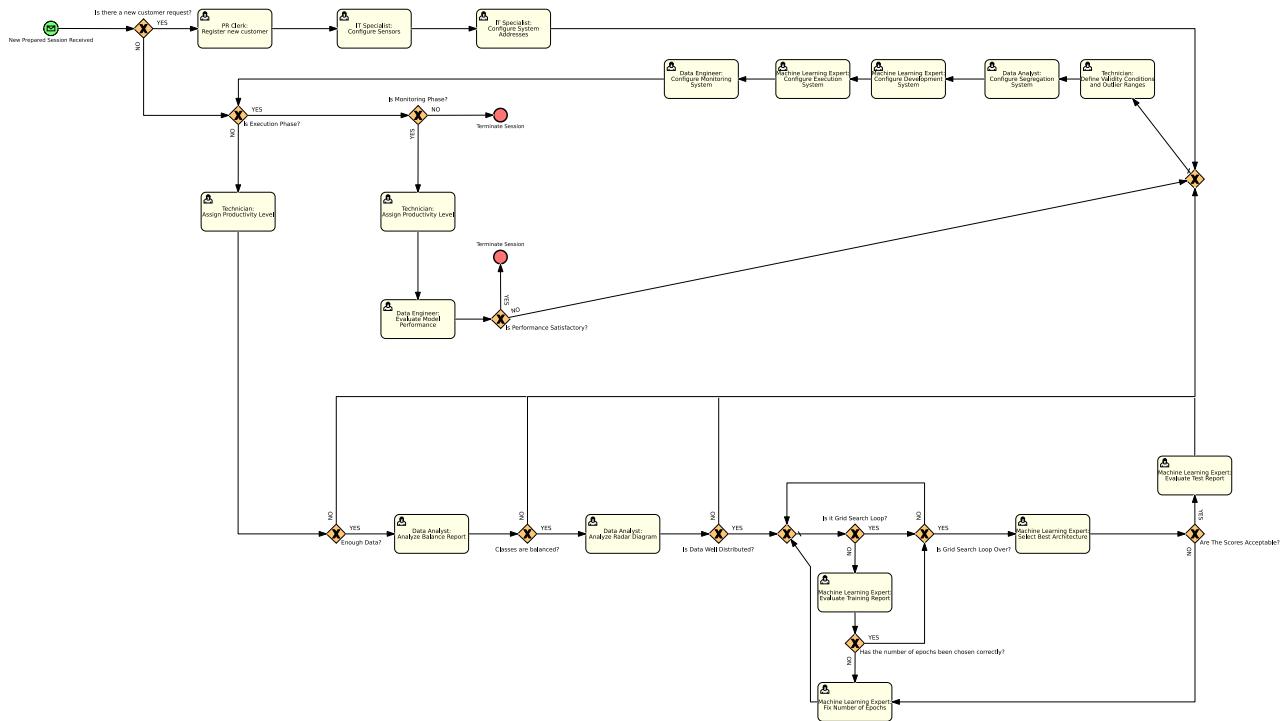
EVALUATE THE TEST REPORT

### Cost Table

Sub-Task	Actor	Step	Cognitive Effort [1-4]	Occurrence	Step Cost
1	Machine Learning Engineer	Open Test Report	Remember (1)	1	2,52
2	Machine Learning Engineer	Evaluate the test Report	Understand (2) Check the test error and compare it with the tolerance	1	5,04
				<b>Total Cost</b>	7,56

## As Is

(All)



## Gateway Probabilities Assumptions

Gateway	YES	NO	Description
Is there a new customer request	0,05%	99,95 %	Of a total of 10.000 sessions, only 5 are a registration of a new customer
Is execution phase	70%	30%	300 tokens for each distinct customer, with 2 retraining for each of them for a total of 3000 tokens in training, the remaining 7000 will go on execution for either monitoring or simple execution.
Is monitoring phase	4,3%	95,7%	60 tokens for each distinct customer filtered out from the total tokens not going in training (7000 in execution of which 300 in monitoring)
Is performance satisfactory	96,67%	3,33%	2 retraining for each customer hence 10 total retraining performed over all the monitoring tokens resulting in 3,33% of probability for retraining
Enough Data	0,5%	99,5%	1 token every 300 will be passed to the production phase
Classes are Balanced	80%	20%	It depends on the number of sessions used for the set building, the more the instances the lower the probability of unbalance, for 200 sessions a probability of 80% for balance has been set

Is data well distributed	80%	20%	It also depends on the number of sessions gathered and has been set to 80% of success
Is it grid search loop	53%	47%	The total amount of tokens in training filtered by the previous 2 gateways as total number of tokens, from which we select 15 (5 distinct customers + 2 retraining for each customer) to set the number of epochs for the grid search.
Has the number of epochs been chosen correctly	70%	30%	Depends on all the factors that might influence the quality of the classifier architecture, such as balance quality and number of samples for each class to be used
Is grid search loop over	2%	98%	Assuming a simple feed forward network we set 5 possible levels of the feed forward for 10 different possible numbers of neurons, so 50 trials for 5 different classifiers to be ranked resulting in a probability of 2% for the loop to end (5 successes every 50*5 tries).
Are the scores acceptable	90%	10%	It depends on the number of architectures to be ranked, and the number of samples giving a higher quality for each architecture.

## Result sheet

BIMP is a fast and simple web-based user interface to simulate business process models using the QBP Simulator.

See the [getting started guide](#) to read more about the features. BIMP can be used for free for academic and trial purposes. Choose the version below:

[Academic](#)   [Trial](#)   [Members](#)

## BIMP - Academic

Academic version of BIMP is supported by University of Tartu and the Estonian Research Council.

Active BPMN file  
AS-IS scenario bpmn ▾  
BPMN Diagram with results heat map   Save results   Download CSV   Save scenario   Back to edit data

## Simulation Results

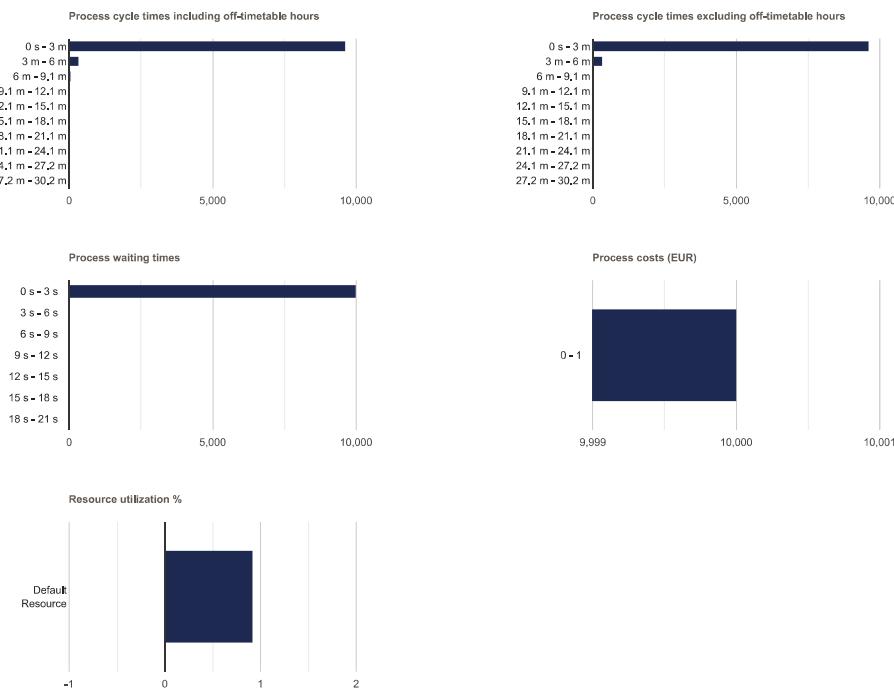
### General information

Completed process instances 10000

Total cost 0 EUR

Total simulation time 70.9 weeks

### Charts



### Scenario Statistics

	Minimum	Maximum	Average
Process instance cycle times including off-timetable hours	0 seconds	30.1 minutes	39.3 seconds
Process instance cycle times excluding off-timetable hours	0 seconds	30.1 minutes	39.3 seconds
Process instance costs	0 EUR	0 EUR	0 EUR

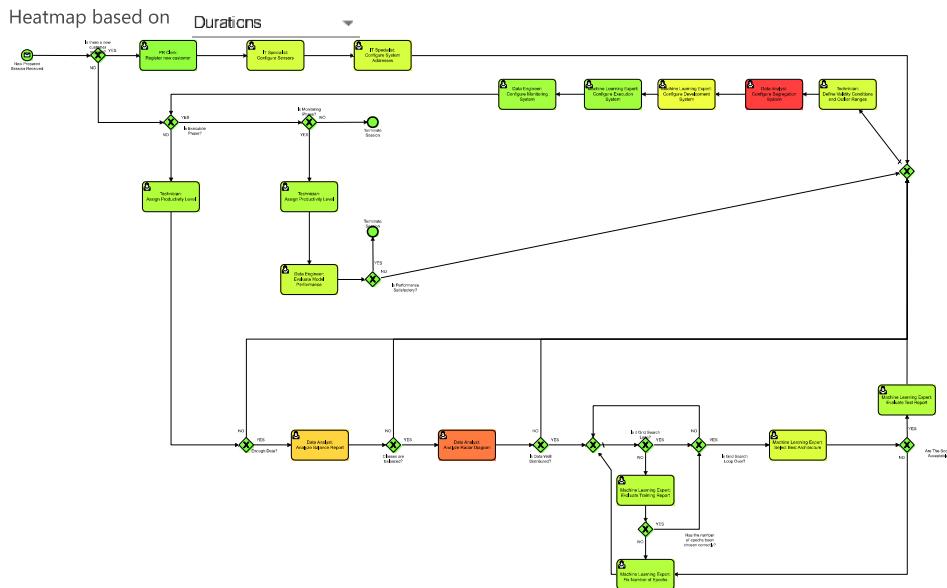
#### Activity Durations, Costs, Waiting times, Deviations from Thresholds

