Process Mining and Intelligence 2022/23

EARLY FOREST FIRE DETECTION

UNIVERSITÀ DI PISA

Team members:

- Elisa De Filomeno
- Stefano Dugo
- Fabiano Pilia
- Valerio Secondulfo

Note

- 4 Records:

1. PHOTOS VECTORIAL -> foto vettorizzate ma non annotate
2. ANNOTATION TOOL (target)
3. ENVIRONMENTAL RECORD -> tipo di incendio, vastità, ...
4. VECTORIAL LAYOUT RECORD-> obstacles, laghi, …

- Missione di ricognizione: fatta da 4 records

- Targets: principi d'incendio (picco singolo e picco multiplo)

- Parametrizzatore basato su rete neurale genetica

- Parametri sono ottimizzati dal parametrizzatore usati per la logica di coordinamento dei droni

- Parametri:

1. RAGGIO DI AZIONE
2. RAGGIO DI RICHIAMO DEL DRONE
3. PERSISTENZA DEL RICHIAMO

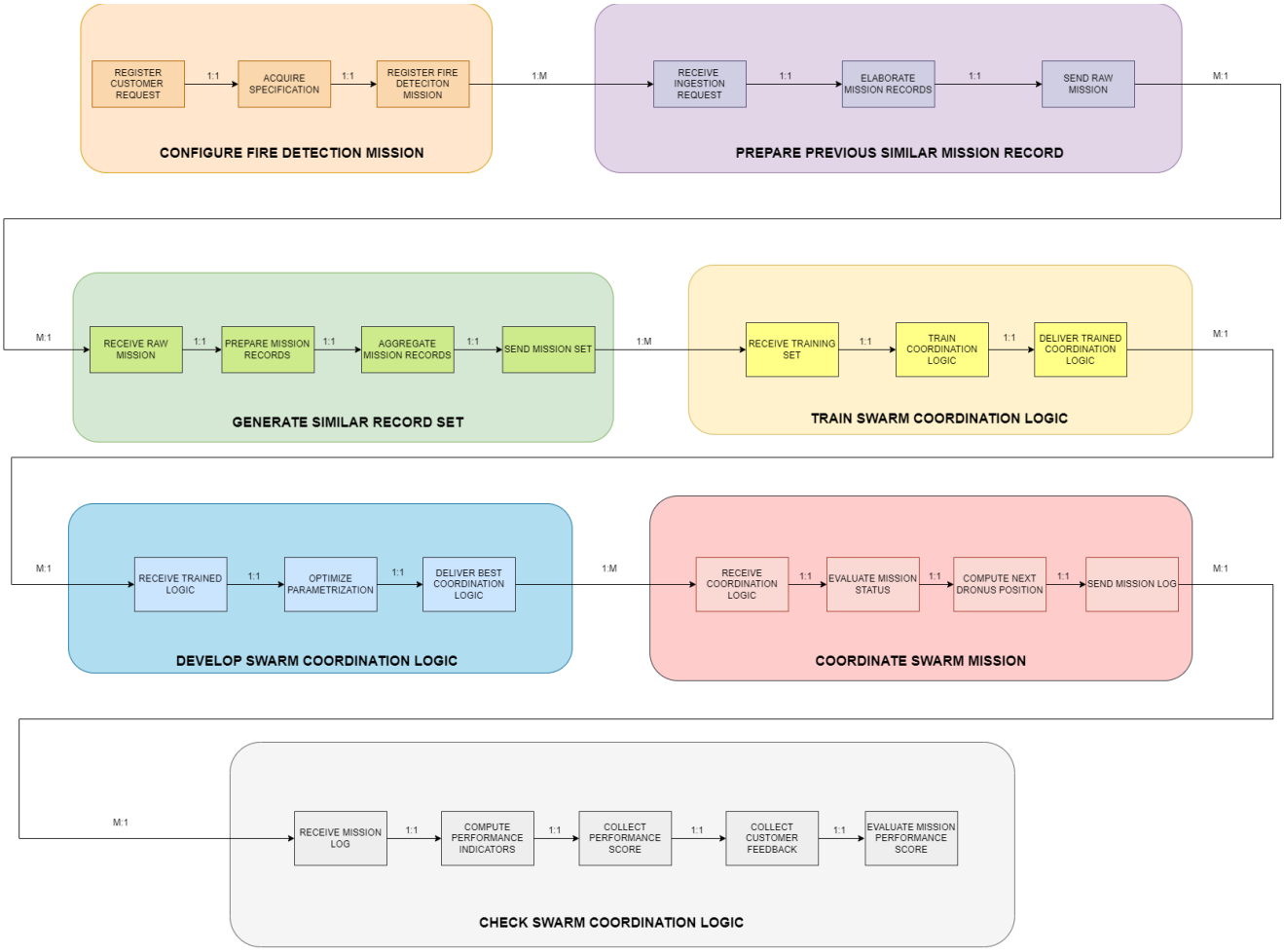
- Iperparametri (scelti dall'azienda):

1. N° DI DRONI
2. RAGGIO DI SCANSIONE (da cui dipende il costo dei droni)

- Fasi:

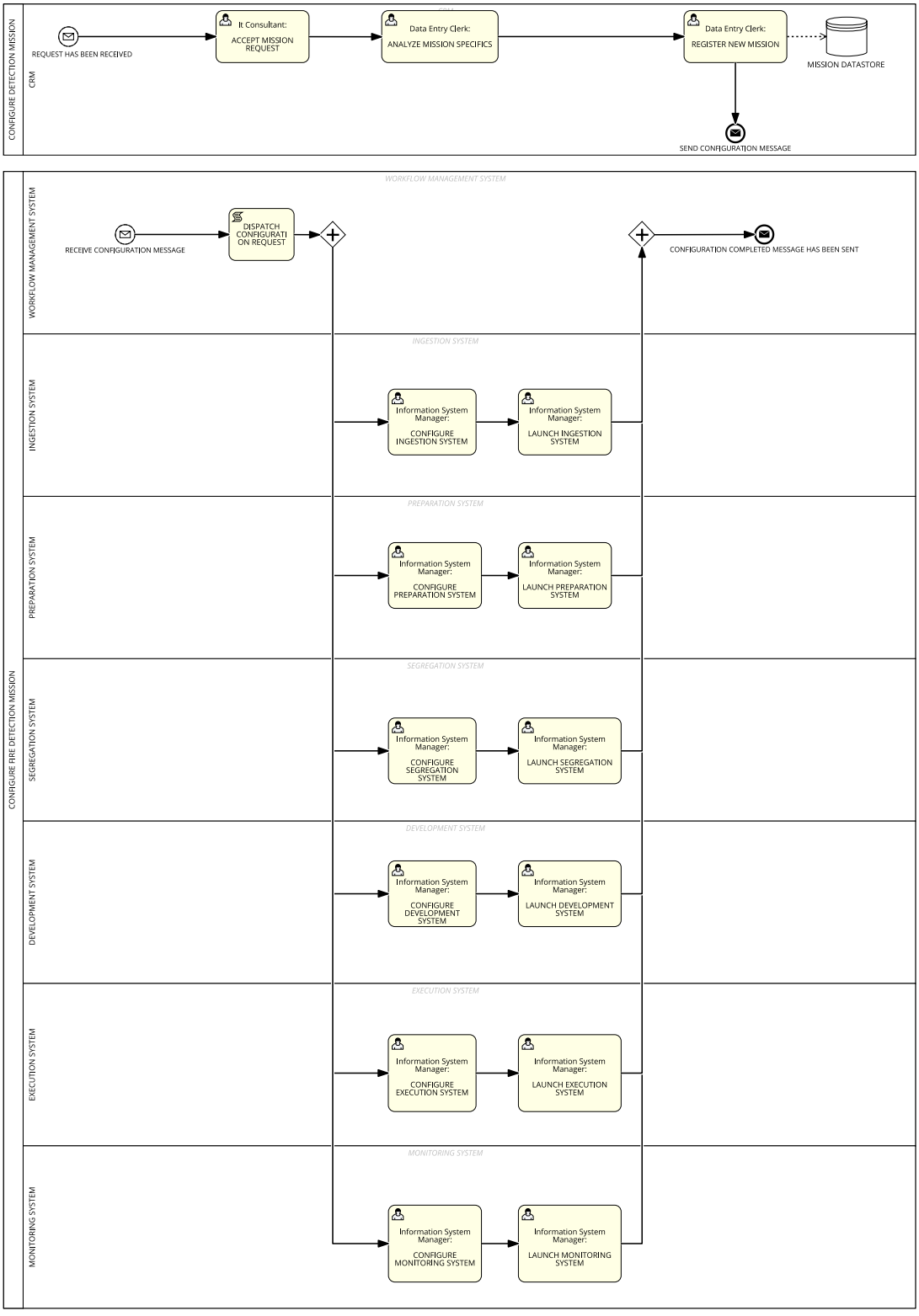
1. DEVELOPMENT PHASE
2. EXECUTION PHASE (3. MONITORING)

Process Landscape

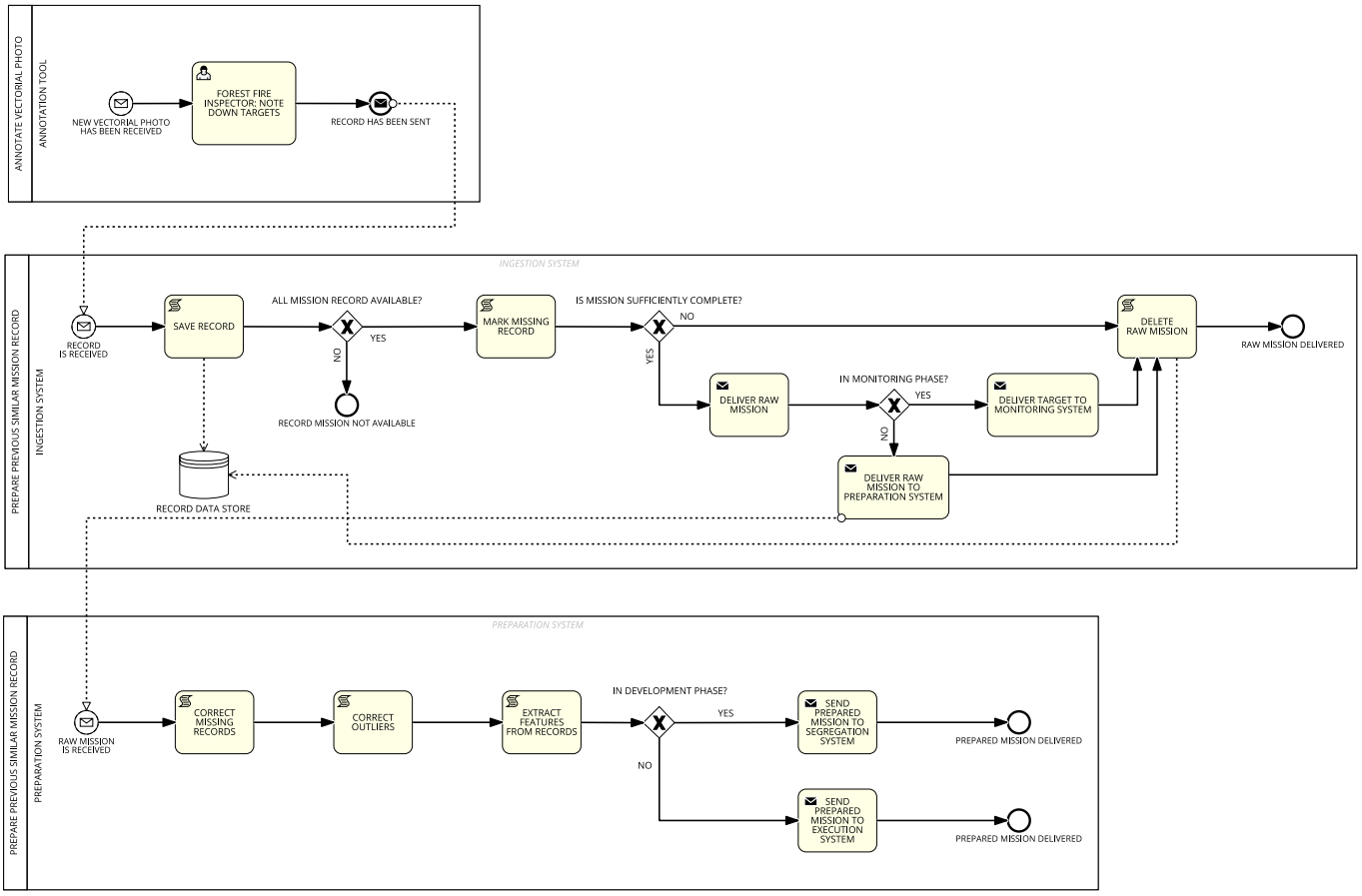


BPMN

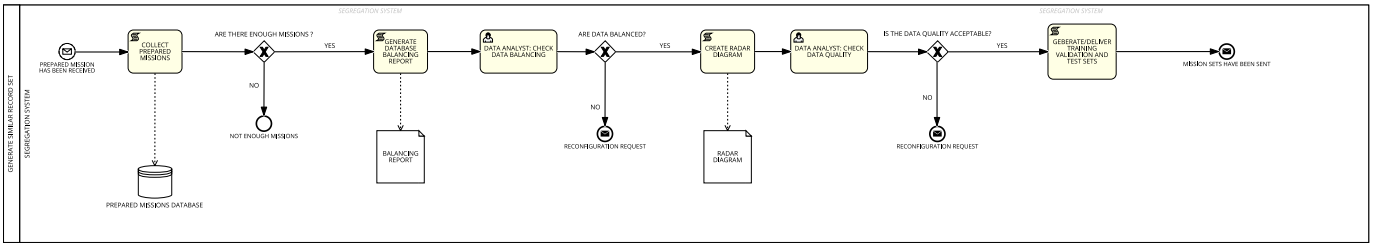
CONFIGURE FIRE DETECTION MISSION (Valerio Secondulfo)



PREPARE PREVIOUS SIMILAR MISSION RECORD (Stefano Dugo, Fabiano Pilia, Elisa De Filomeno only Annotation Tool)

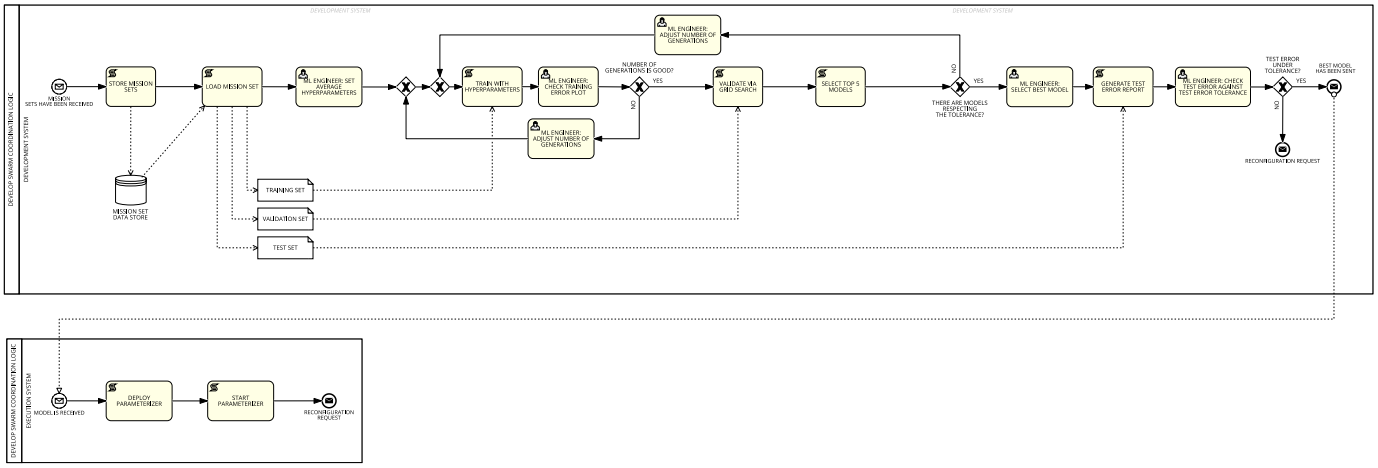


GENERATE SIMILAR RECORD SET (Elisa De Filomeno, Valerio Secondulfo)

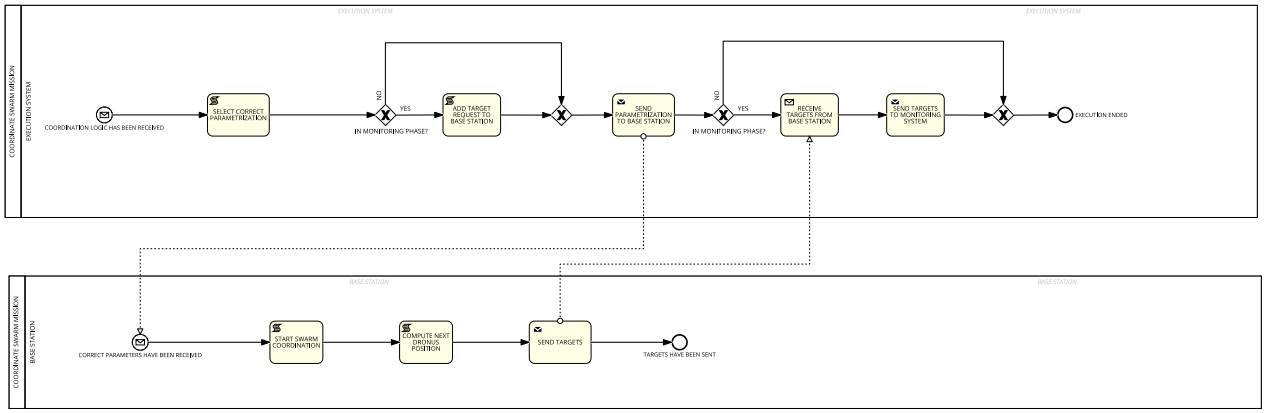


DEVELOP SWARM COORDINATION LOGIC (Fabiano Pilia)