Name	Waiting time				Duration				Duration over threshold			Cost			Cost over threshold		
	Count	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	
&#10;Machine Learning Expert:&#10;Evaluate Training Report	410	0 s	0 s	0 s	7.3 s	8.3 s	9.3 s	0 s	0 s	0 s	0	0	0	0	0	0	
&#10;Machine Learning Expert:&#10;Fix Number of Epochs	134	0 s	0 s	0 s	6.8 s	7.6 s	8.3 s	0 s	0 s	0 s	0	0	0	0	0	0	
Data Analyst:&#10;Analyze Balance Report	19	0 s	0 s	0 s	21.2 s	23.4 s	25.6 s	0 s	0 s	0 s	0	0	0	0	0	0	
Data Analyst:&#10;Analyze Radar Diagram	17	0 s	0 s	0 s	32 s	35.4 s	38.3 s	0 s	0 s	0 s	0	0	0	0	0	0	
Data Analyst:&#10;Configure Segregation System	4323	0 s	0 s	5.3 s	39.3 s	43.6 s	51.5 s	0 s	0 s	0 s	0	0	0	0	0	0	
Data Engineer:&#10;Configure Monitoring System	4323	0 s	0 s	11.3 s	4.7 s	5.2 s	16.5 s	0 s	0 s	0 s	0	0	0	0	0	0	
Data Engineer:&#10;Evaluate Model Performance	443	0 s	0 s	0 s	9.7 s	10.7 s	11.7 s	0 s	0 s	0 s	0	0	0	0	0	0	
IT Specialist:&#10;Configure Sensors	5	0 s	0 s	0 s	11.2 s	11.8 s	12.5 s	0 s	0 s	0 s	0	0	0	0	0	0	
IT Specialist:&#10;Configure System Addresses	5	0 s	0 s	0 s	11.4 s	11.8 s	12.3 s	0 s	0 s	0 s	0	0	0	0	0	0	
Machine Learning Expert:&#10;Configure Development System	4323	0 s	0 s	0 s	13.6 s	15.1 s	16.6 s	0 s	0 s	0 s	0	0	0	0	0	0	
Machine Learning Expert:&#10;Configure Execution System	4323	0 s	0 s	7.3 s	4.5 s	5 s	12.1 s	0 s	0 s	0 s	0	0	0	0	0	0	
Machine Learning Expert:&#10;Evaluate Test Report	11	0 s	0 s	0 s	6.9 s	7.6 s	8.3 s	0 s	0 s	0 s	0	0	0	0	0	0	
Machine Learning Expert:&#10;Select Best Architecture	11	0 s	0 s	0 s	12 s	13 s	13.8 s	0 s	0 s	0 s	0	0	0	0	0	0	
PR Clerk:&#10;Register new customer	5	0 s	0 s	0 s	2.4 s	2.5 s	2.6 s	0 s	0 s	0 s	0	0	0	0	0	0	
Technician:&#10;Assign Productivity Level	4301	0 s	0 s	10.8 s	6.5 s	7.3 s	18.1 s	0 s	0 s	0 s	0	0	0	0	0	0	
Technician:&#10;Assign Productivity Level	443	0 s	0 s	0 s	6.5 s	7.2 s	8 s	0 s	0 s	0 s	0	0	0	0	0	0	
Technician:&#10;Define Validity Conditions and Outlier Ranges	4323	0 s	0 s	4.7 s	10.5 s	11.6 s	16 s	0 s	0 s	0 s	0	0	0	0	0	0	

## Durations Heatmap

### Heatmap

Heatmap based on



### Legend

Color

Value

0 s

5 s

10 s

15 s

19 s

24 s

29 s

34 s

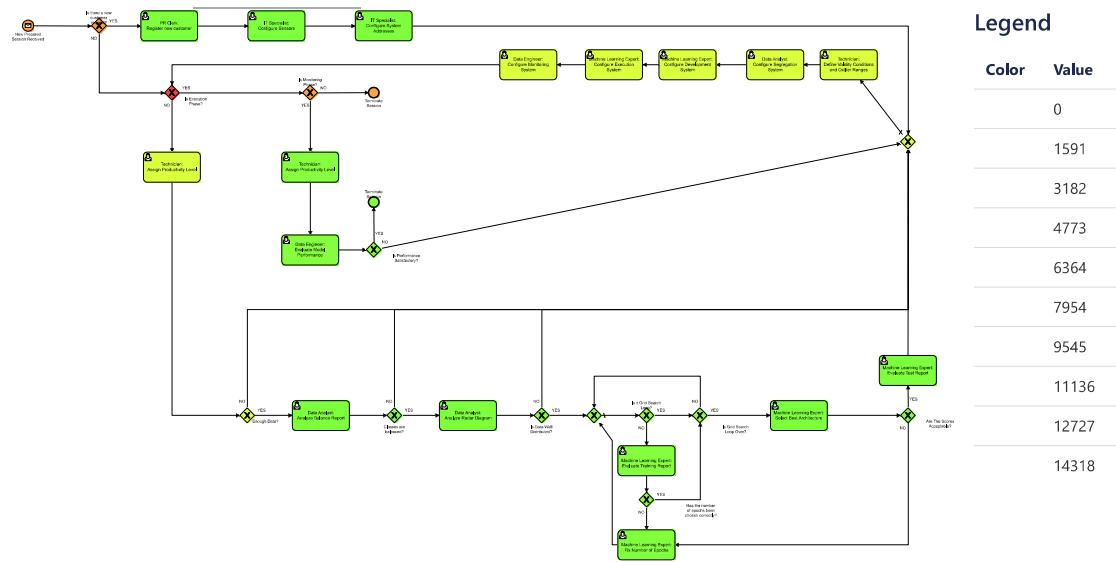
39 s

44 s

## Counts Heatmap

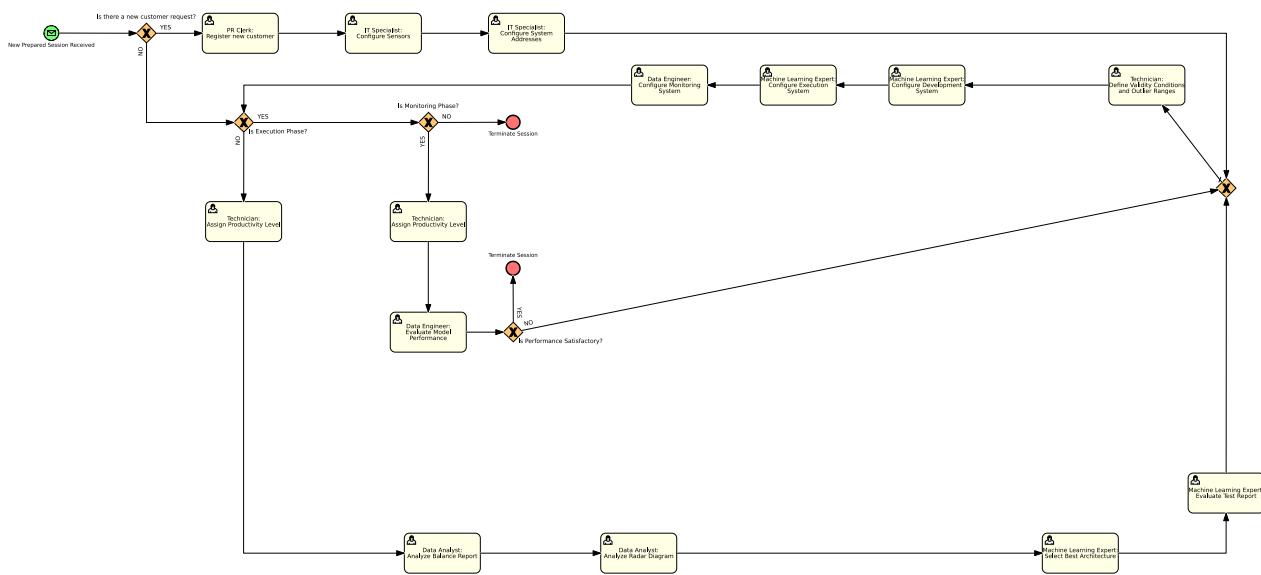
## Heatmap

Heatmap based on Counts



## To Be

(All)



### *Handoff modification:*

During the registration phase, Customers specify the number of rolls produced and types of paper produced. With this information it is possible to tell if two customers have similar requirements and use the raw data from a model built in the past to balance classes for the new customer that may need additional data. For this reason, configuration of thresholds and ranges are imported from external sources reducing the cognitive efforts of configuration steps to Remember (1).

The grid search loop is removed since the optimal parameters are taken from the best architecture of existent customers, which is then fine-tuned on the customer scenario using a fine-tuning script.

“Enough Data” gateway is removed because the “no” condition would cause the script to rebalance the dataset using external sources data, resulting in a 100% of “yes” condition. Same can be done with “Classes are Balanced” and “Is data well distributed” gateways.

“Are the scores Acceptable” gateway is dropped because it has already been assessed on the other configuration

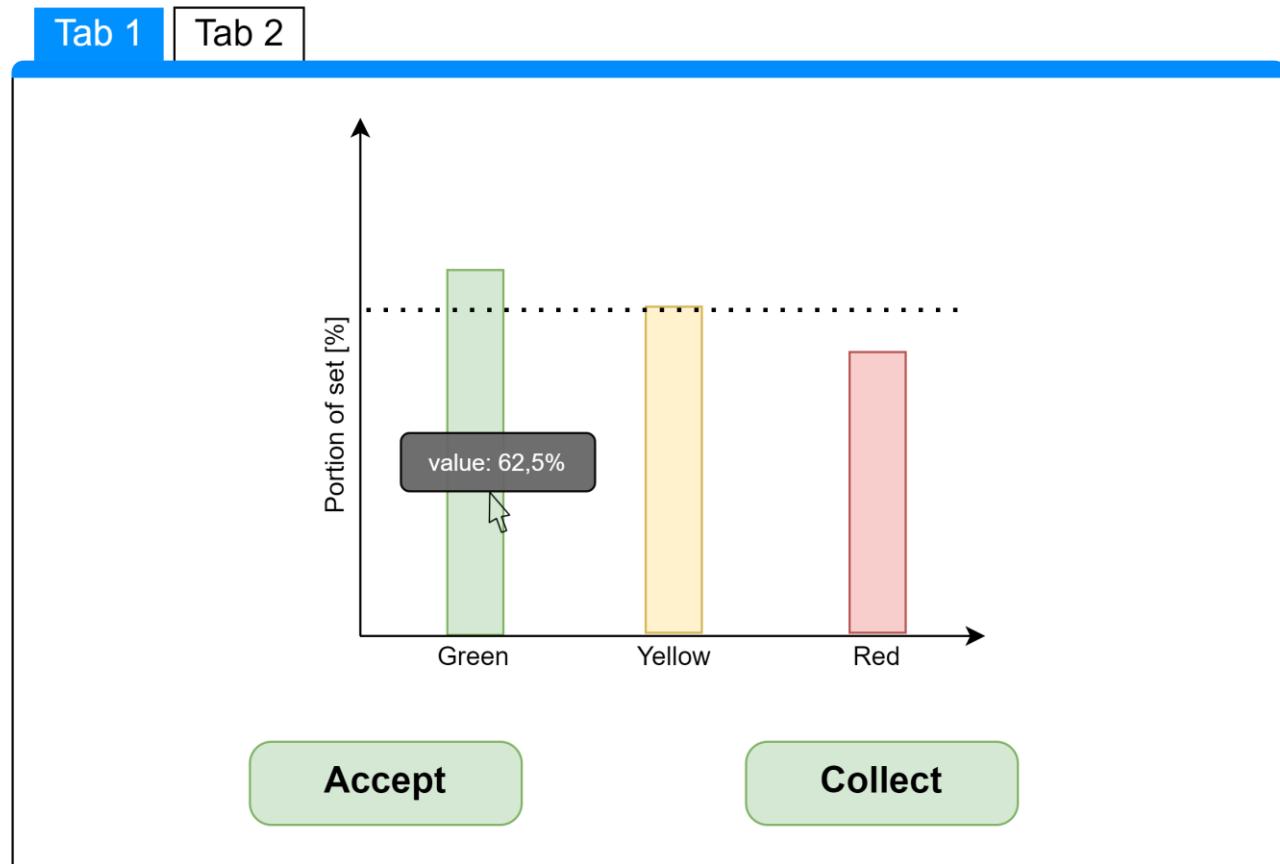
### *Service modification:*

The configuration task of the segregation system has been substituted by a script task since the availability of external data allows to always being able to gather new data from lacking classes and obtain the optimal threshold.

### *Task modification:*

The cognitive effort for the “Analyze Balance Report” is dropped to a simple Understand (2) since there are no more multiple thresholds to be compared but data can be collected to match the number

## Modified Analyze Balance Report Task



Sub-Task	Actor	Step	Cognitive Effort [1 - 4]	Occurrence	Step Cost
1	Data Analyst	Open Balance Report	Remember (1)	1	2,73
	System	Show Balance Report			
2	Data Analyst	Check if Data are Balanced	Understand (2)	1	5,46
IF (20%)		If data are far from threshold			
3.1	Data Analyst	Click "Collect" Button	Remember (1)	0,2	0,55
	System	Collects missing samples from under sampled classes			

<b>ELSE (80%)</b>					
4.1	Data Analyst	Click "Approve" Button	Remember (1)	0,8	2,18
	System	Start generating the production session set			
			<b>Total Cost</b>	€ 10,92	

## Result Sheet

BIMP is a fast and simple web-based user interface to simulate business process models using the QBP Simulator.

See the [getting started guide](#) to read more about the features. BIMP can be used for free for academic and trial purposes. Choose the version below:

[Academic](#)   [Trial](#)   [Members](#)

## BIMP - Academic

Academic version of BIMP is supported by University of Tartu and the Estonian Research Council.

Active BPMN  
file TO-RF bpmn ▾  
BPMN Diagram with results heat map Save results Download CSV Save scenario Back to edit data

## Simulation Results

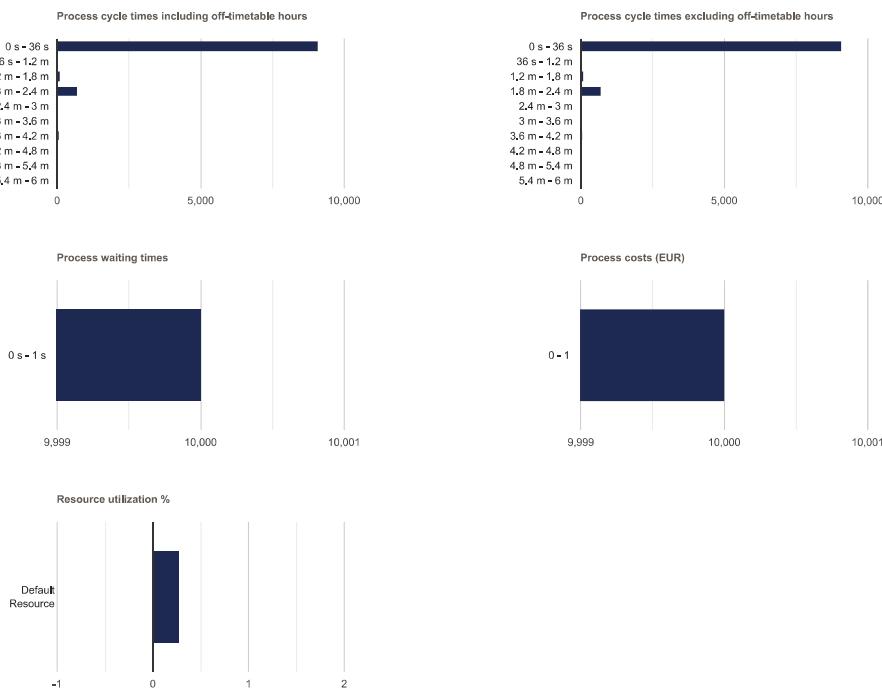
### General information

Completed process instances 10000

Total cost 0 EUR

Total simulation time 70.6 weeks

### Charts



### Scenario Statistics

	Minimum	Maximum	Average
Process instance cycle times including off-timetable hours	0 seconds	5.9 minutes	11.4 seconds
Process instance cycle times excluding off-timetable hours	0 seconds	5.9 minutes	11.4 seconds
Process instance costs	0 EUR	0 EUR	0 EUR

#### Activity Durations, Costs, Waiting times, Deviations from Thresholds

Name	Waiting time				Duration				Duration over threshold			Cost			Cost over threshold		
	Count	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	
Data Analyst:&#10;Analyze Balance Report	967	0 s	0 s	0 s	9.9 s	10.9 s	11.9 s	0 s	0 s	0 s	0	0	0	0	0	0	
Data Analyst:&#10;Analyze Radar Diagram	967	0 s	0 s	0 s	32 s	35.5 s	39 s	0 s	0 s	0 s	0	0	0	0	0	0	
Data Engineer:&#10;Configure Monitoring System	988	0 s	0 s	0 s	4.7 s	5.2 s	5.7 s	0 s	0 s	0 s	0	0	0	0	0	0	
Data Engineer:&#10;Evaluate Model Performance	319	0 s	0 s	0 s	9.7 s	10.7 s	11.7 s	0 s	0 s	0 s	0	0	0	0	0	0	
IT Specialist:&#10;Configure Sensors	4	0 s	0 s	0 s	11.2 s	11.9 s	12.3 s	0 s	0 s	0 s	0	0	0	0	0	0	
IT Specialist:&#10;Configure System Addresses	4	0 s	0 s	0 s	10.9 s	11.5 s	12.6 s	0 s	0 s	0 s	0	0	0	0	0	0	
Machine Learning Expert:&#10;Configure Development System	988	0 s	0 s	0 s	13.6 s	15.1 s	16.6 s	0 s	0 s	0 s	0	0	0	0	0	0	
Machine Learning Expert:&#10;Configure Execution System	988	0 s	0 s	0 s	4.5 s	5 s	5.5 s	0 s	0 s	0 s	0	0	0	0	0	0	
Machine Learning Expert:&#10;Evaluate Test Report	967	0 s	0 s	0 s	6.8 s	7.6 s	8.3 s	0 s	0 s	0 s	0	0	0	0	0	0	
Machine Learning Expert:&#10;Select Best Architecture	967	0 s	0 s	0 s	11.9 s	12.9 s	13.8 s	0 s	0 s	0 s	0	0	0	0	0	0	
PR Clerk:&#10;Register new customer	4	0 s	0 s	0 s	2.4 s	2.6 s	2.8 s	0 s	0 s	0 s	0	0	0	0	0	0	
Technician:&#10;Assign Productivity Level	319	0 s	0 s	0 s	6.5 s	7.3 s	8 s	0 s	0 s	0 s	0	0	0	0	0	0	
Technician:&#10;Assign Productivity Level	967	0 s	0 s	0 s	6.5 s	7.2 s	8 s	0 s	0 s	0 s	0	0	0	0	0	0	
Technician:&#10;Define Validity Conditions and Outlier Ranges	988	0 s	0 s	0 s	10.5 s	11.6 s	12.7 s	0 s	0 s	0 s	0	0	0	0	0	0	

© 2020 QBP Simulator

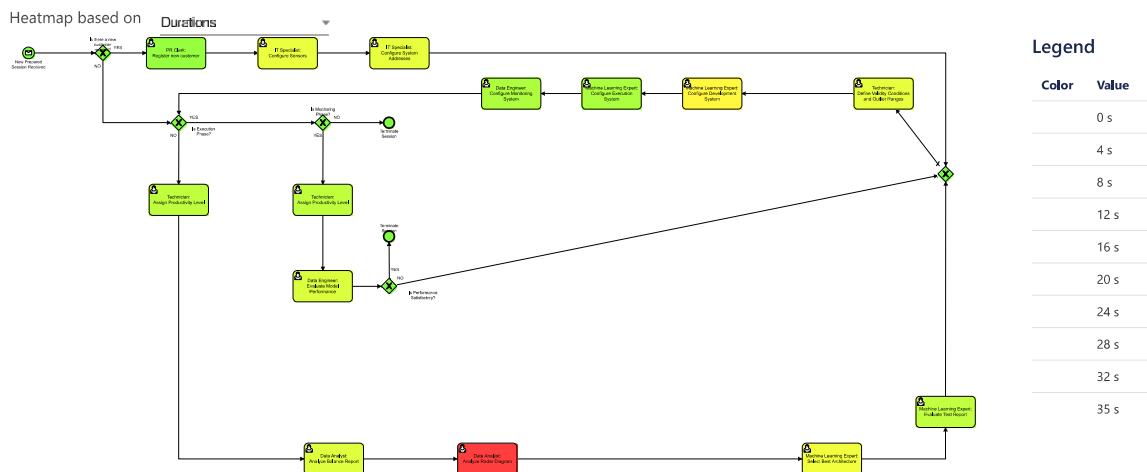


This website uses cookies to enhance your site experience and for analytical purposes. By proceeding on this website you are OK with this.