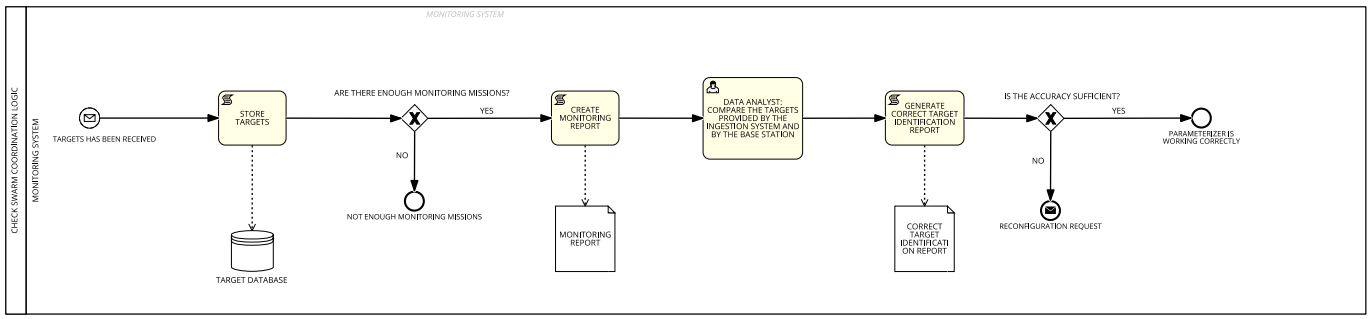
Assunzione: se la tolleranza non viene rispettata si ricomincia il training da capo, perché l'intervallo degli iperparametri è robusto e quindi assumiamo che l'errore sia dovuto all'overfitting



COORDINATE SWARM MISSION (Stefano Dugo)



CHECK SWARM COORDINATION LOGIC (Elisa De Filomeno)



Task Level

Accept mission request -> IT Consultant
Analyze mission specifics -> Data entry clerk
Register new mission -> Data entry clerk

Configure * system -> Information system manager
Launch * system -> Information system manager

Check data balancing -> Data analyst
Check data quality -> Data analyst

Set average hyperparameters -> ML Engineer
Check Training Error Plot -> ML Engineer
Adjust Number of Generations -> ML Engineer
Select best model -> ML Engineer
Check test error against test error tolerance -> ML engineer
Compare The targets provided by the ingestion... -> Data analyst

NORMALIZED SALARIES

Actor	Link	Salary	Normalized Salary
IT consultant	IT consultant job profile | Prospects.ac.uk	$53.4	5.34
Information Systems Manager	Information systems manager job profile | Prospects.ac.uk	$33.14	3.314
ML Engineer	Machine learning engineer job profile | Prospects.ac.uk	$68.21	6.821
Data Analyst	Data analyst job profile | Prospects.ac.uk	$16.3	1.63
Forest Fire Inspector	https://www.bls.gov/oes/current/oes332022.htm	$20.48	2.05
Data Entry Clerk	link	$10	1

Roles main tasks:

-IT consultant:

- work with clients to determine requirements and define the scope of a project

- Data Entry Clerk:

- Transfer data from paper formats into computer files or database systems using keyboards, data recorders or optical scanners
- Type in data provided directly from customers.
- Verify data by comparing it to source documents.
- Update existing data.
- Perform regular backups to ensure data preservation.

- Information Systems Manager:

- research and install new systems and networks.
- implement technology, directing the work of systems and business analysts, developers, support specialists and other computer-related workers.
- secure data from internal and external attack
- schedule upgrades and security backups of hardware and software systems
- evaluate user needs and system functionality, ensuring that IT facilities meet these needs.

- ML Engineer:

- use exceptional mathematical skills, in order to perform computations and work with the algorithms involved in this type of programming.
- analyze large, complex datasets to extract insights and decide on the appropriate technique.
- apply machine learning algorithms and libraries.
- produce project outcomes and isolate the issues that need to be resolved, in order to make programs more effective.

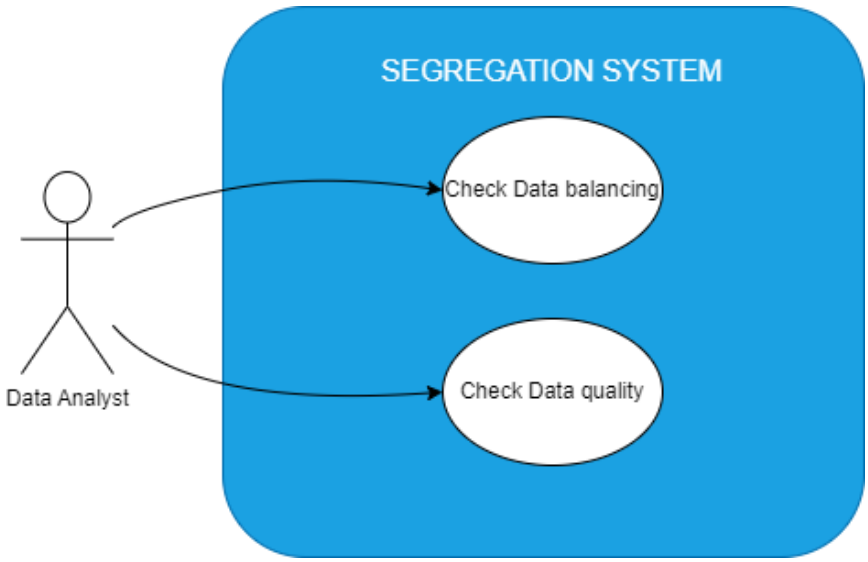
- Data Analyst:

- monitor and audit data quality.

-Forest Fire Inspector:

- Estimate sizes and characteristics of fires, and report findings to base camps by annotation tool.

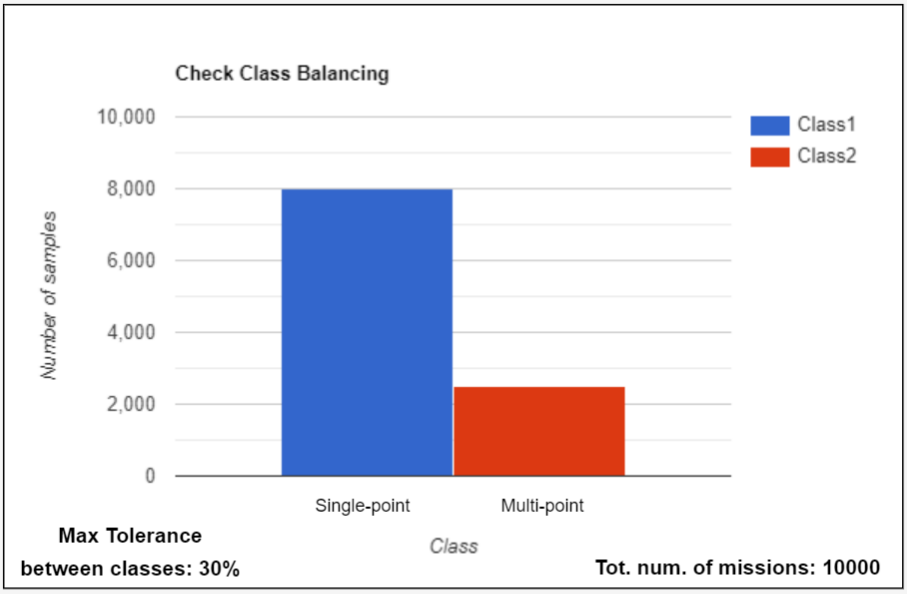
SEGREGATION SYSTEM



CHECK DATA BALANCING(Stefano Dugo)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	Data Analyst	Open the data balancing check report	Remember (1)	1	1*1*1.63 =1.63
2	System	Display the balancing report			
3	Data Analyst	Count difference of missions between the two classes	Apply (3)	1	3*1*1.63 = 4.89
4 – IF (60%)		Difference of samples is over the tolerance			
4.1	Data Analyst	Count number of missions per class	Apply (3)	0.6	3*0.6*1.63 = 2.934
4.2 - IF (60%)		Number of missions are not enough			
4.2.1	Data Analyst	Selects unbalanced class	Remember (1)	0.6*0.6	1*0.6*0.6*1.63 = 0.5868

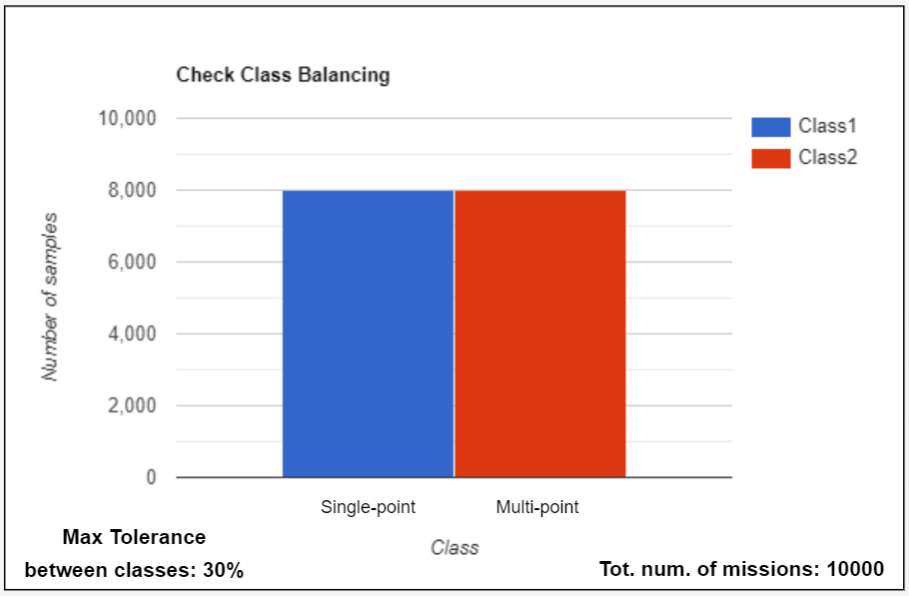
4.2.2	Data Analyst	Set number of missions required	Remember (1)	0.6*0.6	1*0.6*0.6*1.63 = 0.5868
4.2.3	Data Analyst	Clicks on “Send Unbalanced Data Notification”	Remember (1)	0.6*0.6	1*0.6*0.6*1.63 = 0.5868
4.2.4	System	Sends notification to reconfigure the system			
4.3 - ELSE (40%)					
4.3.1	Data Analyst	Remove excess missions	Remember (1)	0.6*0.4	1*0.6*0.4*1.63 = 0.3912
5 – ELSE (40%)					
5.1	Data Analyst	Click on the “Number of samples are enough” button	Remember (1)	0.4	1*0.4*1.63 = 0.652
				TOTAL COST	**12.2576**



Set unbalanced class: ☐ Single-point ☐ Multi-point

Number of samples required:

SEND UNBALANCED DATA NOTIFICATION | REMOVE SAMPLES | NUMBER OF SAMPLES ARE ENOUGH



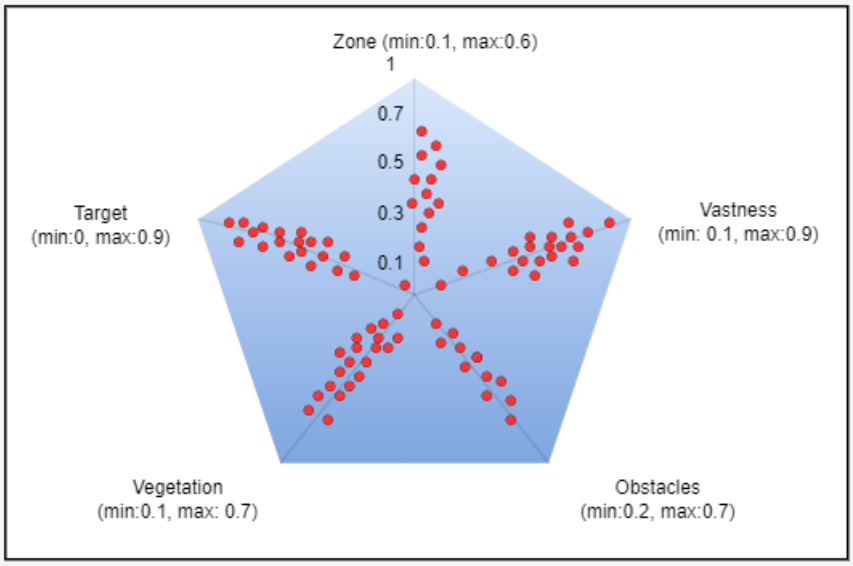
Set unbalanced class: ☐ Single-point ☐ Multi-point

Number of samples required:

SEND UNBALANCED DATA NOTIFICATION | REMOVE SAMPLES | NUMBER OF SAMPLES ARE ENOUGH

CHECK DATA QUALITY(Stefano Dugo)

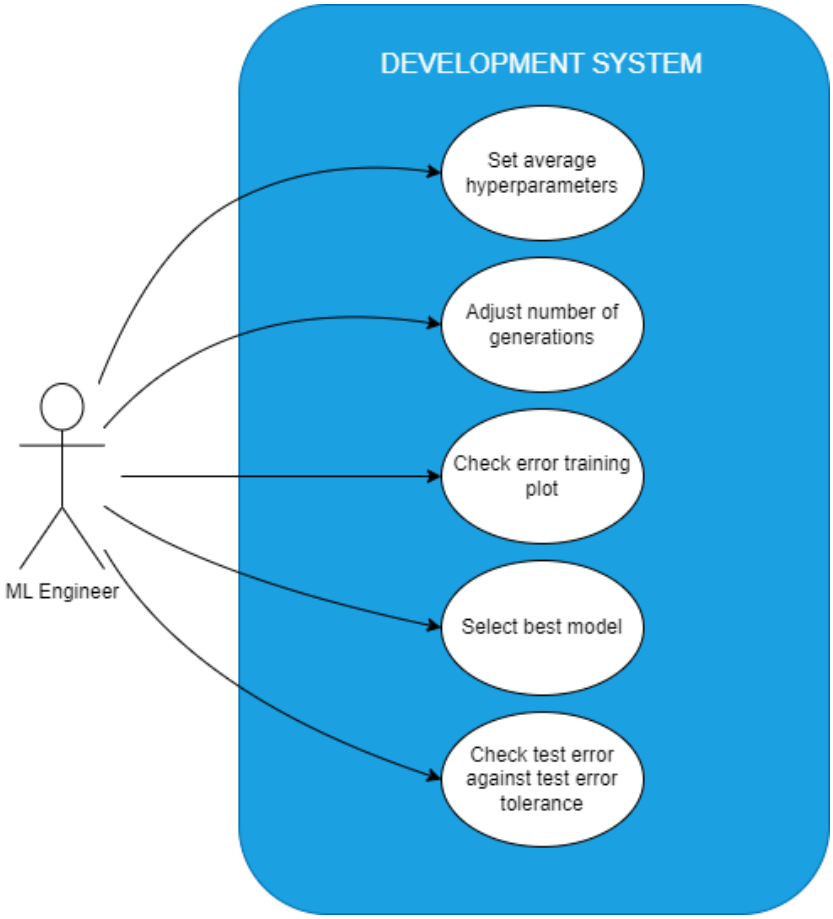
Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	Data Analyst	Open the data quality check report	Remember (1)	1	1*1*1.63 = 1.63
2	System	Display the quality report			
3 – FOR EACH					
3.1	Data Analyst	Check feature distribution based on the values it can get on this context	Analyze (4)	0.6*0.6	4*1.63 = 6.52
3.2 - IF (70%)		Feature can be correctly distributed by ingesting other data			
3.2.1	Data Analyst	Click on the “Send Features not correctly distributed Notification” button	Remember (1)	0.7*0.4	1*0.7*1.63 = 1.141
3.2.2	System	Sends notification to reconfigure the system			
3.2.3		->5			
3.3 – ELSE (30%)					
4	Data Analyst	Click on the button “Approve dataset”	Remember (1)	1	1*1*1.63 =1.63
5	System	END			
				TOTAL COST	**N = 1: 9.291 N = 5: 41.565 AVG: 25.428**



SEND FEATURE NOT CORRECTLY DISTRIBUTED NOTIFICATION

APPROVE DATASET

DEVELOPMENT SYSTEM

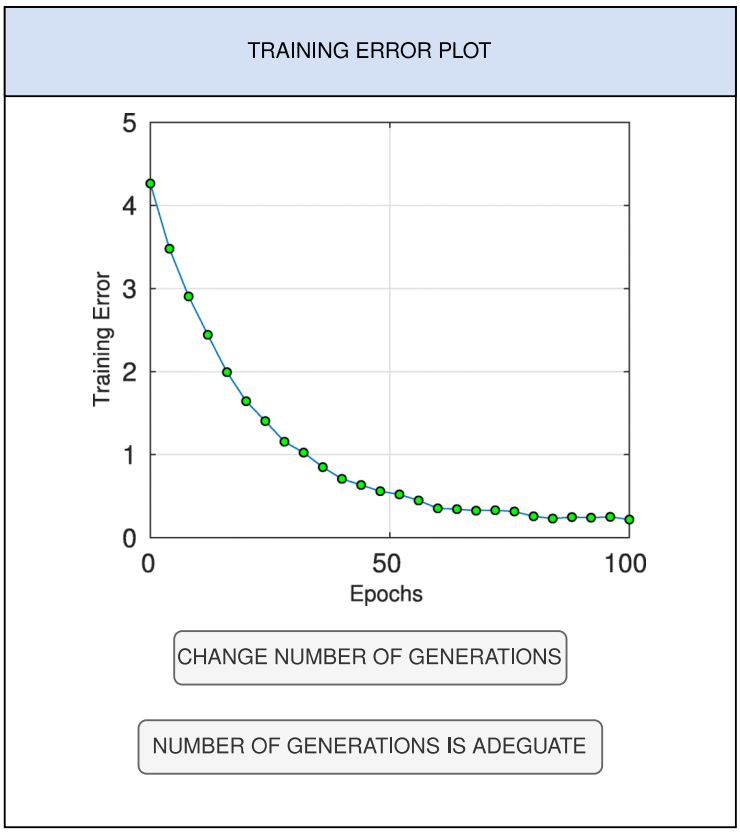


SET AVERAGE HYPERPARAMETERS (Fabiano Pilia)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	ML Engineer	Open the interface to set the hyperparameters	Remember (1)	1	1*1*6.821 = 6.821
2	System	Display the hyperparameters interface			
3	ML Engineer	Set the number of drones	Apply (3)	1	3*1*6.821 = 20.463
4	ML Engineer	Set the radius of scanning	Apply (3)	1	3*1* 6.821 = 20.463
5	ML Engineer	Click on “Start training” button	Remember (1)	1	1*1* 6.821 = 6.821
				TOTAL COST	**54.568**