[Close](#)

## Durations Heatmap

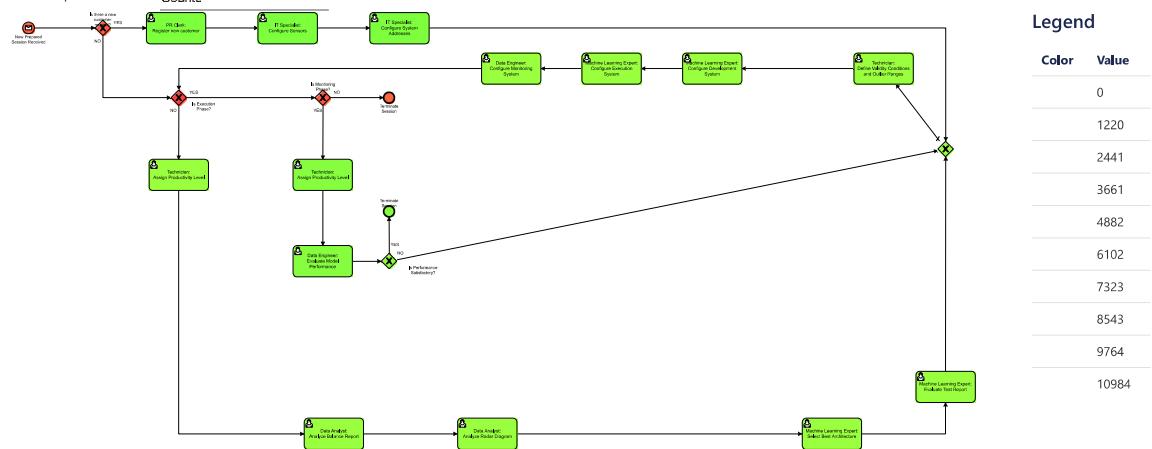
## Heatmap



## Counts Heatmap

## Heatmap

Heatmap based on Counts

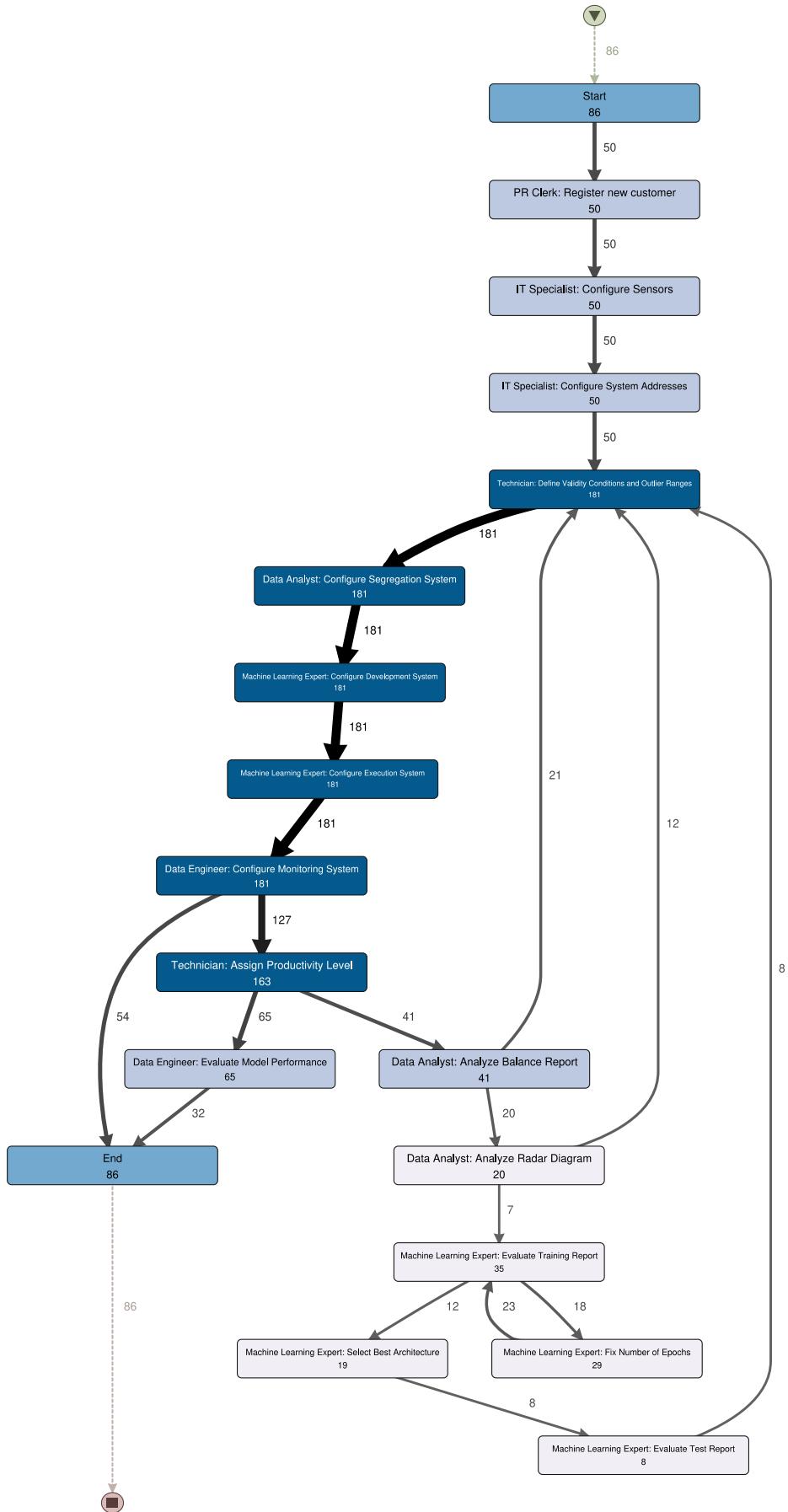


## Process Mining and Conformance Check

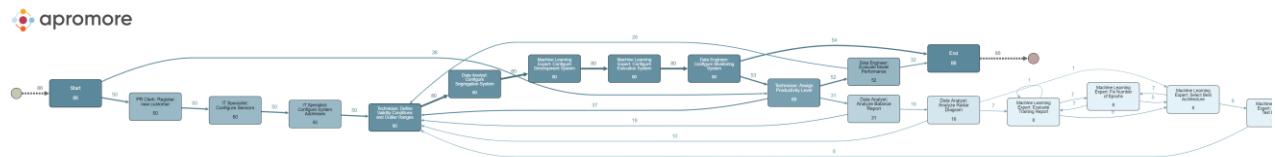
### Transition Maps

(Disco – Luca, Fabio; Apromore – Giulio, Giovanni)

[Transition Map Disco](#)



## Transition Map Apromore



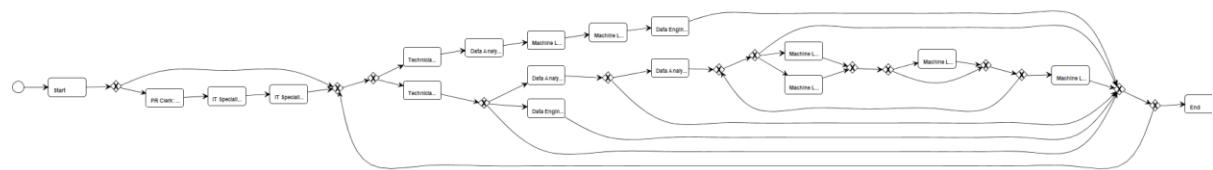
Using disco we managed to obtain commons start and end events, which are reported in both the transition maps.

As differences we can see that Disco counts on the arcs is higher with respect to Apromore, in fact, the Disco way to count tokens in a task is cumulative, and applies loops back causing an higher count on arcs inside a loop whereas Apromore just filter the probabilities at each arc advancing forward only.

## FILTERED BPMN (ProM)

(Luca, Fabio)

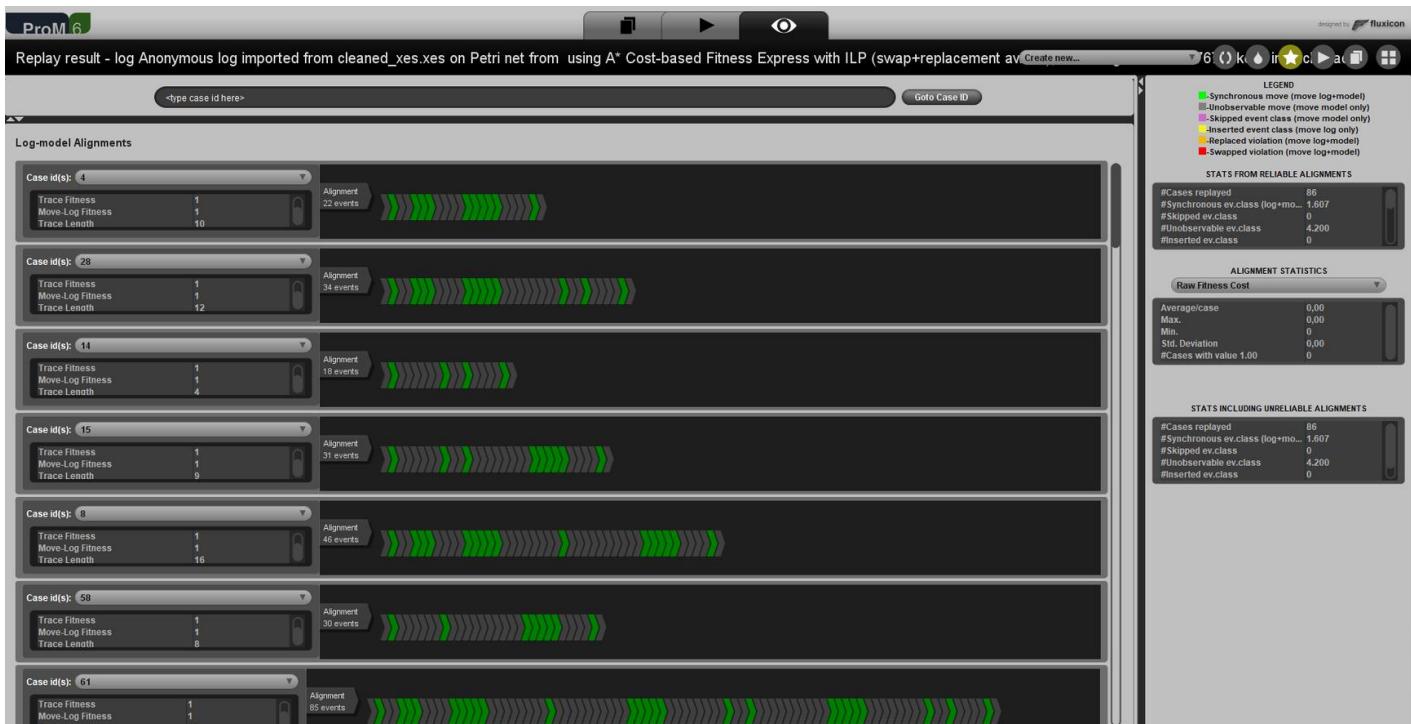
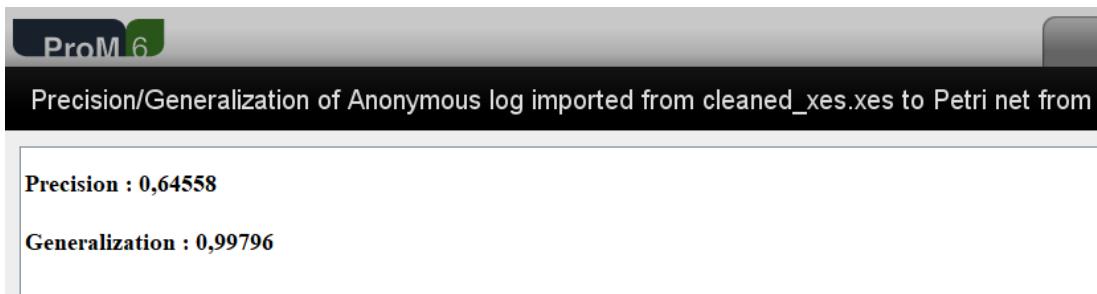
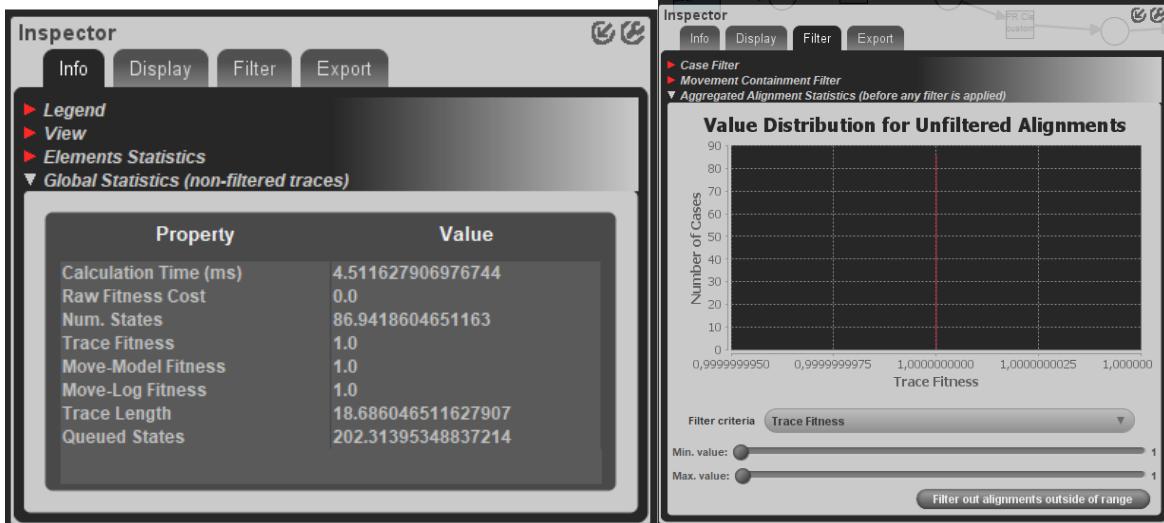
BPMN mined using inductive miner



## Quality measures:

- Precision  $\approx 0,646$

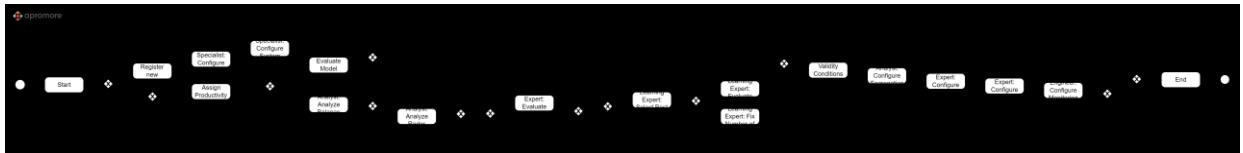
- Generalization  $\approx 0.998$
- Fitness = 1
- Simplicity = #gateways + #sequence flows + #activities = 13+42+18 = 70



## FILTERED BPMN (Apromore)

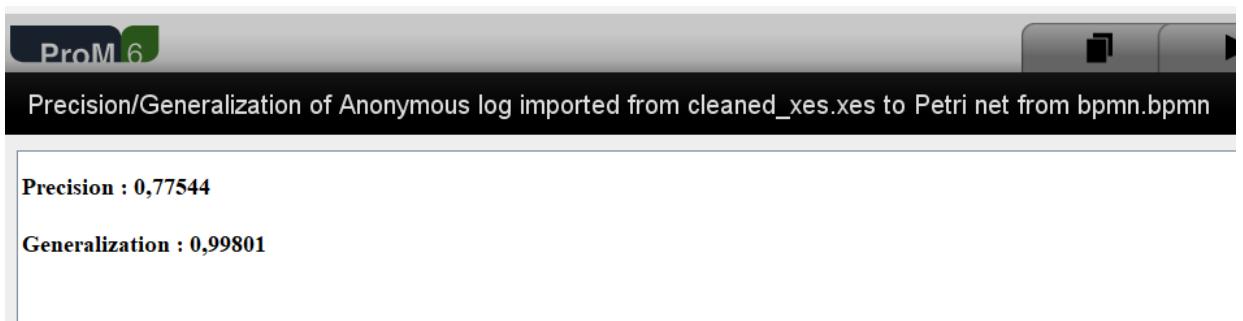
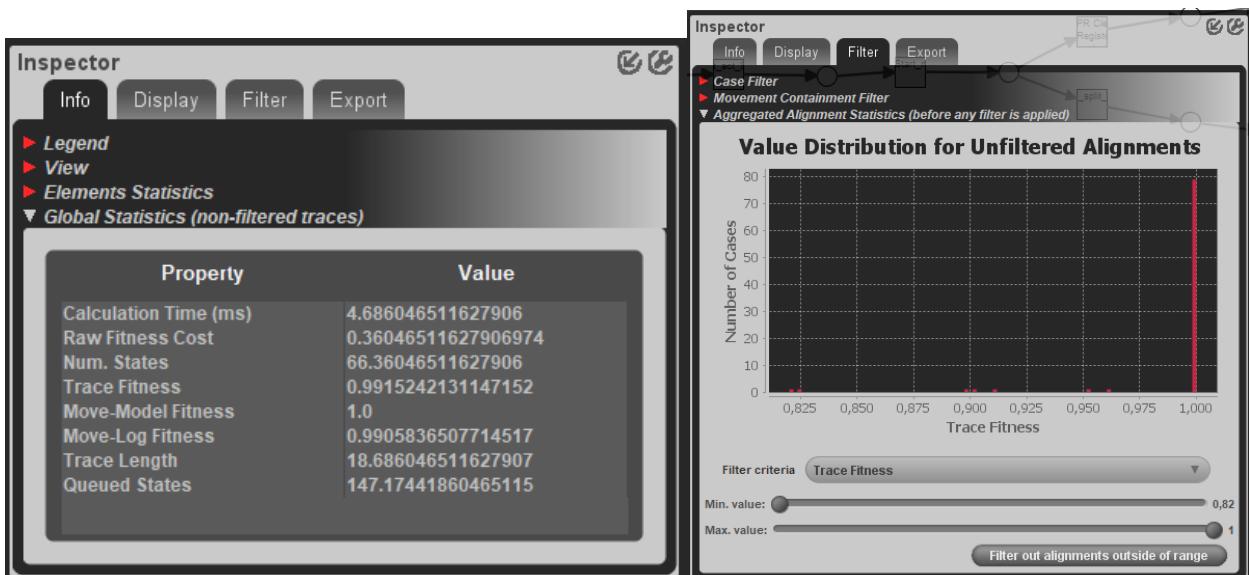
(Giulio, Giovanni)

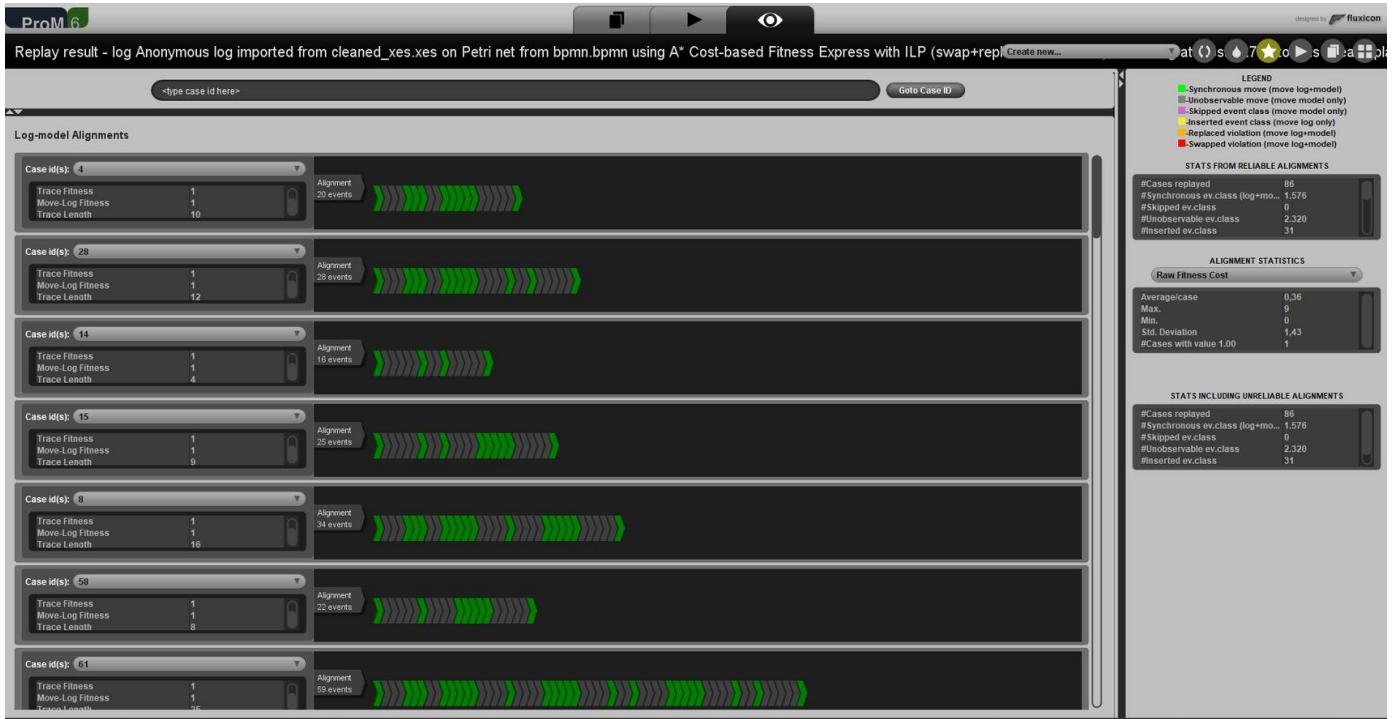
### BPMN mined using apromore



### Quality measures:

- Precision = 0,77
- Generalization = 0,998
- Fitness = 0,99
- Simplicity =  $13 + 42 + 18 = 70$