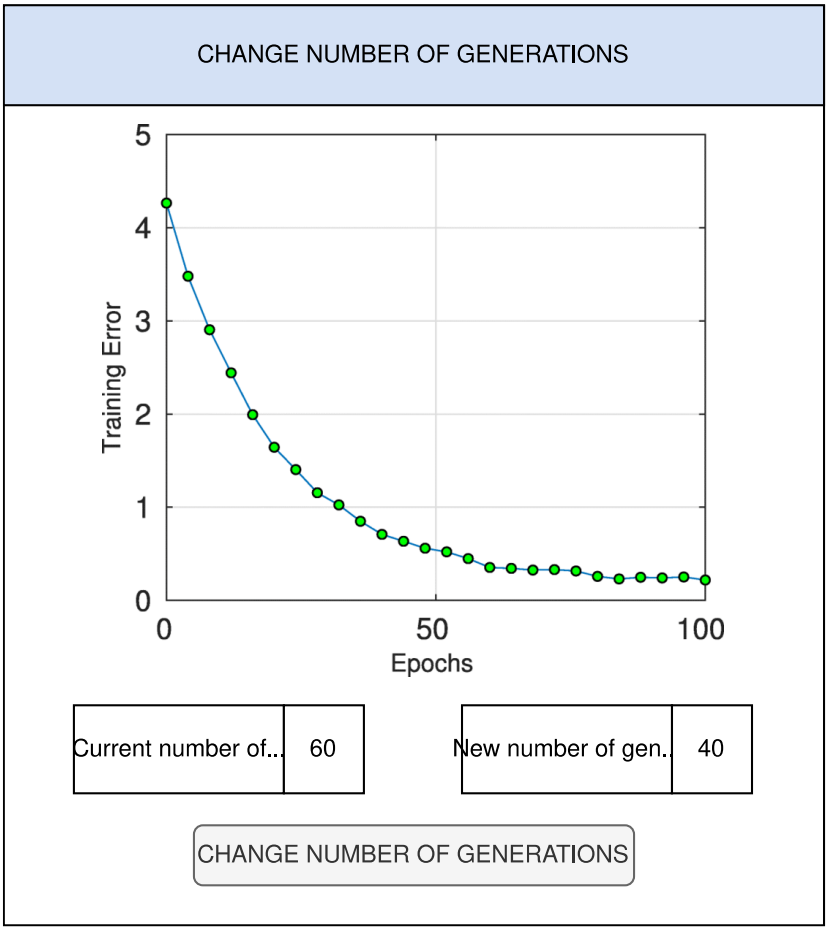
CHECK TRAINING ERROR PLOT (Fabiano Pilia)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	ML Engineer	Open the training error plot over the epochs	Remember (1)	1	1*1*6.821 = 6.821
2	System	Displays the loss over the last generations			
3	ML Engineer	Check if the plot shows overfitting or underfitting	Apply (3)	1	3*1*6.821 = 20.463
4-IF (70%)		If the plot shows overfitting or underfitting			
4.1		Click the “Change number of generations” button	Remember (1)	0.7	1*0.7* 6.821 = 4.775
5-ELSE (30%)		Else the plot shows a good fit			
5.1	ML Engineer	Click the “Number of generations is adequate” button	Remember (1)	0.3	1*0.3*6.821 = 2.046
6	System	END			
			Total Cost:		34.105



ADJUST NUMBER OF GENERATIONS (Stefano Dugo)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	System	Displays the loss over the last generations			
2– IF (50%)		IF the loss is flat for at least half of the generations			
2.1	ML Engineer	Reduce the number of generations by 1/3 to manage overfitting	Apply (3)	0.5	3*0.5*6.821= 10.2315
2.2	ML Engineer	Insert the updated number of generations	Remembe r (1)	0.5	1*0.5*6.821 = 3.4105
2.3	ML Engineer	Click the “Change number of generations” button	Remembe r (1)	0.5	1*0.5*6.821 = 3.4105
3 – ELSE (50%)		ELSE the loss is not flat at the end of the generations			
3.1	ML Engineer	Increase the number of generations by 1/3	Apply (3)	0.5	3*0.5*6.821 = 10.2315
3.2	ML Engineer	Insert the updated number of generations	Remembe r (1)	0.5	1*0.5*6.821 = 3.4105
3.3	ML Engineer	Click the “Change number of generations” button	Remembe r (1)	0.5	1*0.5*6.821 = 3.4105
4	System	END			
				TOTAL COST	**34.105**

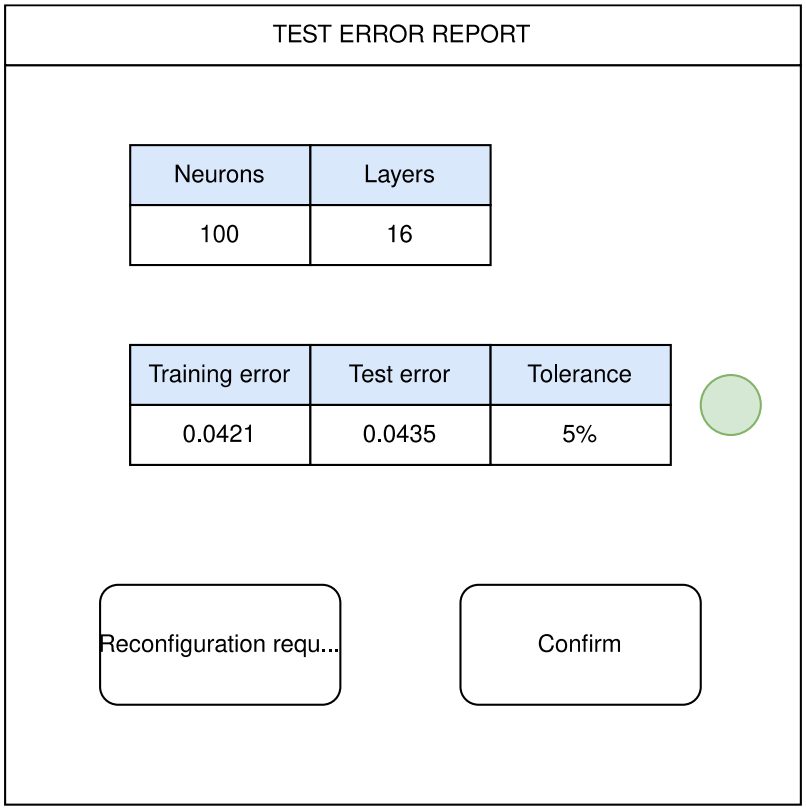


SELECT BEST MODEL (Fabiano Pilia)

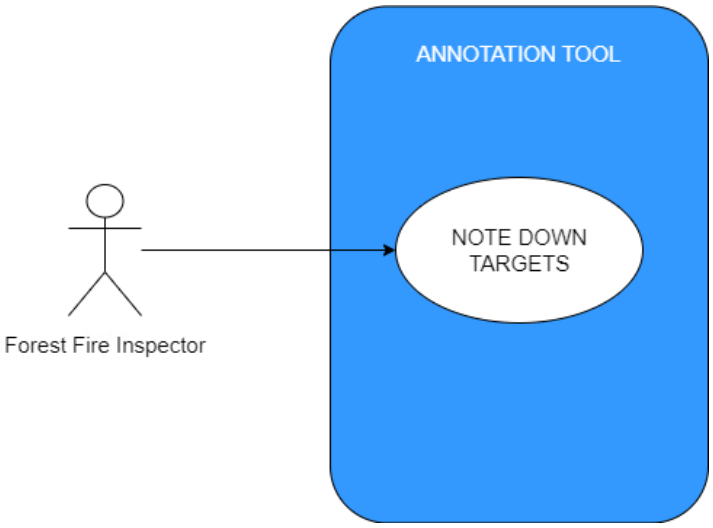
Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	ML Engineer	Open the list of the top 5 best models	Remember (1)	1	1*1*6.821 = 6.821
2	System	Display the validation error, test error, number of neurons and layers of the top 5 best models			
3	ML Engineer	Select the best model based on the tradeoff between the errors and the complexity	Apply (3)	1	3*1*6.821 = 20.463
4	System	END			
			Total Cost:		27.284

CHECK TEST ERROR AGAINST TEST ERROR TOLERANCE (Fabiano Pilia)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	ML Engineer	Open test error report	Remember (1)	1	1*1*6.821 = 6.821
2	System	Show test error report			
3	ML Engineer	Check if the test error is inside the tolerance threshold with respect to the training error	Apply (3)	1	3*1*6.821 = 20.463
4-IF (75%)		Error on the test set out of the tolerance threshold			
4.1	ML Engineer	Click on “reconfiguration request”	Remember (1)	0.75	1*0.75*6.821 = 5.116
4.2	System	Send reconfiguration message			
5-ELSE (25%)		Error on the test set within the tolerance threshold			
5.1	ML Engineer	Click on “Confirm” to mark the model as the best model	Remember (1)	0.25	1*0.25*6.821 = 1.705
5.2	System	Send best model			
6	System	END			
			Total Cost:		34.105



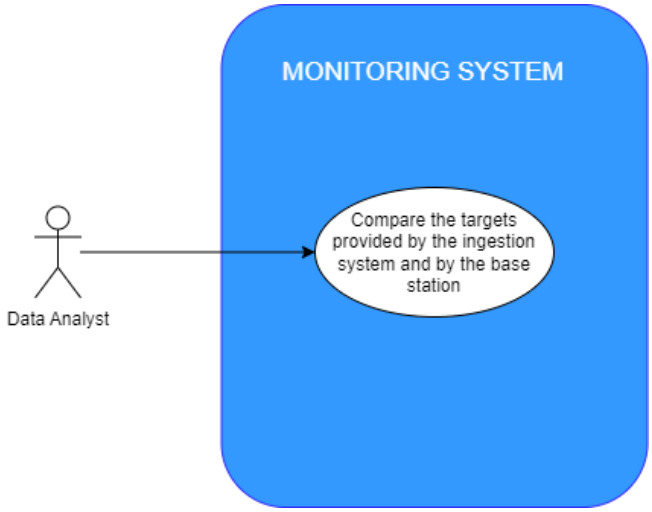
ANNOTATION TOOL



NOTE DOWN TARGETS (Elisa De Filomeno)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	Forest Fire Inspector	Open the vectorial photo	Remember (1)	1	1*1*2.05=2.05
2	Annotation Tool	Show the vectorial photo			
3 – FOR EACH		Area in the vectorial photo		4	
3.1	Forest Fire Inspector	Check at the color of the pixels in the area	Apply (3)	1*4	3*1*4*2.05=24.6
3.2 - IF (90%)		The pixel isn't red, orange or yellow			
3.2.1				0.9*4	
3.3 - ELSE (10%)					
3.3.1	Forest Fire Inspector	Press in the location of the pixel to annotate the target	Remember (1)	0.1*4	1*0.1*4*2.05=0.82
4	Forest Fire Inspector	Press the "send record" button	Remember (1)	1	1*1*2.05=2.05
5	Annotation Tool	END			
			Total Cost:		29.52

MONITORING SYSTEM



COMPARE THE TARGETS PROVIDED BY THE INGESTION SYSTEM AND BY THE BASE STATION (Elisa De Filomeno)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	Data Analyst	Open the Monitoring Report	Remember (1)	1	1*1* 1.63 =1.63
2	System	Shows Monitoring Report			
3	Data Analyst	Compare the average percentage of targets correctly found in each type of mission with the threshold for that type mission	Understand (2)	1	2*1* 1.63 =3.26
4-IF (70%)		The percentage is lower than the threshold			
4.1	Data Analyst	Press the “YES” button	Remember (1)	0.7	1*0.7* 1.63 =1.14
5-ELSE (30%)		The percentage is higher than the threshold			
5.1	Data Analyst	Press the “NO” button	Remember (1)	0.3	1*0.3* 1.63 =0.49
6	System	END			
			Total Cost:		6.52

- M: wildfire with multiple outbreaks
- S: wildfire with a single outbreak

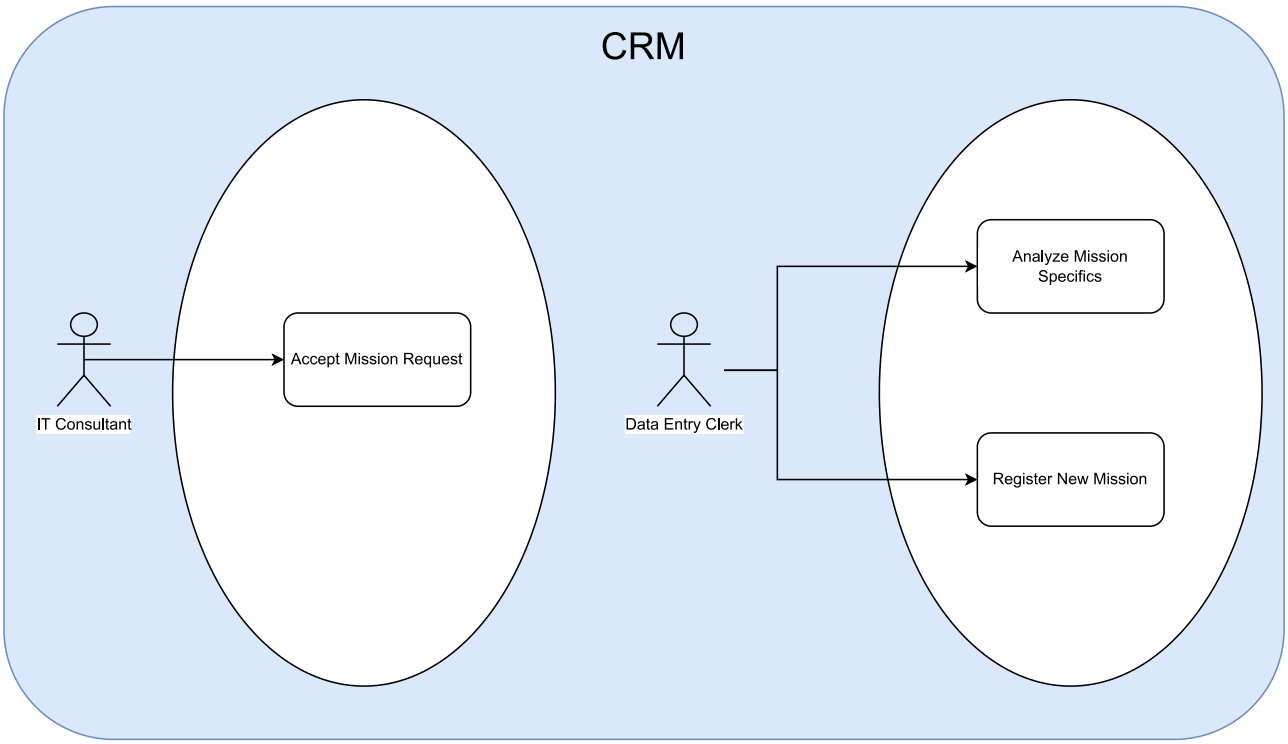
Mission	Type	Perc. of targets correctly found
1	M	74%
2	S	81%
3	M	69%
4	M	91%
5	S	85%

Type	Avg perc. of targets correctly found	Threshold
M	78%	75%
S	83%	80%

Is the accuracy enough?