## Violations to the model (All)

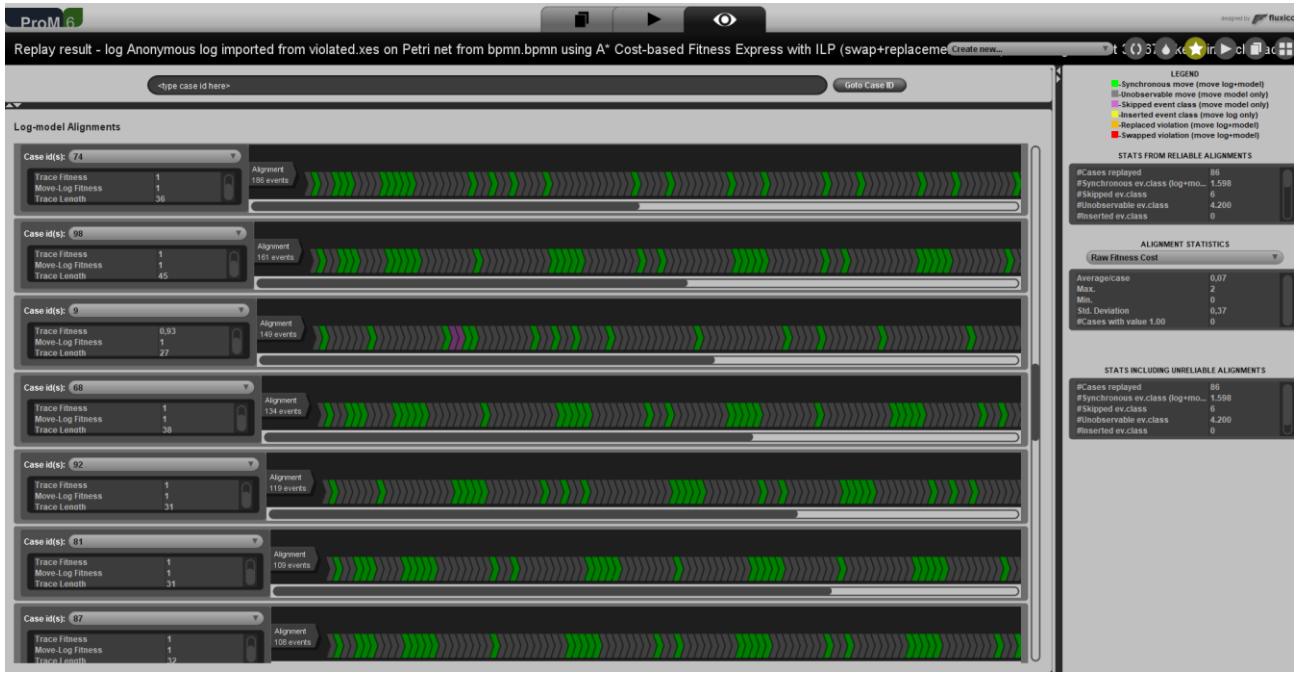
Violations:

- Configure Development system can be removed since once external data is found for our purposes, we can adopt also the same hyperparameters.
- The task fix number of Epochs can be removed because the number of epochs from models trained on similar scenarios can be reused or because of a human mistake the ML Engineer might want to return to training report analysis and skip a single iteration of “fix number of epochs”.
- Configure segregation system can be skipped because the threshold can be set basing on other customer’s similar cases or for similar models.

Conformance check of violations with mined model:

Inspector	
Info	
Display	
Filter	
Export	
▶ Legend	
▶ View	
▶ Elements Statistics	
▼ Global Statistics (non-filtered traces)	
Property	Value
Calculation Time (ms)	4.8023255813953485
Raw Fitness Cost	0.06976744186046513
Num. States	86.30232558139537
Trace Fitness	0.998079474823661
Move-Model Fitness	0.9980235393226571
Move-Log Fitness	1.0
Trace Length	18.581395348837205
Queued States	200.9534883720931

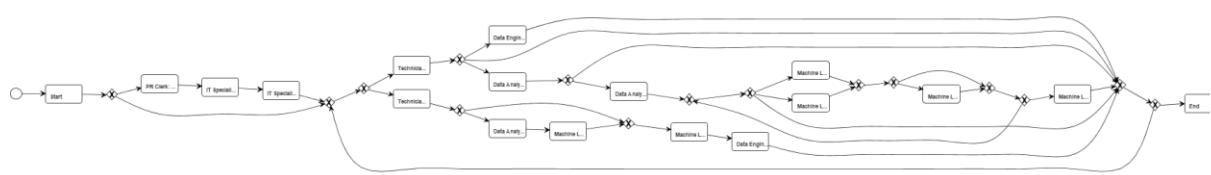
We can notice the fitness is slightly dropped, also in the following image we can notice some purple points on the modified case ID dropping the fitness of the case, while all others (not modified) are reporting a full fitness:



We can also notice a third step should be skipped in the modified log, but it doesn't appear and that may be due to the fact that the generated petrinet might be finding an existent path matching with the sequence not involving the: "fix number of epochs" activity, removed in that log.

ProM Statistics with violations  
(Fabio, Luca)

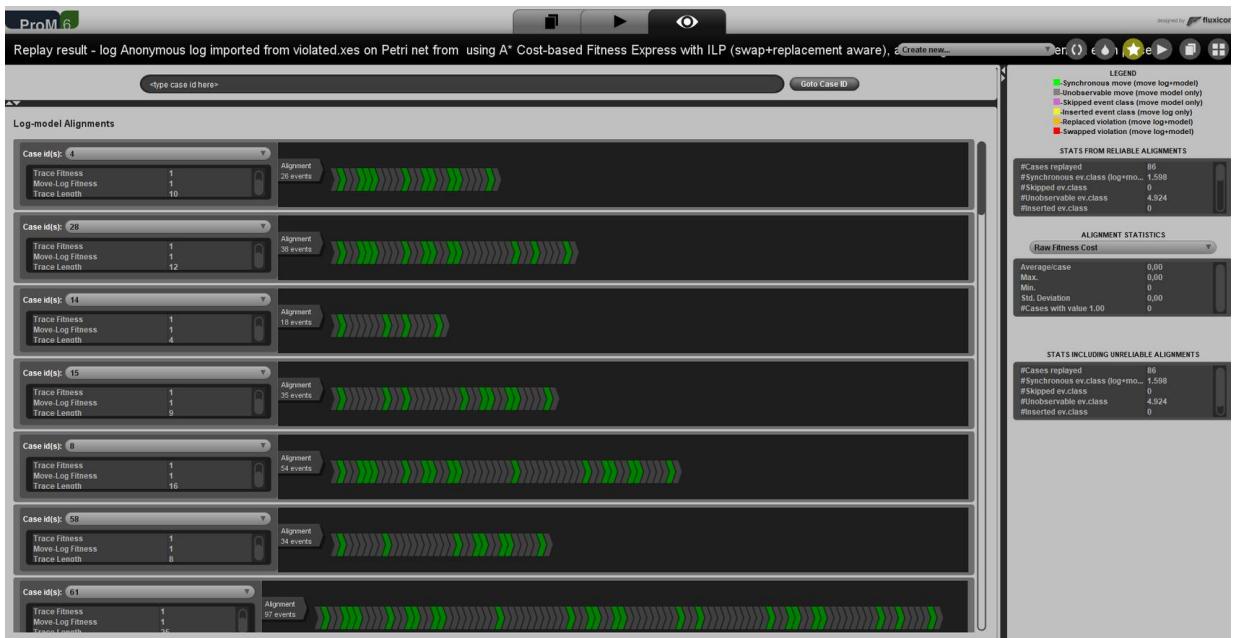
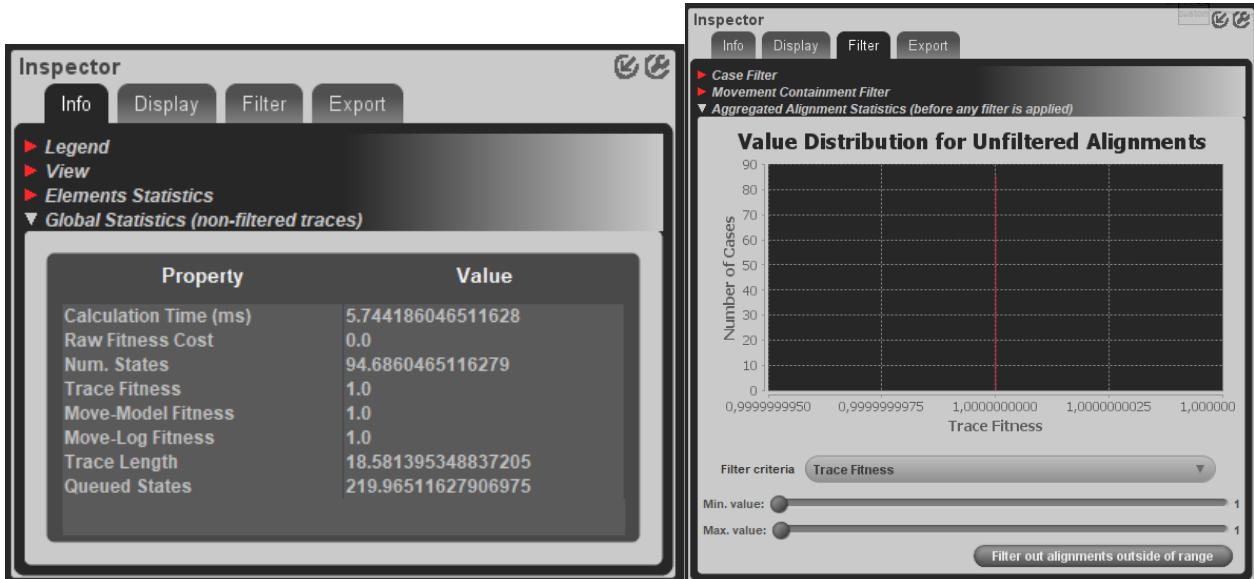
Mined BPMN model using inductive miner on violation logs