YES **NO**

Customer Relationship Management CRM (Secondulfo Valerio)



ACCEPT MISSION REQUEST

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	Client	Present mission request			
2	It Consultant	Ask for mission details	Remember(1)	1	1*1*5.34=5.34
3	Client	Present mission details			
4	It Consultant	Present mission contract	Remember(1)	1	1*1*5.34=5.34
5	Client	Accept and sign mission contract			
				TOTAL COST	10,68

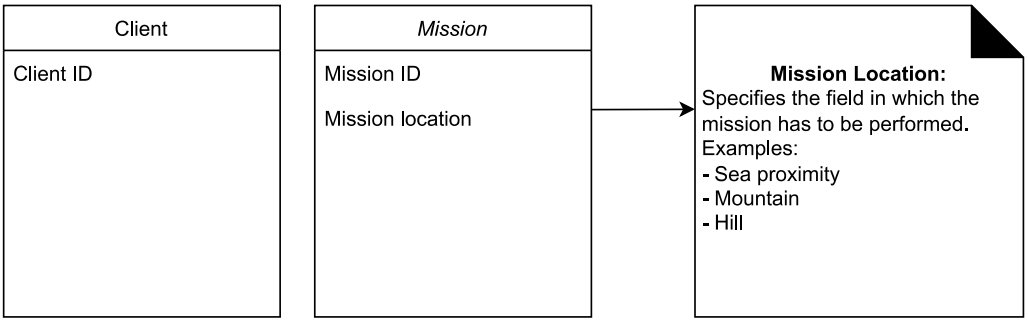
ANALYZE MISSION SPECIFICS (Valerio Secondulfo)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	Data Entry Clerk	Ask client for mission specifics	Remember(1)	1	1*1*1=1
2	Client	Send mission specifics			
3	Data Entry Clerk	Check mission specifics quality and availability	Understand(2)	1	2*1*1=2
4-IF (10%)		Mission specifics and information missing			
4.1	Data Entry Clerk	Ask client for missing specifics	Remember(1)	0.10	1*0,1*1=0,1
5-ELSE (90%)		Mission specifics are correct			
5.1	Data Entry Clerk	Open Mission Registration System	Remember(1)	0.90	1*0.9*1=0.9
				TOTAL COST	4

REGISTER NEW MISSION (Valerio Secondulfo)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1	Data entry clerk	Enter user credentials and select Log In	Remember(1)	1	1*1*1=1
2	Mission Registration System	Display the interface for missions management			
3	Data entry clerk	Select “Register new mission”	Remember(1)	1	1*1*1=1
4	Mission Registration System	Ask for mission informations			
4.1	Data entry clerk	Enter mission informations	Remember(1)	1	1*1*1=1
4.2	Mission Registration System	Ask for registration confirmation			

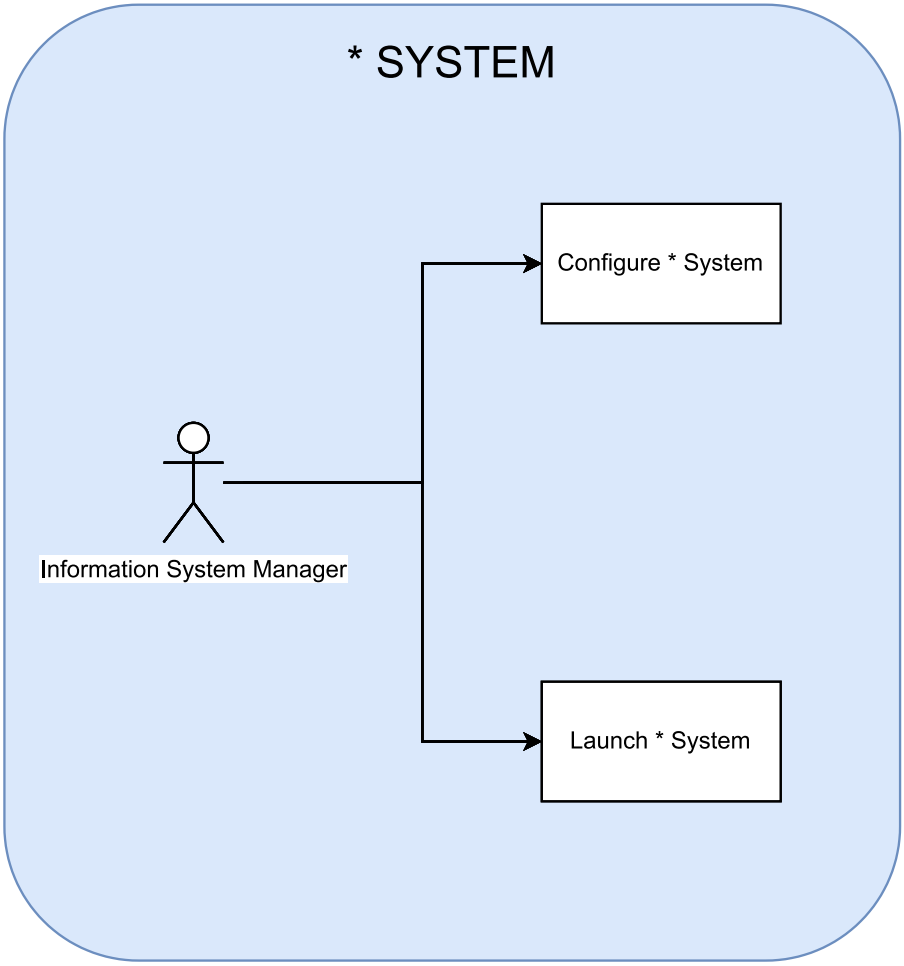
4.3	Data entry clerk	Confirm new mission registration	Remember(1)	1	1*1*1=1
5	Mission Registration System	Add new mission			
5.1	Mission Datastore	Return “Add new mission confirmation”			
6	Mission Registration System	Send configuration message			
7	Mission Registration System	Display registered mission page			
8	Data entry clerk	Log out and exit the system	Remember(1)	1	1*1*1=1
				TOTAL COST	5



Systems Configuration (Valerio Secondulfo)

*Configuration and launch processes of the

- Ingestion System
- Preparation System
- Segregation System
- Development System
- Execution System
- Monitoring System



CONFIGURE * SYSTEM (Valerio Secondulfo)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1-IF (30%)		System is already configured			
1.1	Information System Manager	Reconfigure parameters and thresholds	Remember(1)	0,3	1*0,3*3.314=0,9942
1.2	Information System Manager	Send “system reconfigured” message to the other systems	Remember(1)	0.3	1*0.3*3.314=0,9942
2-ELSE (70%)					
2.1	Information System Manager	Boot the system	Remember(1)	0.7	1*0.7*3.314=2,3198

2.2	Information System Manager	Set initial parameters	Remember(1)	0.7	1*0.7*3.314=2,3198
2.3	Information System Manager	Set parameter's thresholds	Remember(1)	0.7	1*0.7*3.314=2,3198
2.4	Information System Manager	Connect the system with the other systems	Remember(1)	0.7	1*0.7*3.314=2,3198
2.5	Information System Manager	Select "Launch system"	Remember(1)	0.7	1*0.7*3.314=2,3198
				TOTAL COST	13,5874

LAUNCH * SYSTEM (Valerio Secondulfo)

Sub-Task	Actor	Action	Cognitive Effort	Occurrence	Cost
1-IF (30%)		System is already launched and configured			
1.1	Information System Manager	Send "system reconfigured" message to the other systems	Remember(1)	0.3	1*0.3*3.314=0,9942
2-ELSE (70%)					
2.1	Information System Manager	Select "Launch system"	Remember(1)	0.7	1*0.7*3.314=2,3198
				TOTAL COST	3,314

SUMMARY FINAL COSTS

HUMAN TASK	COST
CHECK DATA BALANCING	12.2576
CHECK DATA QUALITY	N = 1: 9.291 N = 5: 41.565 AVG: 25.428
SET AVERAGE HYPERPARAMETERS	54.568
CHECK TRAINING ERROR PLOT	34.105
ADJUST NUMBER OF GENERATIONS	34.105
SELECT BEST MODEL	27.284
CHECK TEST ERROR AGAINST TEST ERROR TOLERANCE	34.105
NOTE DOWN TARGETS	29.52
COMPARE THE TARGETS PROVIDED BY THE INGESTION SYSTEM AND BY THE BASE STATION	6.52
ACCEPT MISSION REQUEST	10.68
ANALYZE MISSION SPECIFICS	4
REGISTER NEW MISSION	5
CONFIGURE * SYSTEM	13.5874
LAUNCH * SYSTEM	3.314

BIMP Simulation

-AS-IS

Assumptions:

-number of missions: 10000

-number of models to be deployed: 5

-number of missions to be registered: 5

-prob. Registration: 5/10.000

-number of trainings per model: 3

-number of deployments: 5*3

-number of missions to train a model: 1000

-number of missions in database 5000

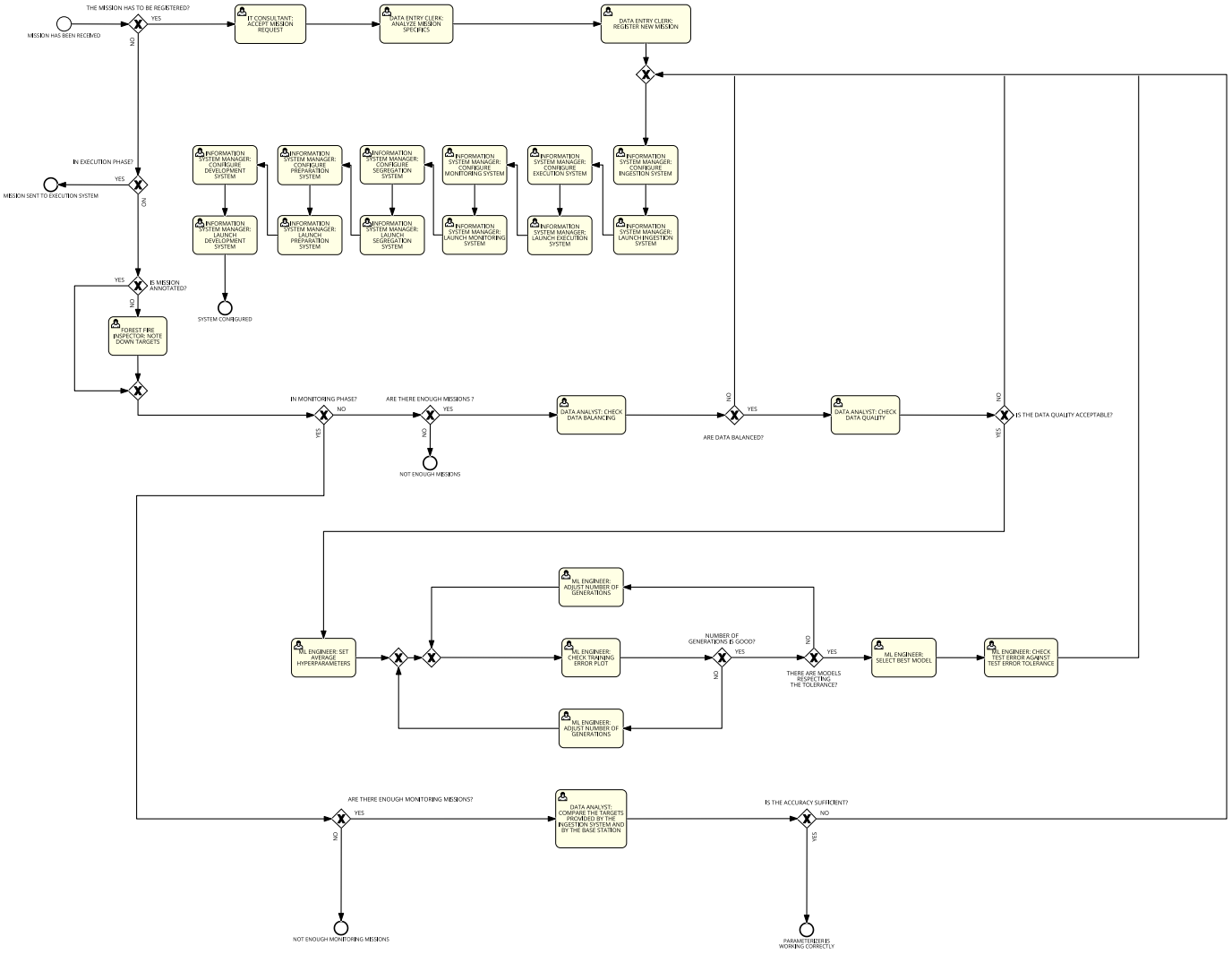
-number of missions in monitoring phase: 5*100

-number of missions in development phase: 5*1000

-number of missions in execution phase: tot – mission_in_dev – mission_in_mon = 4500

Other simulation assumptions:

- We added one gateway at the start to specify if a mission is registered or not, and, in that case, we need to configure the system, otherwise the mission goes to the ingestion system.
- We added a gateway before the note down task, since we assume that we have 5000 missions already in the database and these missions are annotated.
- The configure and launch tasks for each system are here shown in series and not in parallel because we cannot insert a parallel gateway in the simulation.
- After the task “Check test error against test error tolerance” we go to the reconfiguration system because in one case we need to reconfigure because the test error is below the tolerance, in the other cases we also need to send a reconfiguration message but only for setting a flag to switch from development phase to monitoring/execution phase.
- We have missions as tokens, so we removed from the simulation diagram the gateway that checks if there are missing records, because in that case we would discard complete missions, and we assumed that the token is complete.
- We shifted the gateway of the check of the execution phase before the annotation of the images because, if we are in execution phase, we don’t need to annotate the image.



PERCENTAGES GATEWAYS:

GATEWAY	YES	NO
THE MISSION HAS TO BE REGISTERED?	**0.05%**	**99.95%**
IN EXECUTION PHASE?	**45%**	**55%**
IS MISSION ANNOTATED?	**91%**	**9%**
IN MONITORING PHASE?	**9%**	**91%**
ARE THERE ENOUGH MISSIONS?	**0.01%**	**99.9%**

ARE DATA BALANCED?	**75%**	**25%**
IS THE DATA QUALITY ACCEPTABLE?	**90%**	**10%**
NUMBER OF GENERATIONS IS GOOD?	**34%**	**66%**
THERE ARE MODELS RESPECTING THE TOLERANCE?	**60%**	**40%**
ARE THERE ENOUGH MONITORING MISSIONS?	**1%**	**99%**
IS THE ACCURACY SUFFICIENT?	**98%**	**2%**

- "THE MISSION HAS TO BE REGISTERED?": we assumed that we want to build 5 models, so we need 5 registrations, so we need to register a mission approximately only 0.05% of the time

YES: **5/10000 = 0.0005 = 0.05%**

NO: 1 – YES = **99.95%**

- "IN EXECUTION PHASE?": we assumed that we have 4500 missions going in execution phase, the total of mission evaluated are 9995 (total missions – the ones that must be registered)

YES: **4500/9995 = 0.4502 ~= 0.45 = 45%**
NO: 1 – YES = **55%**

- "IS MISSION ANNOTATED?": Since we have **5500** missions going in monitoring or development phase and we've assumed that 5000 mission are already present in the database, then only 500 of the overall mission must be annotated, so we've obtained that:

YES: **5000/5500 = 0.9090 => 0.91 = 91%**
NO: **1 – YES = 0.09 = 9%**

- "IN MONITORING PHASE?": we assumed that **5000** missions are used in the development phase to build the training, 1000 for each model so

5*1000 = 5000, validation and test sets. Only 500 will be used to monitor the model accuracy, 100 for each model => **5*100 = 500**.

YES: **500/5500 = 0.09 = 9%**
NO: **1 – YES = 0.91 = 91%**

- "ARE THERE ENOUGH MISSIONS?": Since we've assumed the need of **1000** missions for the development of a model, then only **1** token every **1000** passes through the yes branch.

YES: **1/1000 = 0.001 = 0.1%**
NO: **1 – YES = 0.999 = 99.9%**

- "ARE DATA BALANCED?": we assumed that there is a prob of 75% that in the database we have for both classes a difference of missions below the threshold, which is 30%. The unbalancing is given from the missions that are not in the database.

YES: **75%**
NO: **25%**

- "IS THE DATA QUALITY ACCEPTABLE?": considering that most of data are taken from the database there is small probability that the missions are not of good quality. The data that are not of good quality are the ones that are not in the database "prepared missions database".

YES: **90%**
NO: **10%**

- "NUMBER OF GENERATIONS IS GOOD?": we assumed that there is **1/3** of cases that there is underfitting, **1/3** overfitting and **1/3** of cases are good.

YES: **34%**
NO: **66%**

- “THERE ARE MODELS RESPECTING THE TOLERANCE?”: we assumed that from the 5 best models at least 3 are respecting the tolerance, considering a tolerance value large enough (ex. 5%).

YES: **60%**
NO: **40%**

- “ARE THERE ENOUGH MONITORING MISSIONS?”: we decided to wait to have enough missions in order to proceed to the testing of the monitoring mission. For the monitoring are needed **100** tokens, so **1 token** will pass every 100 tokens.

YES: **1%**
NO: **99%**

- “IS THE ACCURACY SUFFICIENT?”: the thresholds of the accuracies to be accepted by the data analyst are 75% and 80% for the 2 types of wildfires. We assumed that almost of the time (98%) the accuracies are over the threshold.

YES: **98%**
NO: **2%**

RESULTS AS-IS:

Simulation Results

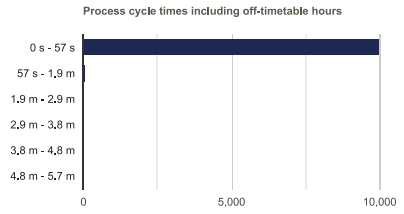
General information

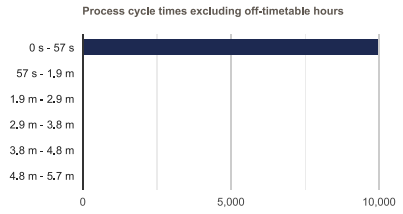
Completed process instances 10000

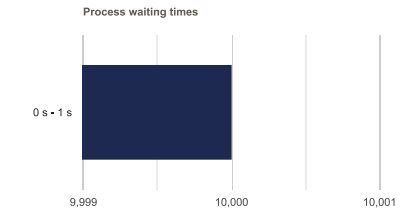
Total cost 0 EUR

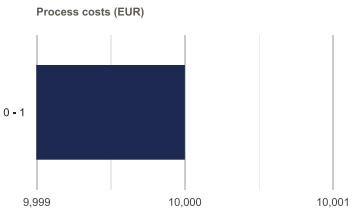
Total simulation time 52 weeks

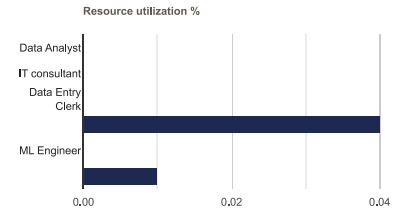
Charts











The interarrival time used was 750.

Scenario Statistics

	Minimum	Maximum	Average
Process instance cycle times including off-timetable hours	0 seconds	9.4 minutes	2 seconds
Process instance cycle times excluding off-timetable hours	0 seconds	9.4 minutes	2 seconds
Process instance costs	0 EUR	0 EUR	0 EUR

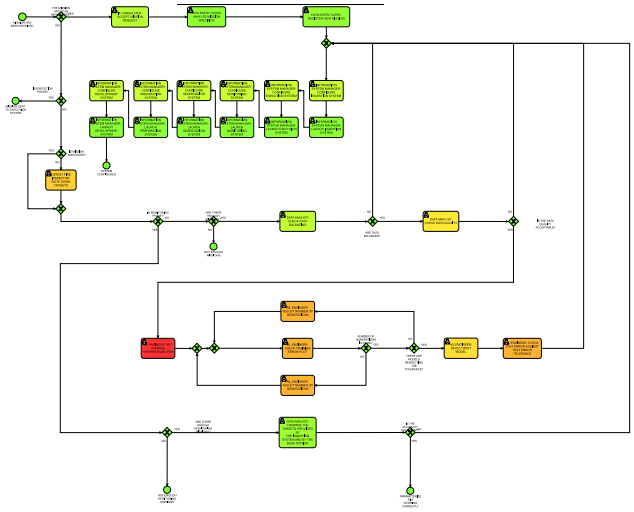
Activity Durations, Costs, Waiting times, Deviations from Thresholds

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: CHECK DATA BALANCING	4	0 s	0 s	0 s	12.3 s	12.3 s	12.3 s	0 s	0 s	0 s	0	0	0	0	0	0
DATA ANALYST: CHECK DATA QUALITY	4	0 s	0 s	0 s	25.4 s	25.4 s	25.4 s	0 s	0 s	0 s	0	0	0	0	0	0
DATA ANALYST: COMPARE THE TARGETS PROVIDED BY THE INGESTION SYSTEM AND BY THE BASE STATION	2	0 s	0 s	0 s	6.5 s	6.5 s	6.5 s	0 s	0 s	0 s	0	0	0	0	0	0
DATA ENTRY CLERK:
ANALYZE MISSION SPECIFICS	37	0 s	0 s	0 s	4 s	4 s	4 s	0 s	0 s	0 s	0	0	0	0	0	0
DATA ENTRY CLERK:
REGISTER NEW MISSION	37	0 s	0 s	0 s	5 s	5 s	5 s	0 s	0 s	0 s	0	0	0	0	0	0
FOREST FIRE INSPECTOR: NOTE DOWN TARGETS	479	0 s	0 s	0 s	29.5 s	29.5 s	29.5 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE DEVELOPMENT SYSTEM	41	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE EXECUTION SYSTEM	41	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE INGESTION SYSTEM	41	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE MONITORING SYSTEM	41	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE PREPARATION SYSTEM	41	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE SEGREGATION SYSTEM	41	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH DEVELOPMENT SYSTEM	41	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH EXECUTION SYSTEM	41	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH INGESTION SYSTEM	41	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH MONITORING SYSTEM	41	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH PREPARATION SYSTEM	41	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH SEGREGATION SYSTEM	41	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
IT CONSULTANT:
ACCEPT MISSION REQUEST	37	0 s	0 s	0 s	10.7 s	10.7 s	10.7 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: ADJUST NUMBER OF GENERATIONS	7	0 s	0 s	0 s	34.1 s	34.1 s	34.1 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: ADJUST NUMBER OF GENERATIONS	1	0 s	0 s	0 s	34.1 s	34.1 s	34.1 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: CHECK TEST ERROR AGAINST TEST ERROR TOLERANCE	4	0 s	0 s	0 s	34.1 s	34.1 s	34.1 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: CHECK TRAINING ERROR PLOT	12	0 s	0 s	0 s	34.1 s	34.1 s	34.1 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: SELECT BEST MODEL	4	0 s	0 s	0 s	27.3 s	27.3 s	27.3 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: SET AVERAGE HYPERPARAMETERS	4	0 s	0 s	0 s	54.6 s	54.6 s	54.6 s	0 s	0 s	0 s	0	0	0	0	0	0

HEATMAP AS-IS:

Heatmap

Heatmap based on Durations

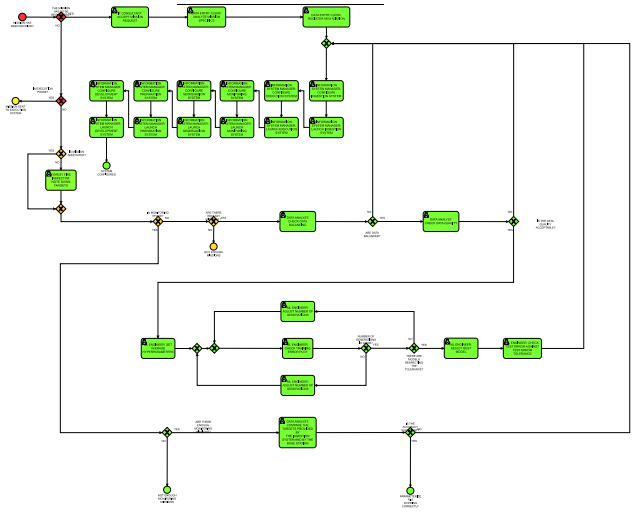


Legend

Color	Value
	0 s
	6 s
	12 s
	18 s
	24 s
	30 s
	36 s
	42 s
	49 s
	55 s

Heatmap

Heatmap based on Counts



Legend

Color	Value
	0
	1111
	2222
	3333
	4444
	5556
	6667
	7778
	8889
	10000

TO BE IMPROVEMENTS:

-handoff level:

Data balancing: if data are not balanced take data from similar clients to balance them and we go on instead of going back and reconfigure the system.
Consequence: change the percentage of “Are data balanced?” from yes: 75%, no: 25% to yes: 90% no: 10%.

-service level:

Set average hyperparameters: skip hyperparameterization since we assume that starting from the hyperparameter of similar missions we can infer the optimal set of hyperparameters for the current mission. This is possible by usings a software that automatically computes the best set of hyperparameters and set them.

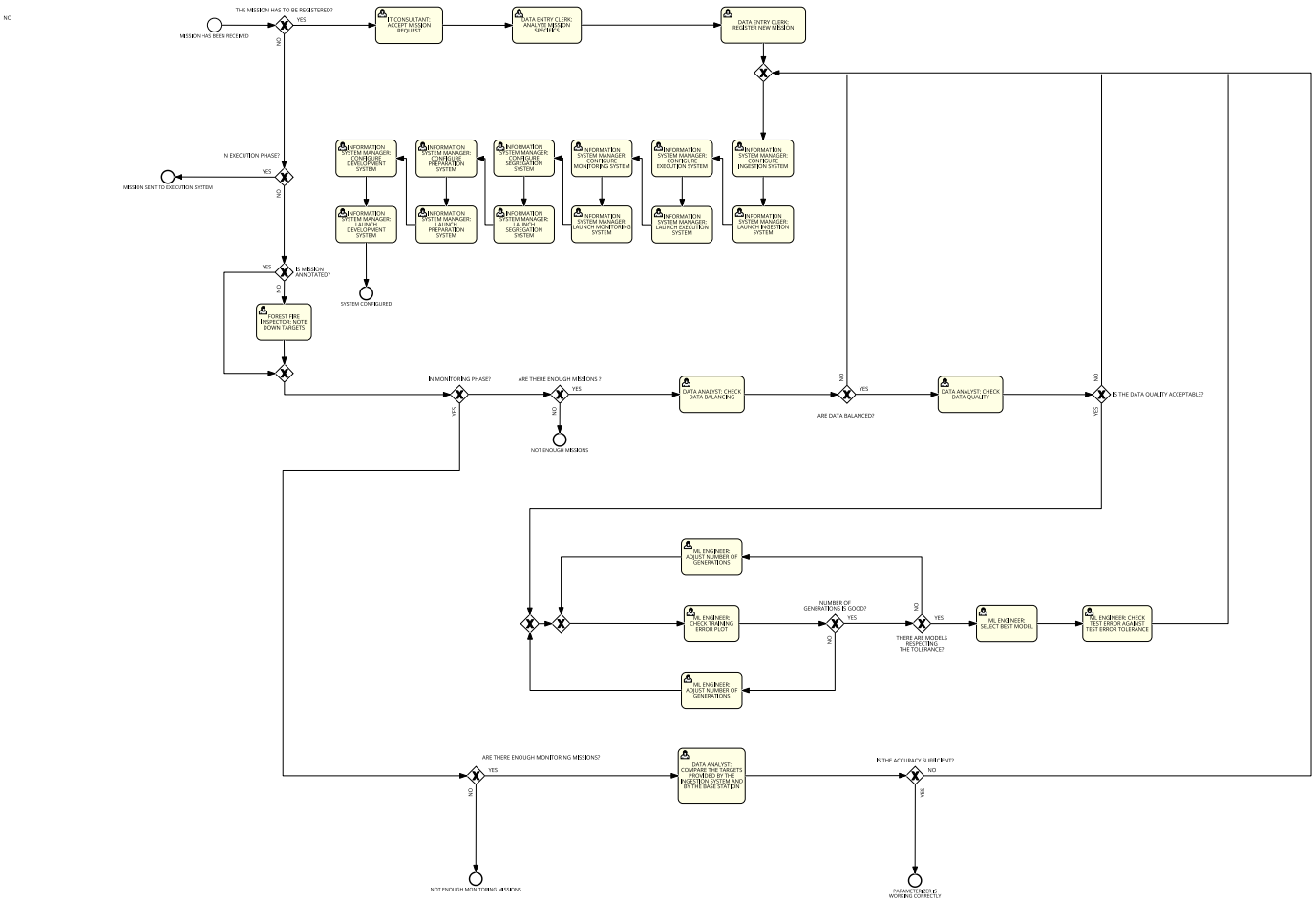
-task level:

Adjust number of generations: We’ve reduced the cognitive effort from 3 to 2, since we assume that the computation of the new number of generations is done by the system. In this way the system decides if it must increase or decrease the number of generations simplifying the task to just a visual inspection of the suggested epochs and to accept or to decline the new number. The updated cost for the task is 27.284.

(Other possible modifications)

- Skip selection of the best model.
- Remove the reconfiguration after data quality.
- Reduce the cognitive effort in the check test error.

-TO-BE:


NO
THE MISSION HAS TO BE REGISTERED?
MISSION HAS BEEN RECEIVED
YES
NO
IT CONSULTANT: ACCEPT MISSION REQUEST
DATA ENTRY CLERK: ANALYZE MISSION SPECIFICS
DATA ENTRY CLERK: REGISTER NEW MISSION
IN EXECUTION PHASE?
YES
NO
MISSION SENT TO EXECUTION SYSTEM
INFORMATION SYSTEM MANAGER: CONFIGURE DEVELOPMENT SYSTEM
INFORMATION SYSTEM MANAGER: CONFIGURE PREPARATION SYSTEM
INFORMATION SYSTEM MANAGER: CONFIGURE SEGREGATION SYSTEM
INFORMATION SYSTEM MANAGER: CONFIGURE MONITORING SYSTEM
INFORMATION SYSTEM MANAGER: CONFIGURE EXECUTION SYSTEM
INFORMATION SYSTEM MANAGER: CONFIGURE INGESTION SYSTEM
INFORMATION SYSTEM MANAGER: LAUNCH DEVELOPMENT SYSTEM
INFORMATION SYSTEM MANAGER: LAUNCH PREPARATION SYSTEM
INFORMATION SYSTEM MANAGER: LAUNCH SEGREGATION SYSTEM
INFORMATION SYSTEM MANAGER: LAUNCH MONITORING SYSTEM
INFORMATION SYSTEM MANAGER: LAUNCH EXECUTION SYSTEM
INFORMATION SYSTEM MANAGER: LAUNCH INGESTION SYSTEM
SYSTEM CONFIGURED
IS MISSION ANNOTATED?
FOREST FIRE INSPECTION: NOTE DOWN TARGETS
IN MONITORING PHASE?
ARE THERE ENOUGH MISSIONS ?
NOT ENOUGH MISSIONS
DATA ANALYST: CHECK DATA BALANCING
ARE DATA BALANCED?
DATA ANALYST: CHECK DATA QUALITY
IS THE DATA QUALITY ACCEPTABLE?
ML ENGINEER: ADJUST NUMBER OF GENERATIONS
ML ENGINEER: CHECK TRAINING ERROR PLOT
NUMBER OF GENERATIONS IS GOOD?
THERE ARE MODELS RESPECTING THE TOLERANCE?
ML ENGINEER: SELECT BEST MODEL
ML ENGINEER: CHECK TEST ERROR AGAINST TEST ERROR TOLERANCE
ML ENGINEER: ADJUST NUMBER OF GENERATIONS
ARE THERE ENOUGH MONITORING MISSIONS?
NOT ENOUGH MONITORING MISSIONS
DATA ANALYST: COMPARE THE TARGETS FROM THE MISSION SPECIFICATION SYSTEM AND BY THE BASE STATION
IS THE ACCURACY SUFFICIENT?
PARAMETERS FILE WORKING CORRECTLY

RESULTS TO_BE:

Simulation Results

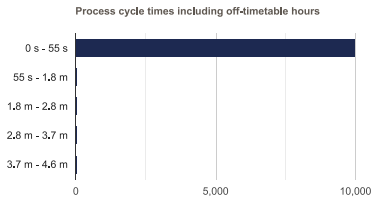
General information

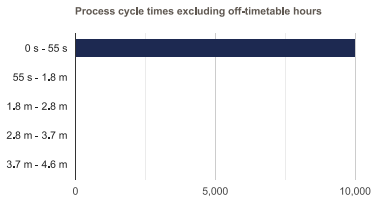
Completed process instances 10000

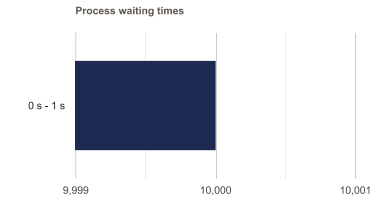
Total cost 0 EUR

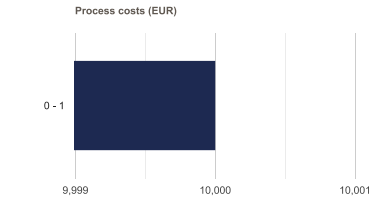
Total simulation time 49.6 weeks

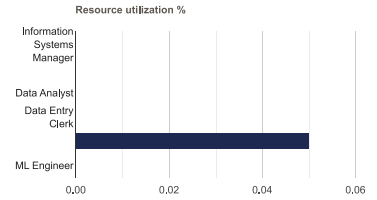
Charts











Scenario Statistics

	Minimum	Maximum	Average
Process instance cycle times including off-timetable hours	0 seconds	9 minutes	1.7 seconds
Process instance cycle times excluding off-timetable hours	0 seconds	9 minutes	1.7 seconds
Process instance costs	0 EUR	0 EUR	0 EUR

The interarrival time used was 720.

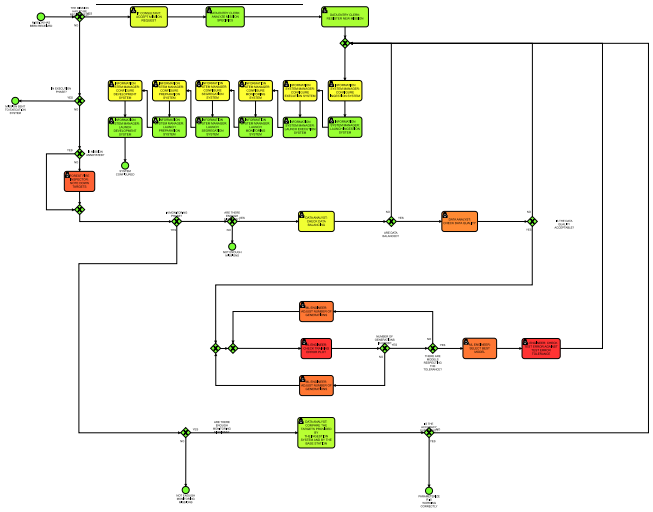
Activity Durations, Costs, Waiting times, Deviations from Thresholds

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: CHECK DATA BALANCING	3	0 s	0 s	0 s	12.3 s	12.3 s	12.3 s	0 s	0 s	0 s	0	0	0	0	0	0
DATA ANALYST: CHECK DATA QUALITY	3	0 s	0 s	0 s	25.4 s	25.4 s	25.4 s	0 s	0 s	0 s	0	0	0	0	0	0
DATA ANALYST: COMPARE THE TARGETS PROVIDED BY THE INGESTION SYSTEM AND BY THE BASE STATION	6	0 s	0 s	0 s	6.5 s	6.5 s	6.5 s	0 s	0 s	0 s	0	0	0	0	0	0
DATA ENTRY CLERK:
ANALYZE MISSION SPECIFICS	6	0 s	0 s	0 s	4 s	4 s	4 s	0 s	0 s	0 s	0	0	0	0	0	0
DATA ENTRY CLERK:
REGISTER NEW MISSION	6	0 s	0 s	0 s	5 s	5 s	5 s	0 s	0 s	0 s	0	0	0	0	0	0
FOREST FIRE INSPECTOR: NOTE DOWN TARGETS	519	0 s	0 s	0 s	29.5 s	29.5 s	29.5 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE DEVELOPMENT SYSTEM	9	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE EXECUTION SYSTEM	9	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE INGESTION SYSTEM	9	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE MONITORING SYSTEM	9	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE PREPARATION SYSTEM	9	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE SEGREGATION SYSTEM	9	0 s	0 s	0 s	13.6 s	13.6 s	13.6 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH DEVELOPMENT SYSTEM	9	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH EXECUTION SYSTEM	9	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH INGESTION SYSTEM	9	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH MONITORING SYSTEM	9	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH PREPARATION SYSTEM	9	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH SEGREGATION SYSTEM	9	0 s	0 s	0 s	3.3 s	3.3 s	3.3 s	0 s	0 s	0 s	0	0	0	0	0	0
IT CONSULTANT:
ACCEPT MISSION REQUEST	6	0 s	0 s	0 s	10.7 s	10.7 s	10.7 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: ADJUST NUMBER OF GENERATIONS	9	0 s	0 s	0 s	27.3 s	27.3 s	27.3 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: ADJUST NUMBER OF GENERATIONS	1	0 s	0 s	0 s	27.3 s	27.3 s	27.3 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: CHECK TEST ERROR AGAINST TEST ERROR TOLERANCE	3	0 s	0 s	0 s	34.1 s	34.1 s	34.1 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: CHECK TRAINING ERROR PLOT	13	0 s	0 s	0 s	34.1 s	34.1 s	34.1 s	0 s	0 s	0 s	0	0	0	0	0	0
ML ENGINEER: SELECT BEST MODEL	3	0 s	0 s	0 s	27.3 s	27.3 s	27.3 s	0 s	0 s	0 s	0	0	0	0	0	0

HEATMAP TO-BE:

Heatmap

Heatmap based on Durations

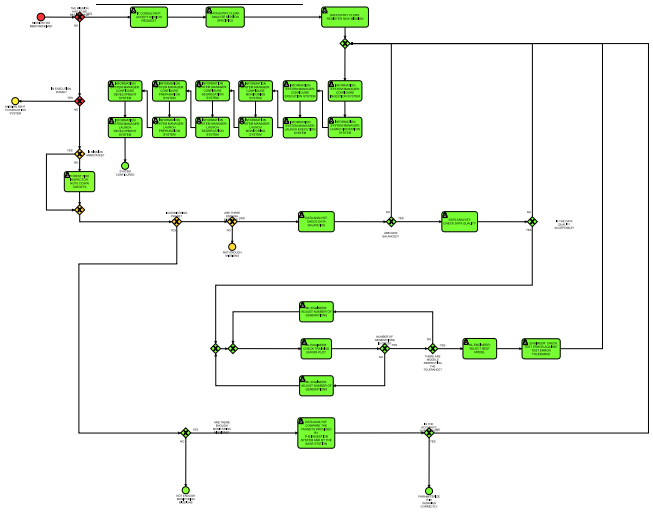


Legend

Color	Value
	0 s
	4 s
	8 s
	11 s
	15 s
	19 s
	23 s
	27 s
	30 s
	34 s

Heatmap

Heatmap based on Counts



Legend

Color	Value
	0
	1111
	2222
	3333
	4444
	5556
	6667
	7778
	8889
	10000

PROCESS MINING

-NORMATIVE PROCESS

The first thing we did was to simulate the AS-IS as a normative model, setting all the probabilities to 50%, costs to 1 eur, times to 1s and giving 10 instances to each lane, with 100 input tokens.