## Quality measures:

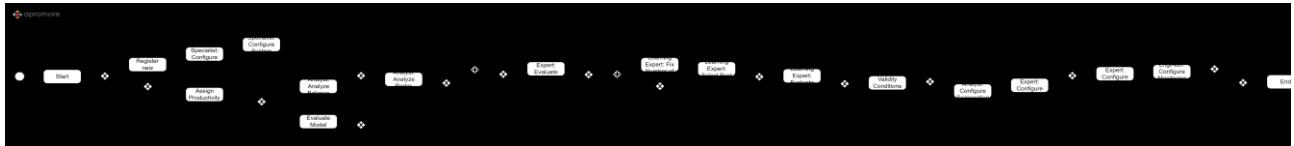
- Precision = 0,541
- Generalization = 0,998

- Fitness = 1
- Simplicity =  $15 + 44 + 18 = 77$



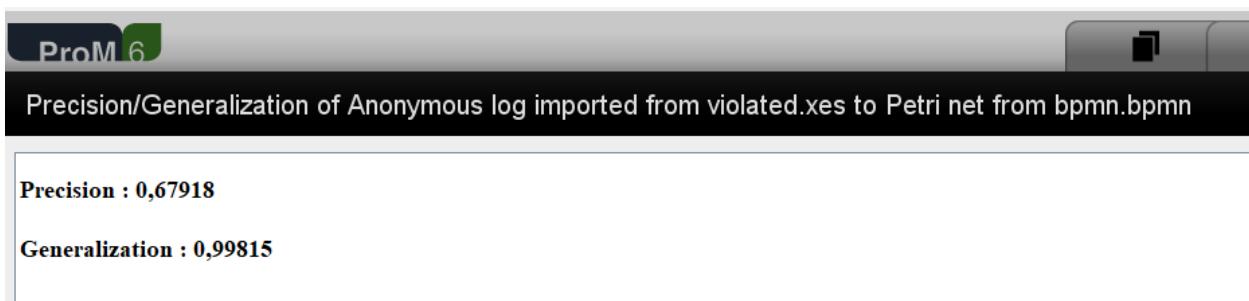
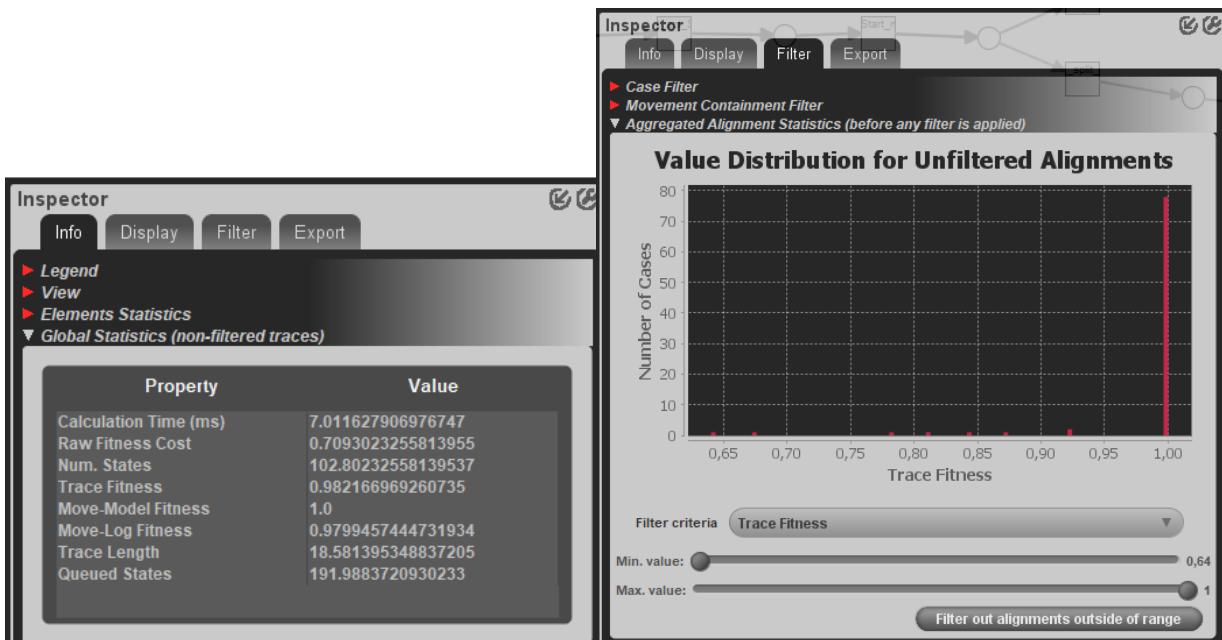
Apronmore Statistics with violations  
(Giovanni, Giulio)