Simulation Results

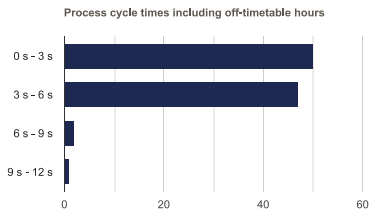
General information

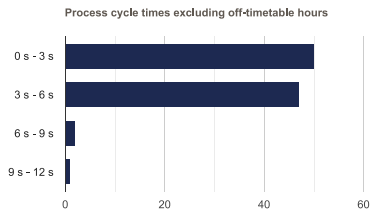
Completed process instances 100

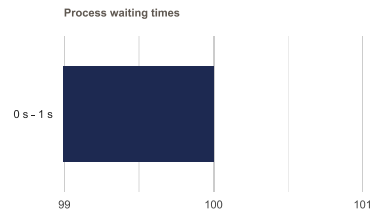
Total cost 794 EUR

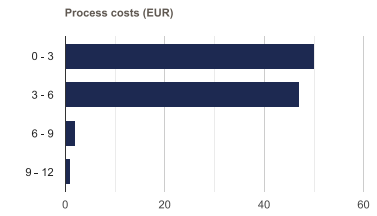
Total simulation time 4.2 days

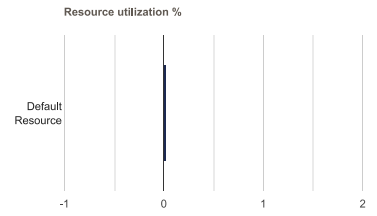
Charts











Scenario Statistics

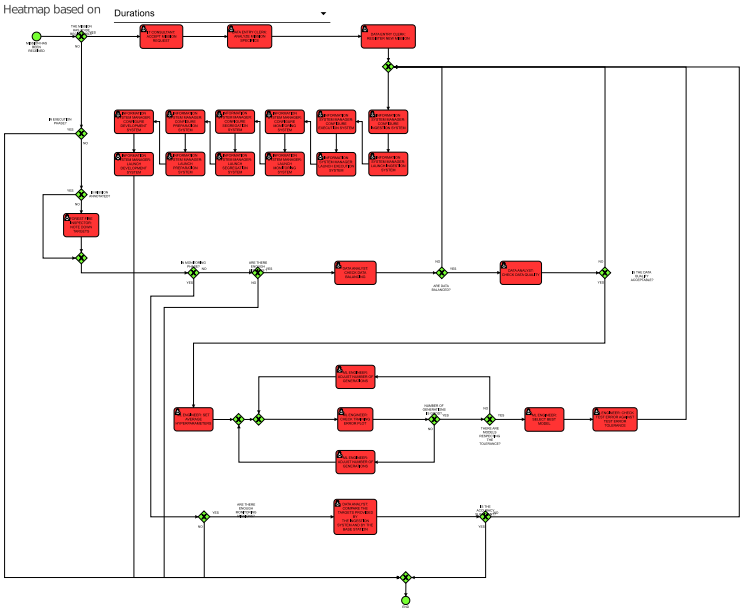
	Minimum	Maximum	Average
Process instance cycle times including off-timetable hours	0 seconds	27 seconds	7,9 seconds
Process instance cycle times excluding off-timetable hours	0 seconds	27 seconds	7,9 seconds
Process instance costs	0 EUR	27 EUR	7.9 EUR

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: CHECK DATA BALANCING	4	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
DATA ANALYST: CHECK DATA QUALITY	4	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
DATA ANALYST: COMPARE THE TARGETS PROVIDED BY THE INGESTION SYSTEM AND BY THE BASE STATION	5	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
DATA ENTRY CLERK:
ANALYZE MISSION SPECIFICS	46	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
DATA ENTRY CLERK:
REGISTER NEW MISSION	46	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
FOREST FIRE INSPECTOR: NOTE DOWN TARGETS	17	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE DEVELOPMENT SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE EXECUTION SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE INGESTION SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE MONITORING SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE PREPARATION SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
CONFIGURE SEGREGATION SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH DEVELOPMENT SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH EXECUTION SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH INGESTION SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH MONITORING SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH PREPARATION SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
INFORMATION SYSTEM MANAGER:
LAUNCH SEGREGATION SYSTEM	50	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
IT CONSULTANT:
ACCEPT MISSION REQUEST	46	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
ML ENGINEER: ADJUST NUMBER OF GENERATIONS	5	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
ML ENGINEER: ADJUST NUMBER OF GENERATIONS	2	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
ML ENGINEER: CHECK TEST ERROR AGAINST TEST ERROR TOLERANCE	3	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
ML ENGINEER: CHECK TRAINING ERROR PLOT	10	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
ML ENGINEER: SELECT BEST MODEL	3	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0
ML ENGINEER: SET AVERAGE HYPERPARAMETERS	3	0 s	0 s	0 s	1 s	1 s	1 s	0 s	0 s	0 s	1	1	1	0	0	0

Heatmap

Heatmap based on Durations

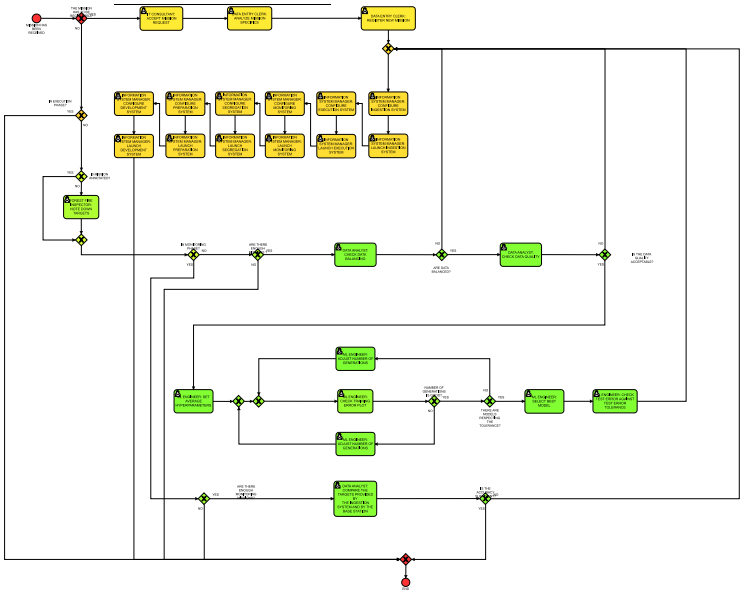


Legend

Color	Value
	0 s
	0 s
	0 s
	0 s
	0 s
	1 s
	1 s
	1 s
	1 s
	1 s

Heatmap

Heatmap based on Counts

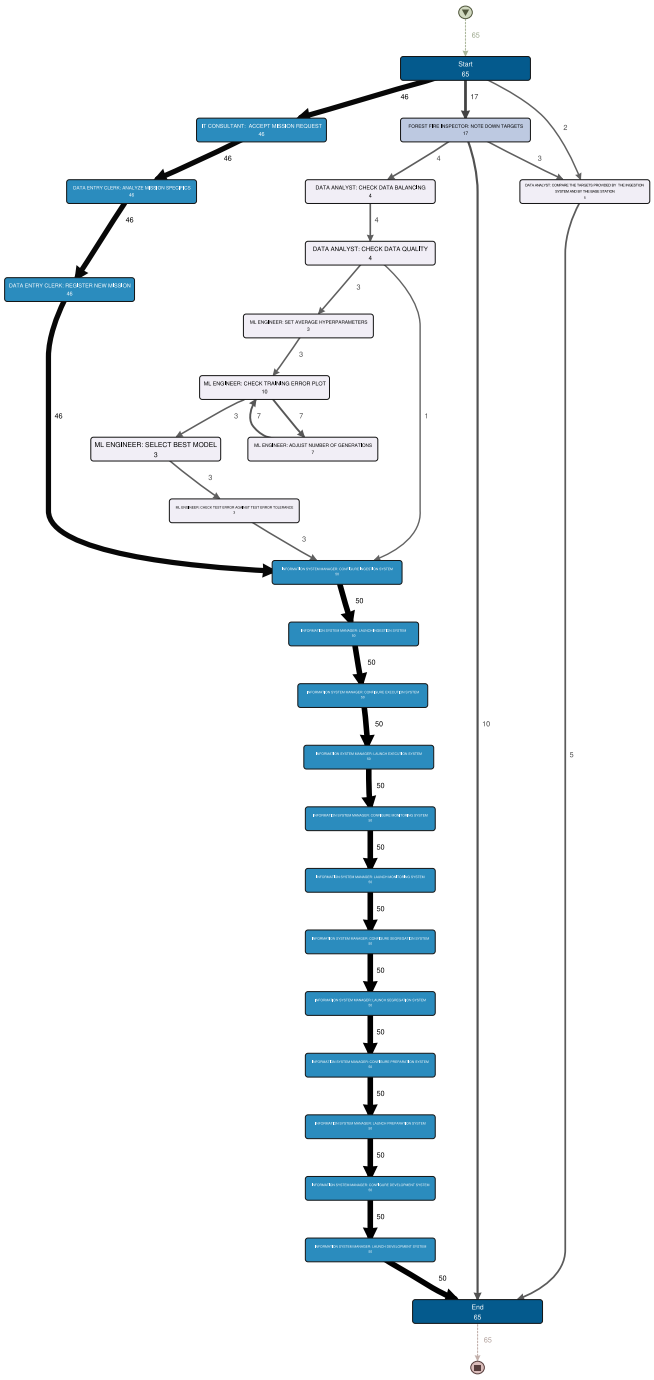


Legend

Color	Value
	0
	11
	22
	33
	44
	56
	67
	78
	89
	100

DISCO

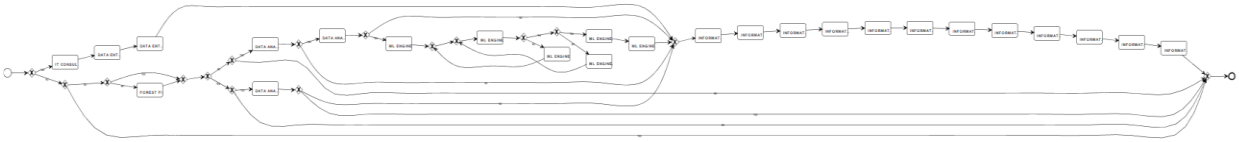
We noticed during mining that we needed a single end event in the normative model, so we updated it and simulated it. We imported the log (*.mxml*) obtained from the simulation of the normative model in Disco. We exported a *csv* file (without endpoints) and then we imported it again on Disco. We removed the unnecessary columns and set as timestamp the starting timestamp. Then we exported the *senior_log.xes* adding the endpoints, after removing the start event and the end event from the activity section.
We've also deleted the rows containing events with durations equal to zero.



PROM (Stefano, Fabiano)

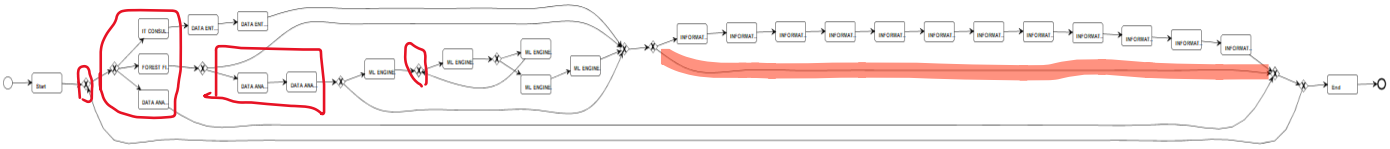
Normative model (Stefano, Fabiano)

We used the “Select BPMN Diagram” tool to display on ProM the normative model.



Inductive miner: (Stefano, Fabiano)

We’ve used the BPMN miner tool of ProM selecting the Inductive Miner and given in input the file senior_log.xes obtained from disco, we’ve obtained the following BPMN diagram as result. The setting used in the tool were the default ones.



We can see some differences from the original diagram: there is a new gateway at the beginning, in which there is a new flow incoming from a gateway before the end event; there is a split after the second gateway which is directed to the first three tasks, without other gateways before them; the two task of the data analyst after the annotation task are executed in sequence, and the rightmost highlighted gateway is a single gateway instead than two. Moreover, there is a new flow after the gateway before the reconfiguration, which is parallel to the reconfiguration itself

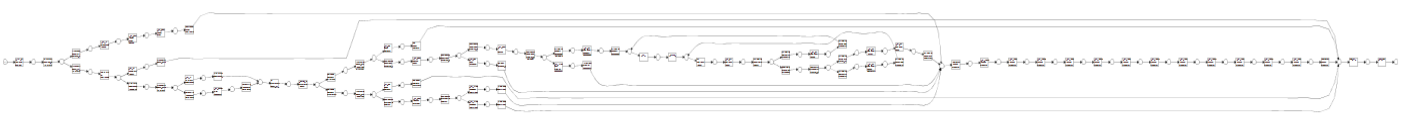
For the conformance checking we had to set to *NONE* the Start and End activities added by the miner, in order to match all the events in the log (since Disco added the endpoints for the Start and End event and so during the conformance checking, one of them was skipped).

Petri-net from mined bpmn:

The BPMN diagram resulting from the inductive mining was converted to the following Petri Net.

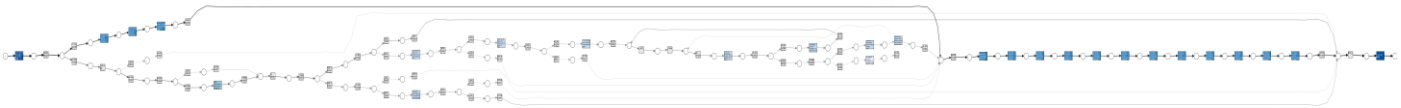
Petri-net from the bpmn of the normative model:

In order to obtain the petri-net from the normative model we’ve downloaded the bpmn, converted it to a bpmn diagram using the “SELECT BPMN DIAGRAM” obtaining an object BPMN Diagram. Then the diagram was converted to following Petri-Net:

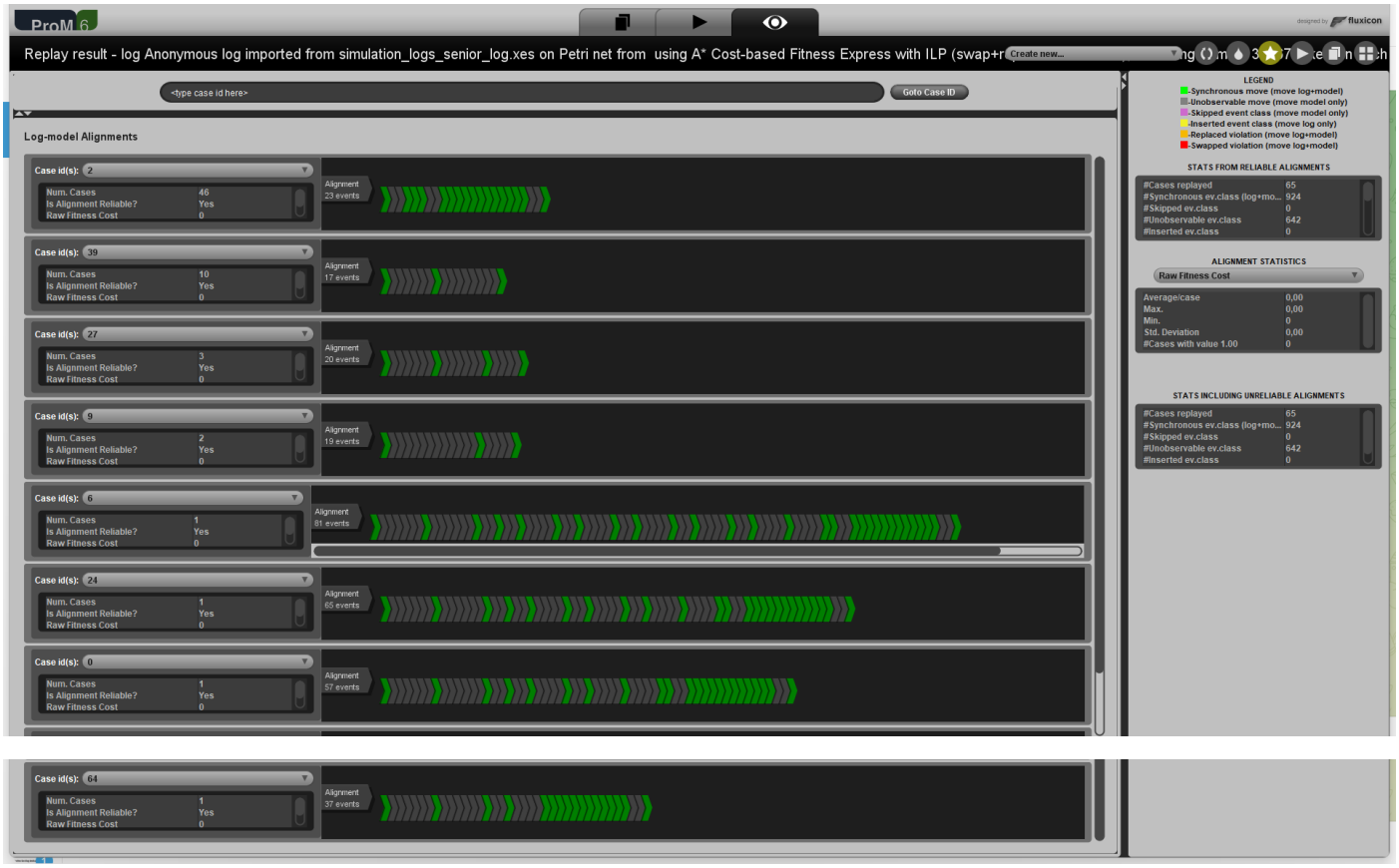


Replay the log on the Petri-net of the normative model: (Stefano, Fabiano)

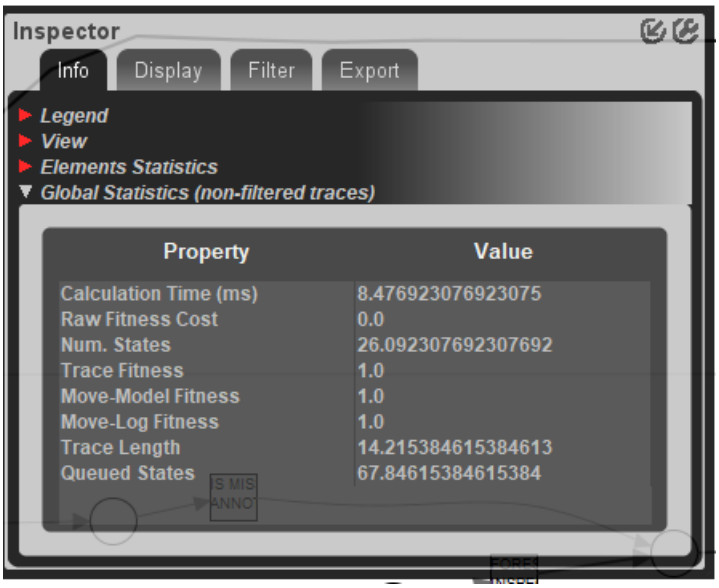
We've used the default setting, the classifier event name and set all the fields that doesn't start with "t_" as **NONE**, in order to skip them. In input we've **the Petri-net of the normative model** and the **senior log**. The result is in the following image:



Then we've created a new project alignment to log, obtaining the following result:



The information about the replay, containing the fitness can be seen in the next image:



Metrics for the normative model:

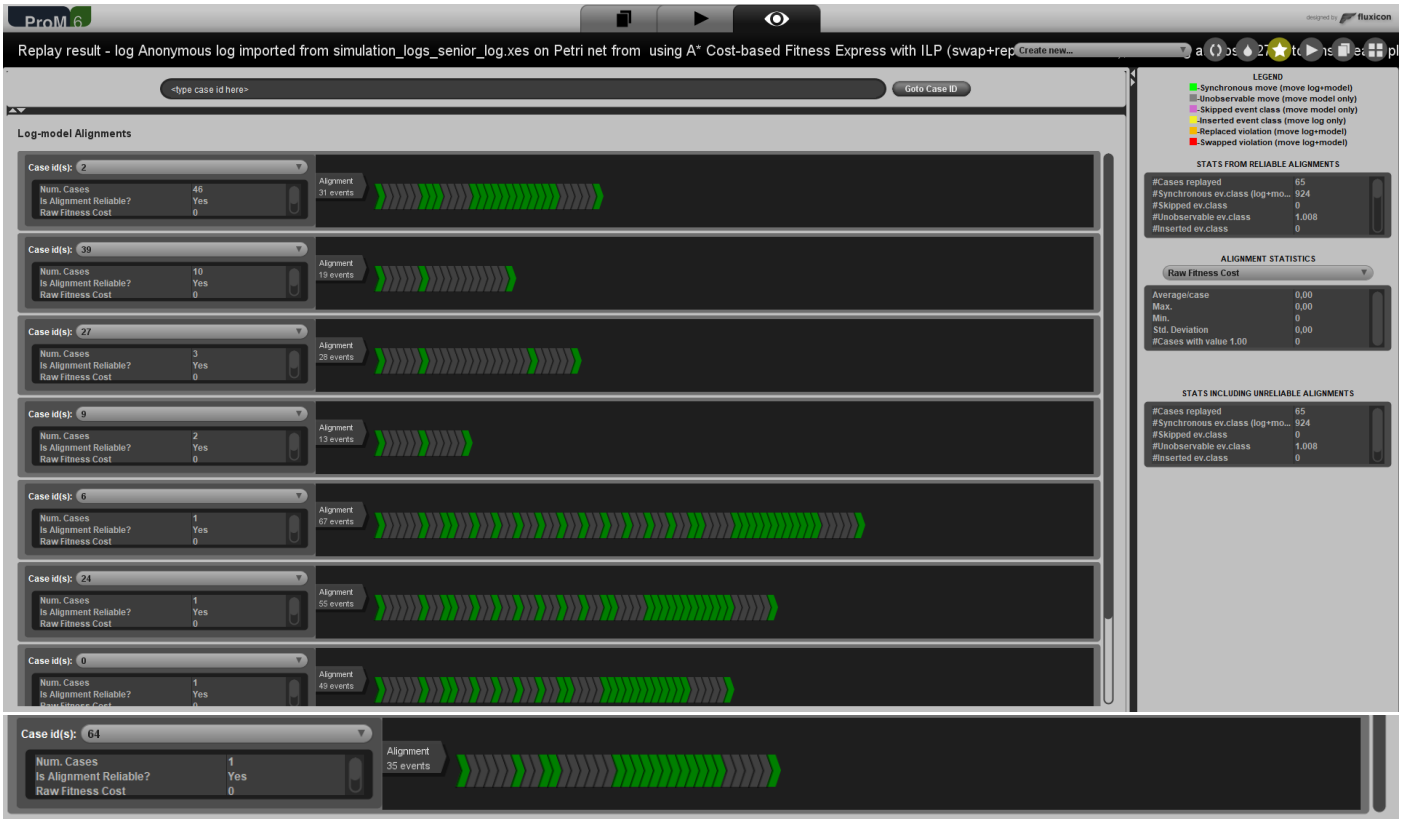
- **Fitness: 1 (obtained by conformance checking)**
- **Simplicity**: 16 gateways + 25 events + 53 flows = 94
- **Precision:** 0,83857
- **Generalization:** 0,86001

Precision and generalization were computed by using the *Measure Precision/Generalization* tool of ProM (valid for each metric computed)

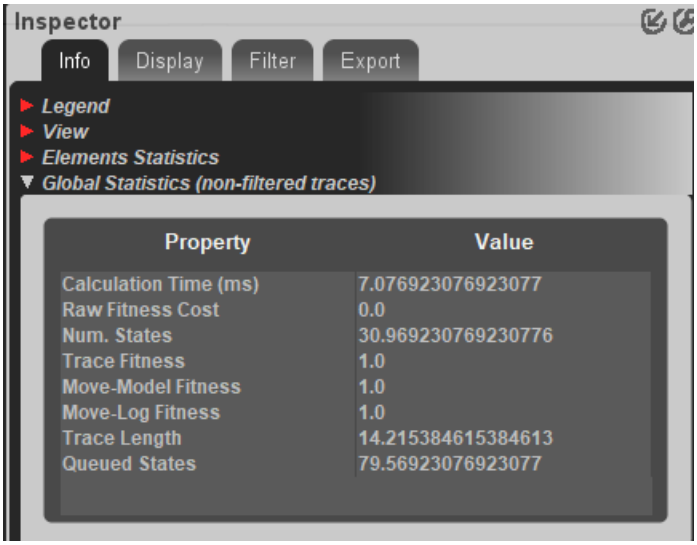
Replay the log on the Petri-net of the bpmn mined: (Stefano, Fabiano)

For this replay we've used in input the **petri-net of the bpmn obtained from the inductive mined** and the **senior_log**. Since the bpmn generated with the inductive miner adds two activities: one for the start and one for the end, so we had to put the **t_act_start** and **t_act_end** as **NONE** in the tool for replay. The result is presented in the following image:

Then we've created a new project alignment to log obtaining the following result:



The image of the inspector is the following:



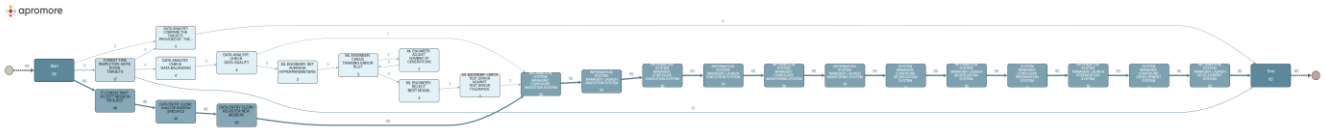
Metrics for the bpmn mined with inductive mining:

- **Fitness: 1.0**
- **Simplicity**: 10 gateways + 26 activities + 43 flows = 79;
- **Precision:** 0.81523
- **Generalization:** 0.86001

-APROMORE (Valerio, Elisa)

We've uploaded on Apromore's website the senior log and the bpmn normalized model obtained from disco.
We then proceeded to create both the transition map and the BPMN model.

Transition map



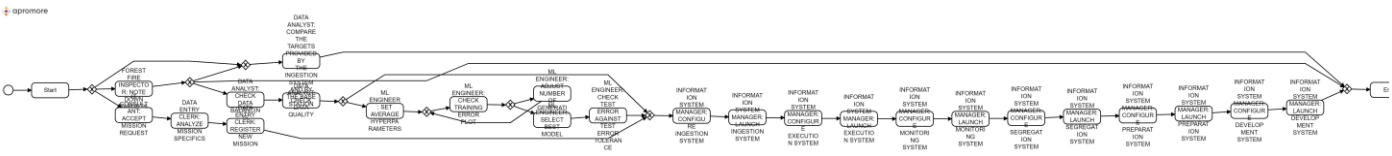
The Transition Map obtained with Apromore is exactly like the one obtained with Disco, with some differences: we can see that in some cases the number of tokens in the flows and the tasks is different; in fact we can see that in Disco the number of tokens in the loop after "check training error plot" is the total number of tokens that are needed in that loop, obtained by summing all the cases. In Apromore, instead, there are always the same number of tokens, without summing the tokens in the loop.

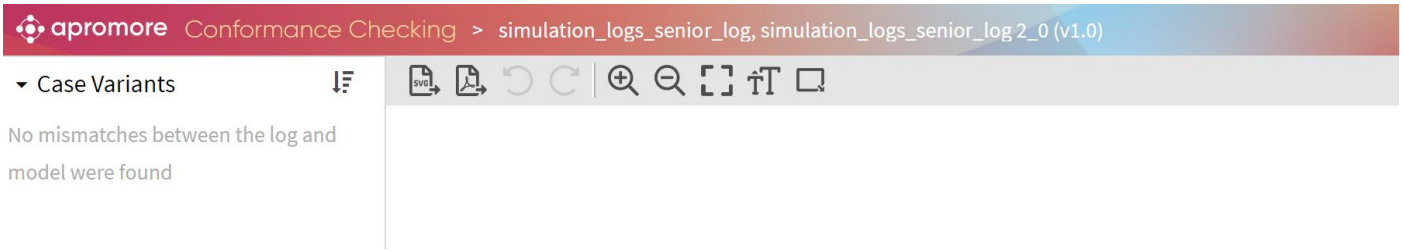
BPMN Model

Here we show the differences with the BPMN of the normative model:

The Forest Fire Inspector task is done directly after the first gateway, since Apromore removed two gateways before it; the gateways after this task have also been removed, along with arrows that pointed directly to the end event; Apromore also removed a gateway between the "Check data Balancing" and "Check Data quality tasks", which also included a flow directed to the configuration tasks; it also reduced to one the gateways between "Set average hyperparameters" and "Check Training error plot", and between "Check Training error plot" and "Select Best Model".

Conformance Checking Apromore

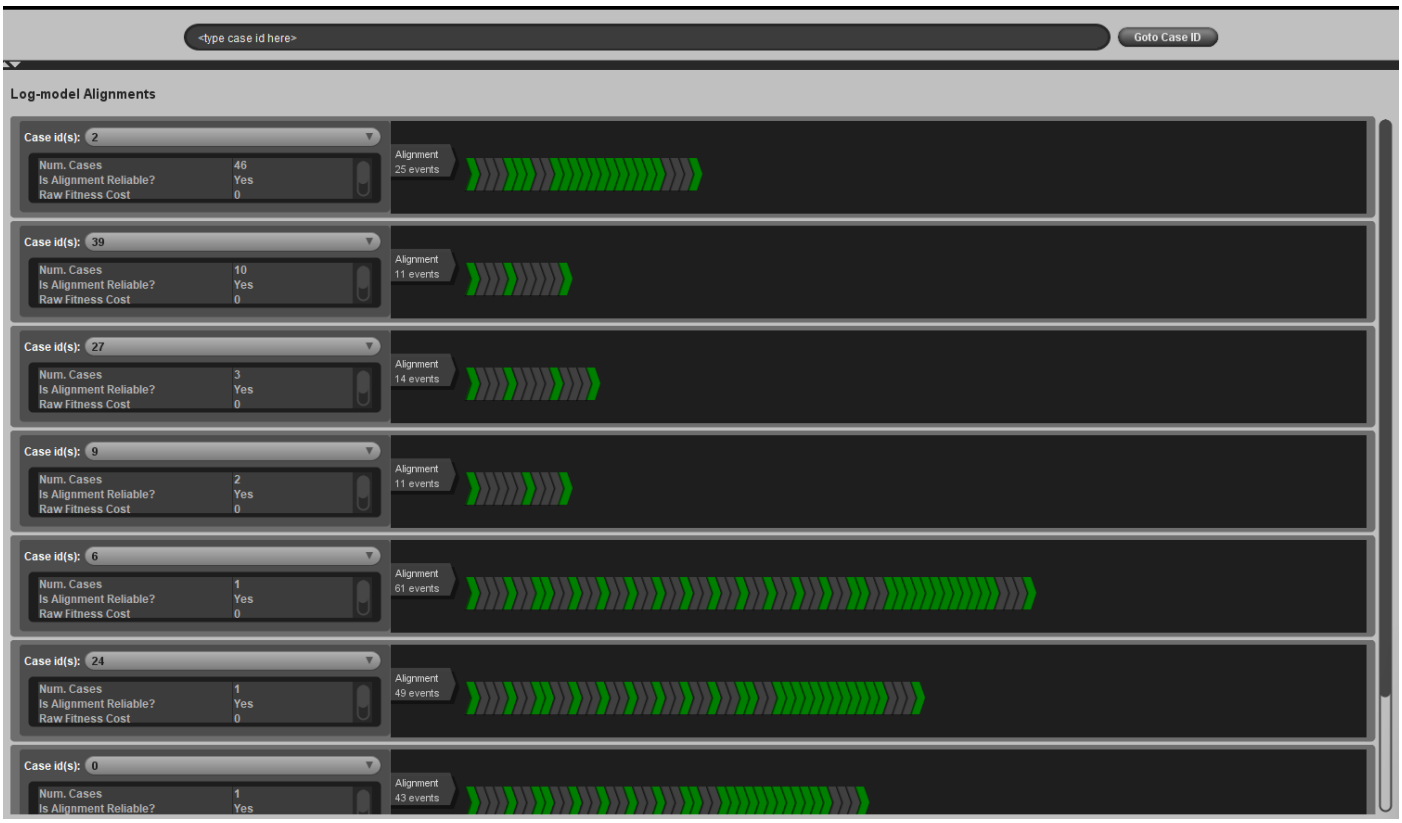


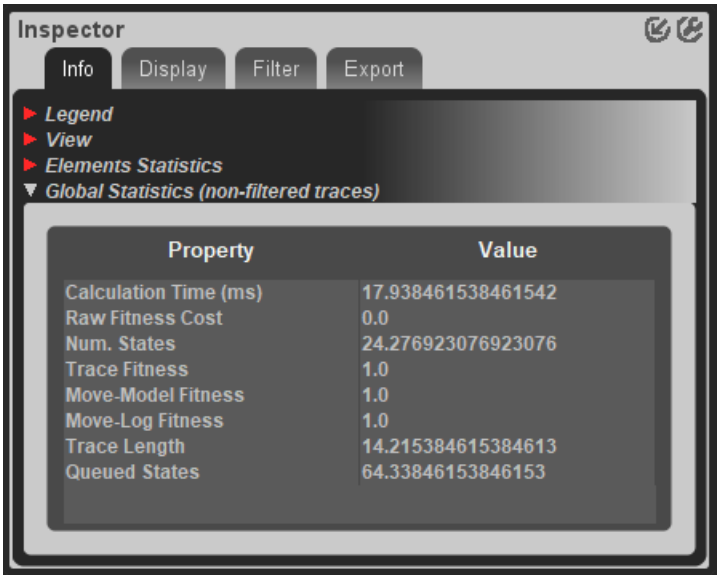


From the conformance checking we can see that there is no difference from the log and the model mined by Apromore.