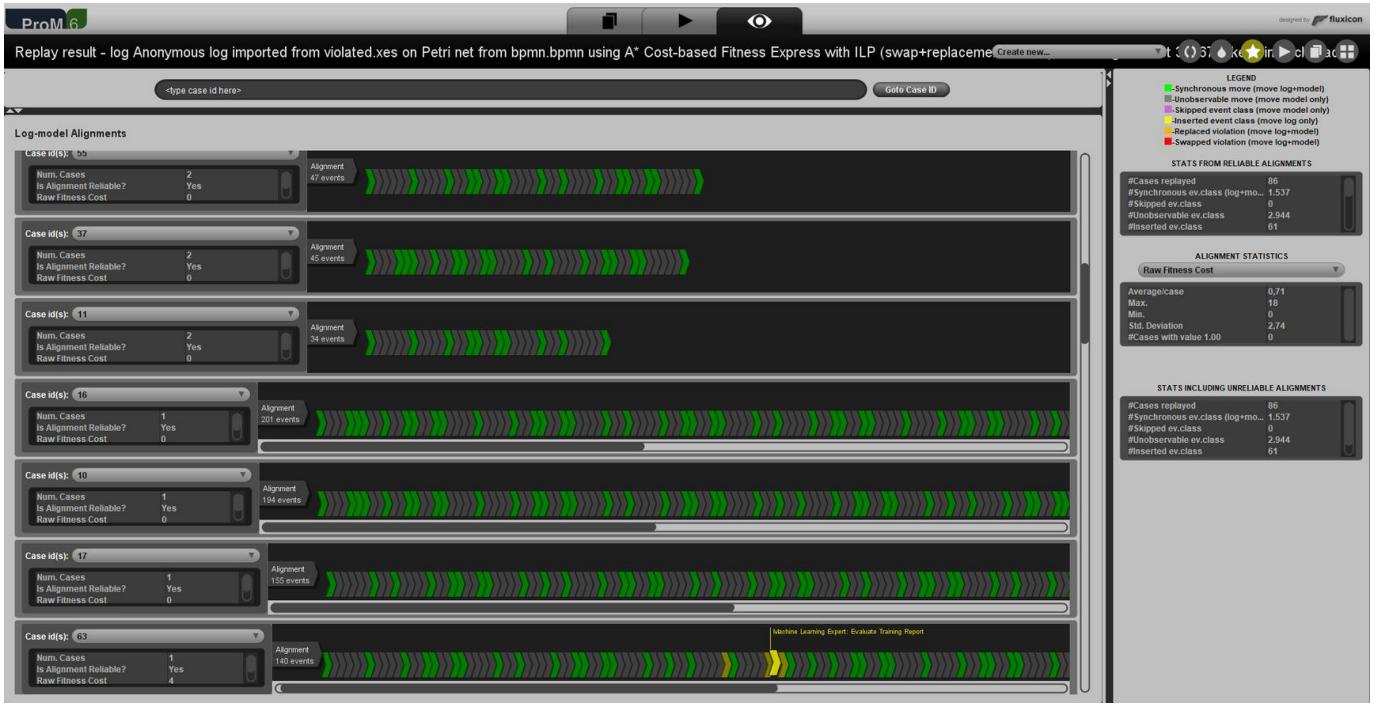
## Mined BPMN model from violation logs



### Quality Measures:

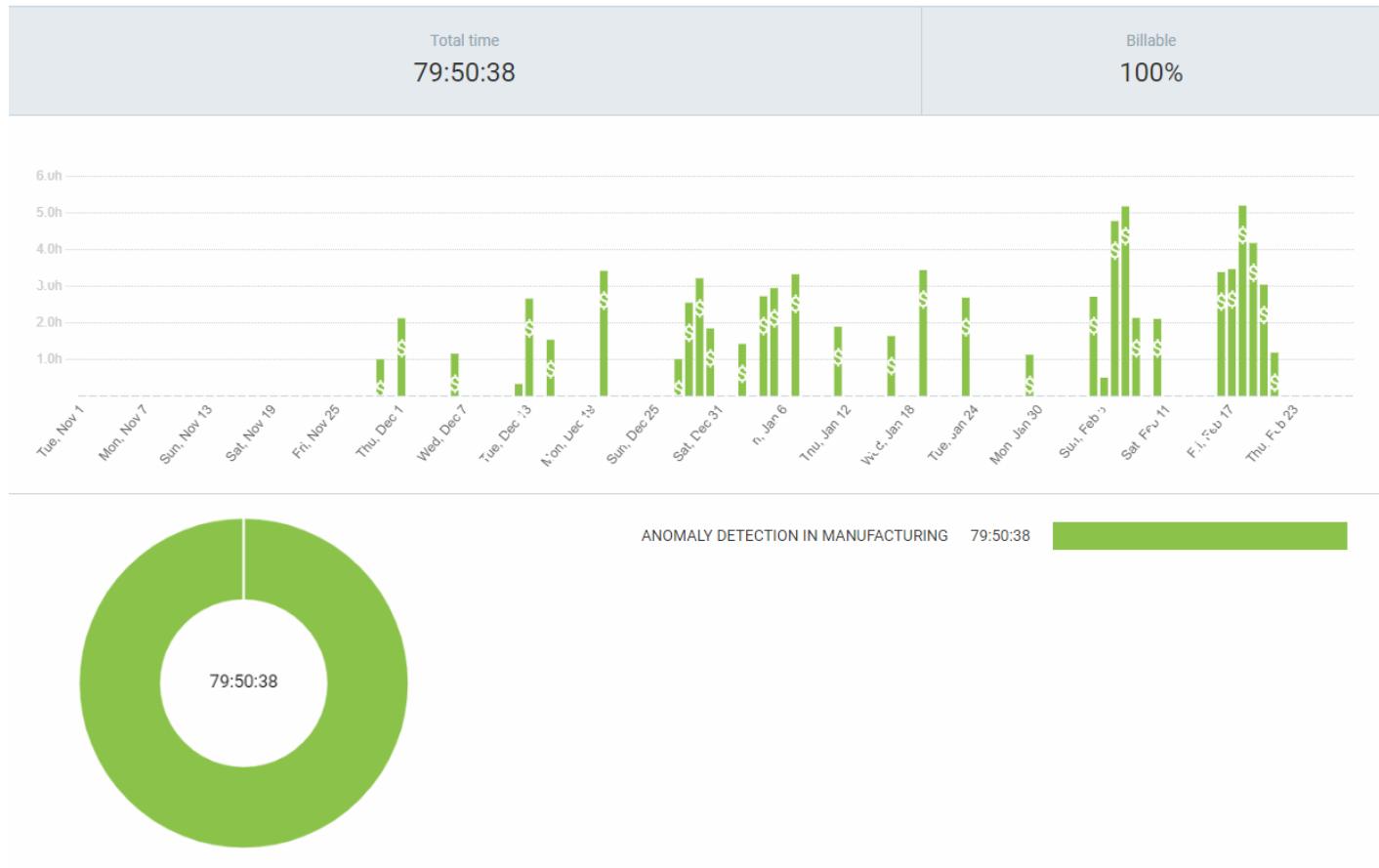
- Precision = 0,679
- Generalization = 0,998
- Fitness = 0,982
- Simplicity =  $15 + 48 + 17 = 80$



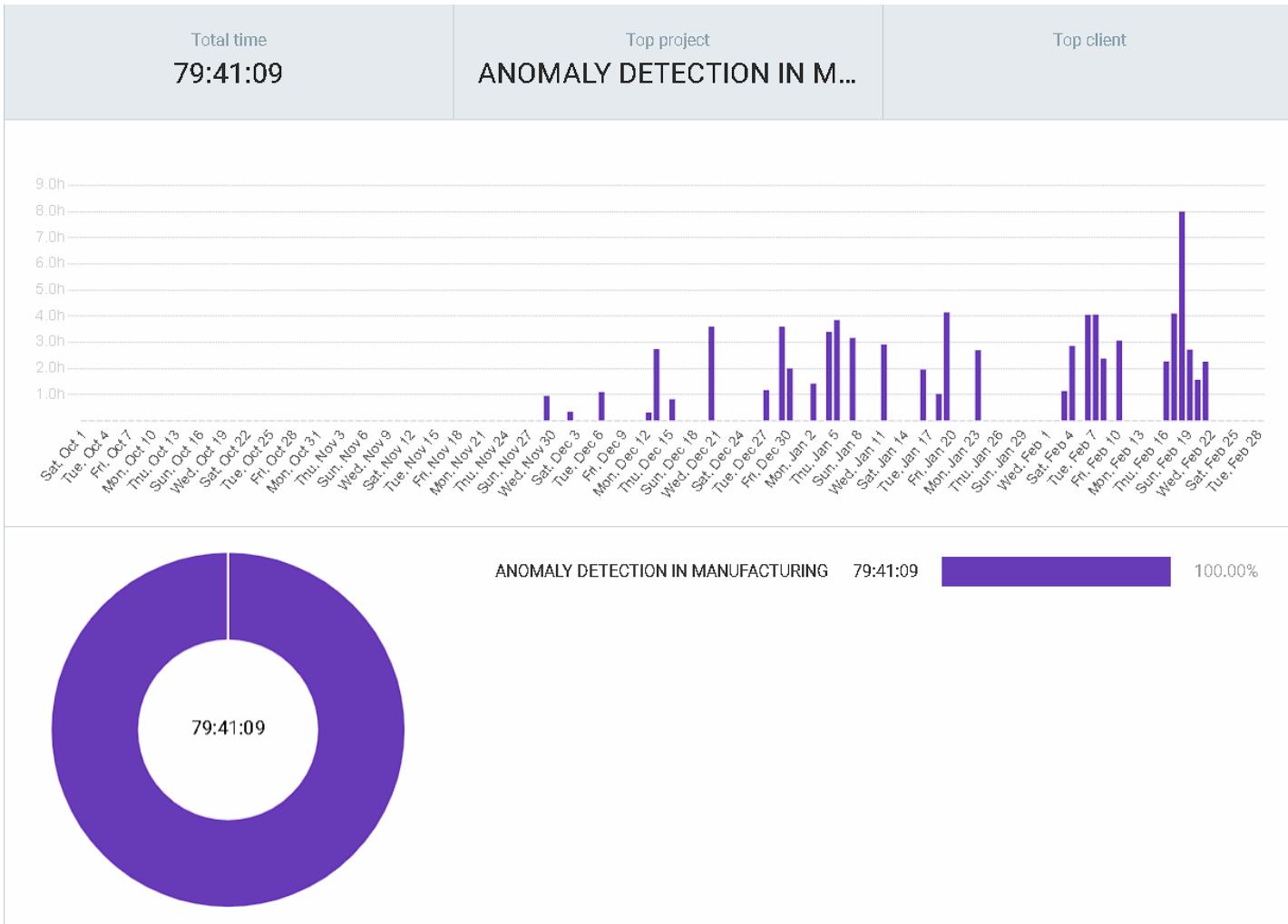


## Trello Dashboard

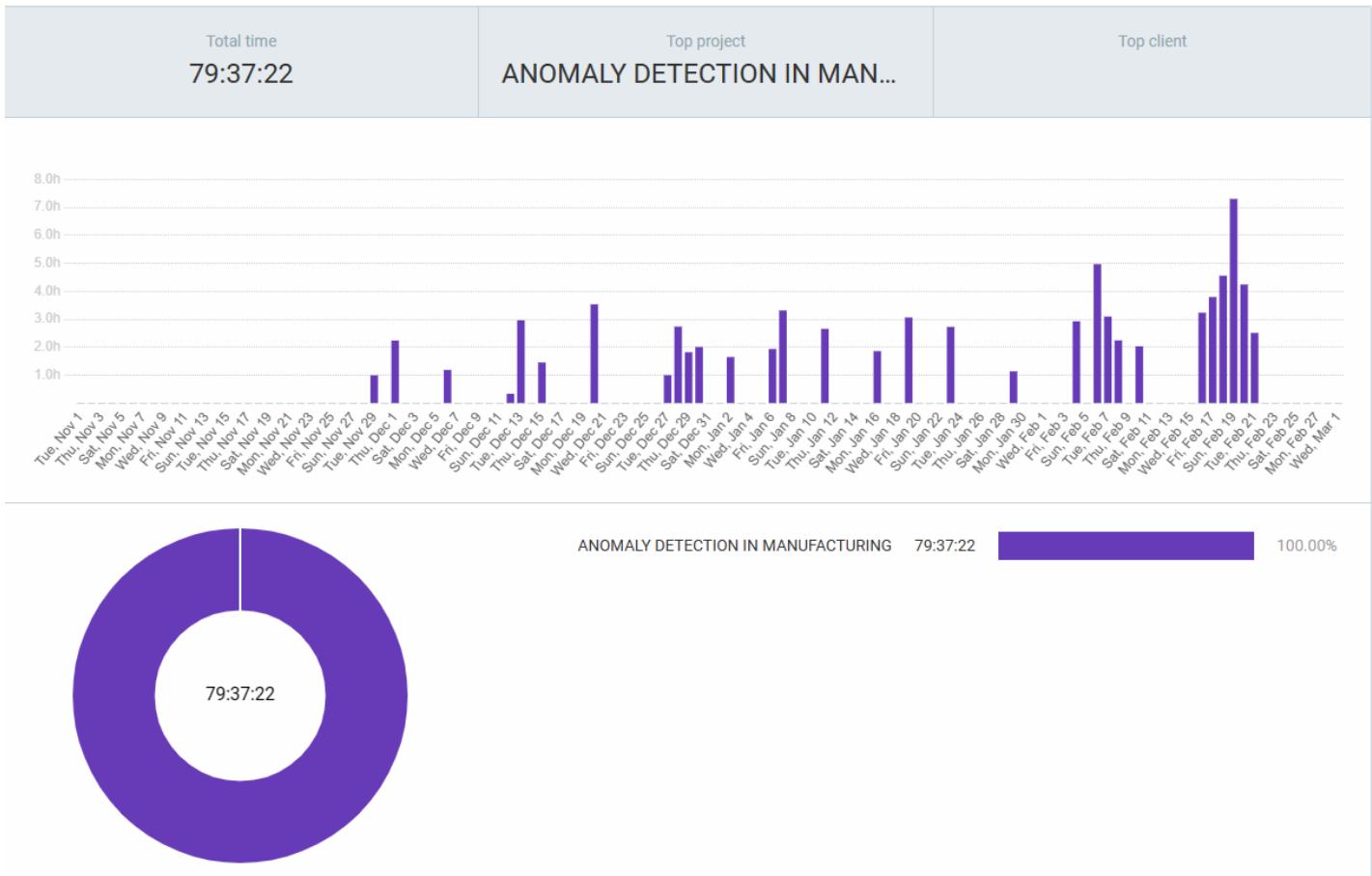
Luca



Fabio



Giovanni



## Giulio