From the APROMORE's results we then proceeded to calculate the metric's using ProM.

The steps were the same as before.





Metrics:

- **Fitness: 1**
- **Simplicity**: 26 events + 8 gateways + 41 flows = 75;
- **Precision:** 0,91763
- **Generalization:** 0,86001

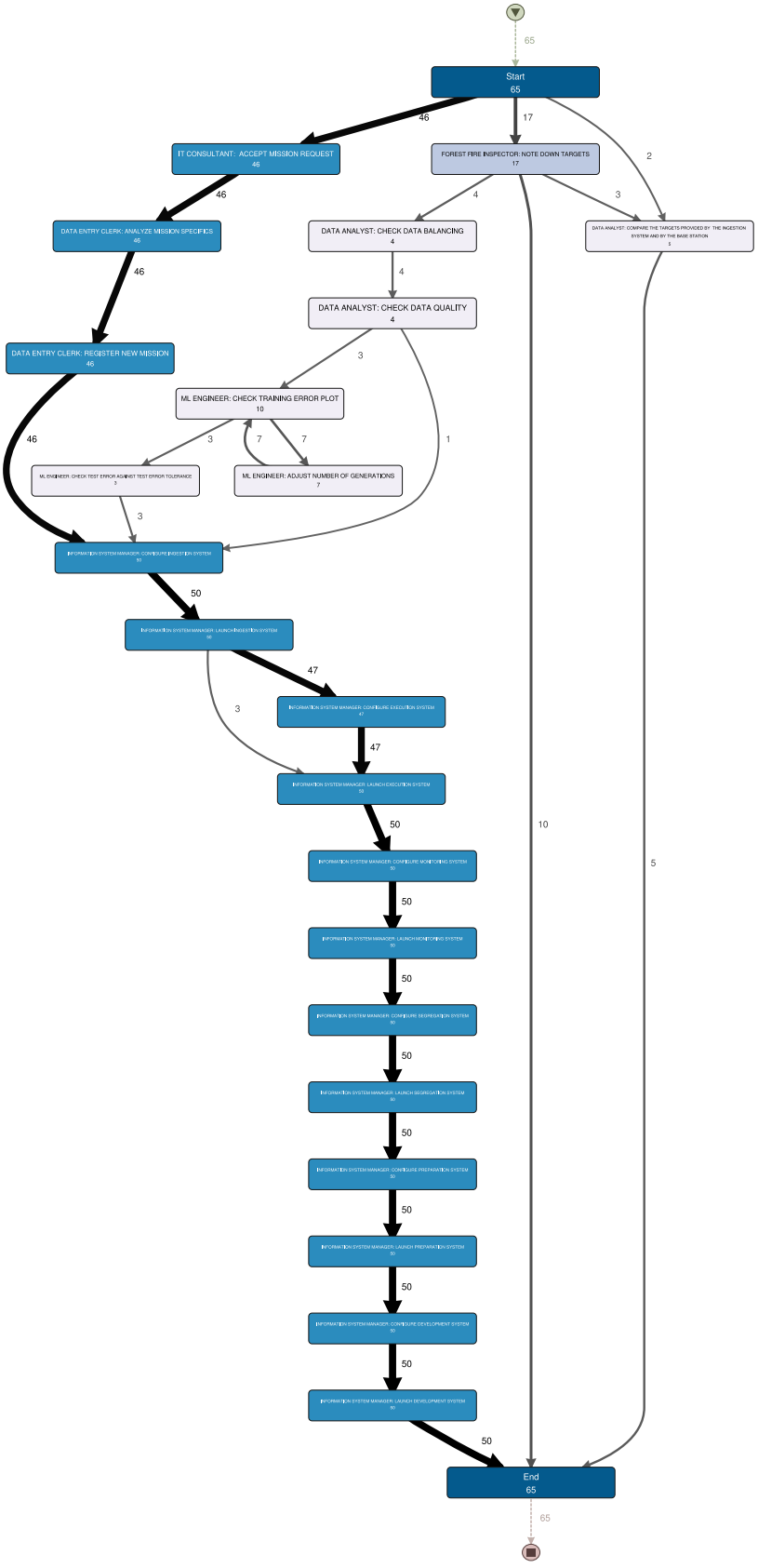
We can see that, with the model mined by Apromore, we have less extra behavior with respect to the log, compared to the inductive miner model, which has a lower precision. It is also more simple.

VIOLATIONS

After checking the normative model and verifying that it matches the event log, we added three violations to the log:

- Skip “CONFIGURE THE EXECUTION SYSTEM”: we assumed that this task can be skipped in some cases because the configuration of the execution system is not mandatory during the development phase, so we removed the rows 8, 25 and 45 of the *senior_log.csv* file obtained by Disco to apply this violation in 3 instances of the new case.
- Skip “SET AVERAGE HYPERPARAMETERS”: we assumed that this task can be skipped because we can take the average hyperparameters directly from similar previous clients, which missions are stored in the database, so we deleted the rows 112, 178 and 780 from the log.
- Skip “SELECT BEST MODEL”: we assumed that the model selected for deploying the parameterizer is the first one given by the system, which will have the highest training and validation accuracy, so we deleted the rows 122, 182 and 786 from the log.

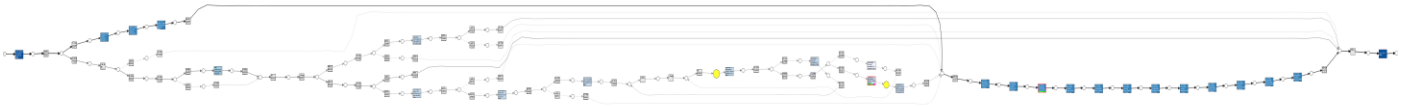
Transition Map with violations In Disco

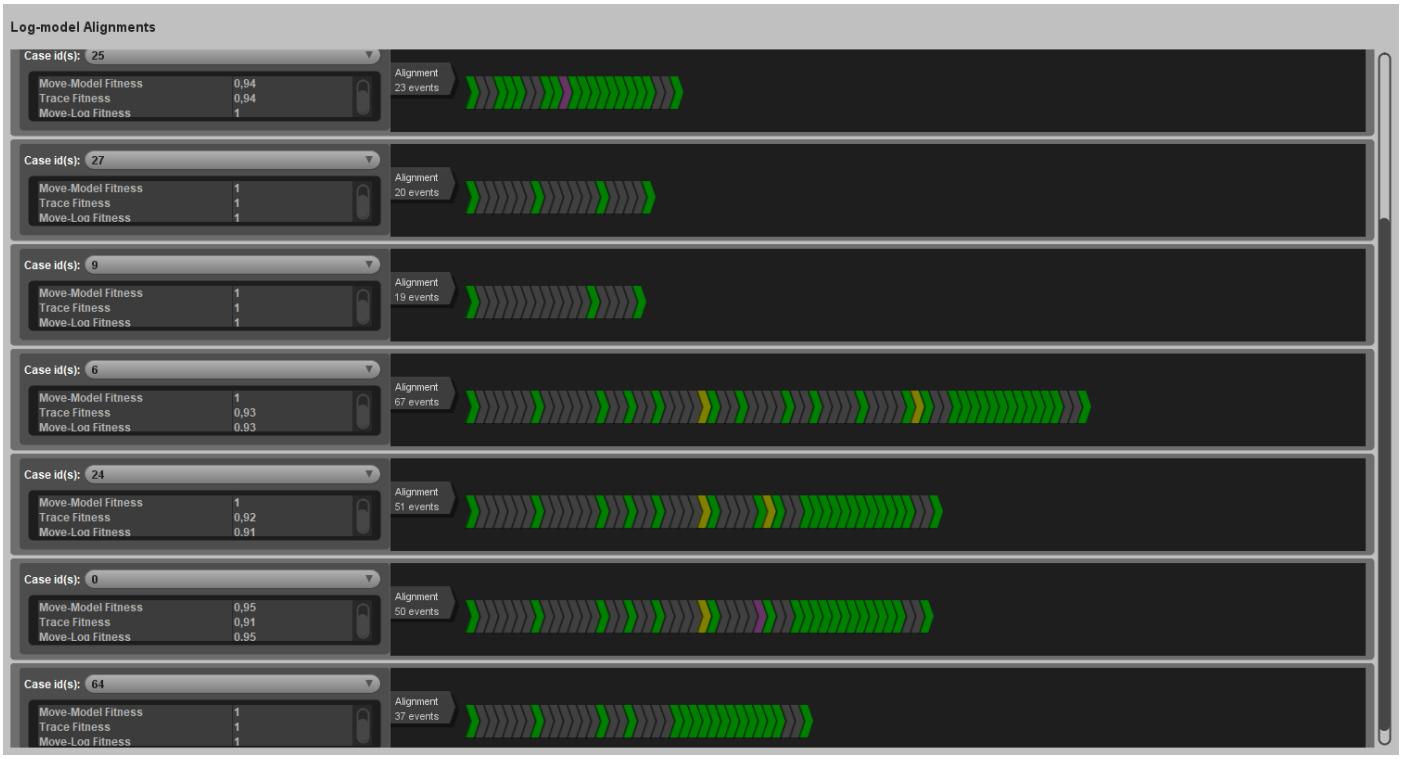


PROM WITH VIOLATIONS (Stefano, Fabiano)

Conformance Checking with the normative model: (Stefano)

For measuring the metrics for the normative model against the violated log, we did the same exact steps as before: selected the normative bpmn model, converted to a petri net and then replayed the log on the petri net for conformance checking. These are the results:





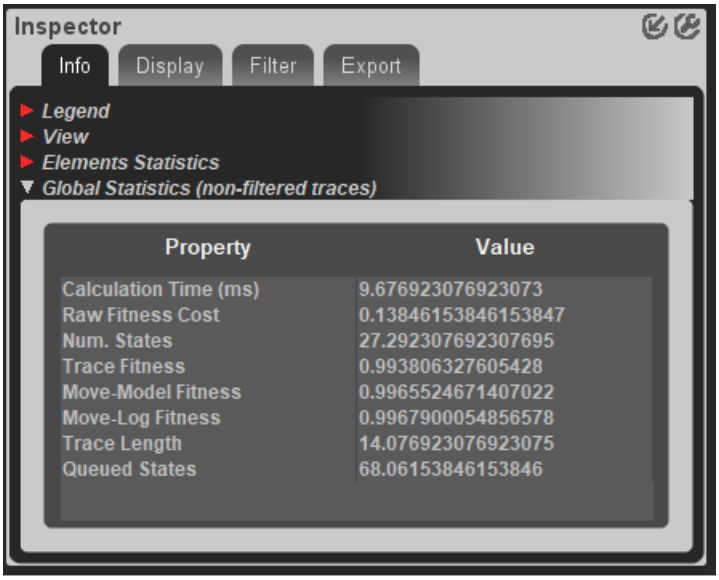
We can see that with the violated log, there are two events that are skipped (not in the log) and others that are not included on the model.

Skipped events: CONFIGURE EXECUTION SYSTEM (case 25) and SELECT BEST MODEL (case 0)

Inserted events: ADJUST NUMBER OF GENERATIONS (in all the three cases displayed) and CHECK TRAINING ERROR PLOT (in the first and second case with inserted events)

The skipped events are the one we skipped to violate the original log; the inserted events are generated by the third violation, in which we skipped the SET AVERAGE HYPERPARAMETERS task; instead, here the event is present because the model generalizes and skips the other tasks listed before; for the “open world assumption”, the activity on the normative model is replayed since it was not present in the log.

Global statistics:

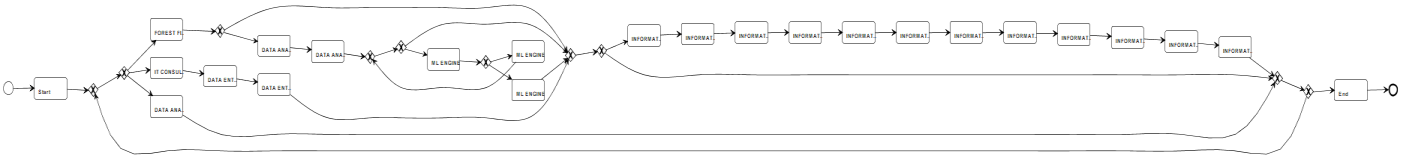


Metrics for the normative model:

- **Fitness:** 0.9938
- **Simplicity**: 16 gateways + 25 events + 53 flows = 94
- **Precision:** 0,85143
- **Generalization:** 0,86357

Mining a BPMN with Inductive Miner from the log with violations: (Fabiano)

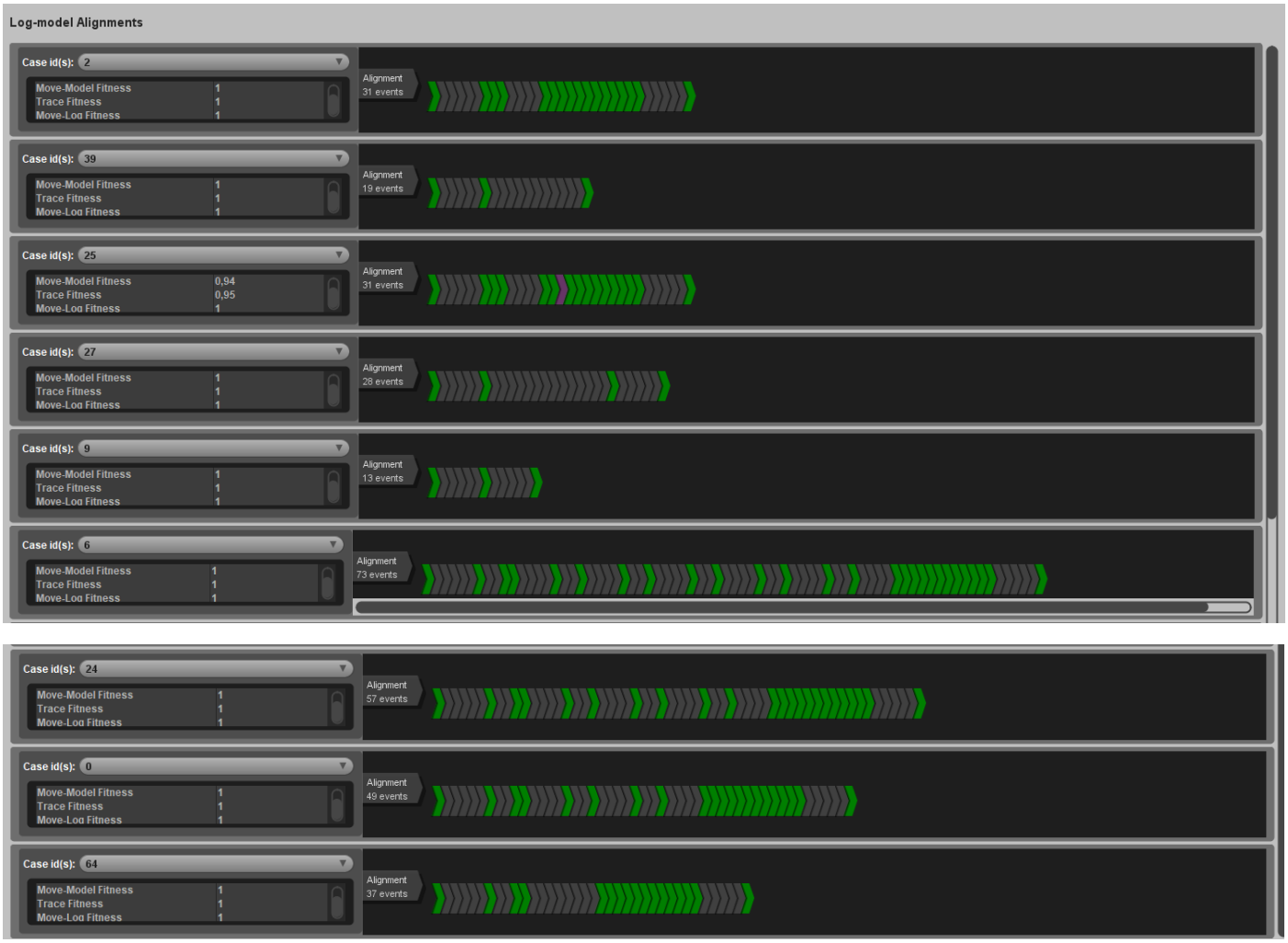
Using the BPMN miner tool we've mine a BPMN from the log with violations, the result is the following:



Petri-net from the mined BPMN from the log with violations:

Conformance check of the mined BPMN:

For the replay of the Petri-net against the low we've used the same approach using the log without violations, since the mining adds two endpoint (a start and an end) we've to skip this transition during the replay, since we've already a start and an end event (otherwise they're duplicated and appear always as skipped). The results are the following:



As we can see we've a skip in the case id 25 in corresponding of the activity "CONFIGURE EXECUTION SYSTEM" since in the model the activity is modeled, but in the log due to the violation applied three times, it's not present in three cases. This is the reason for the skipped event class transition.

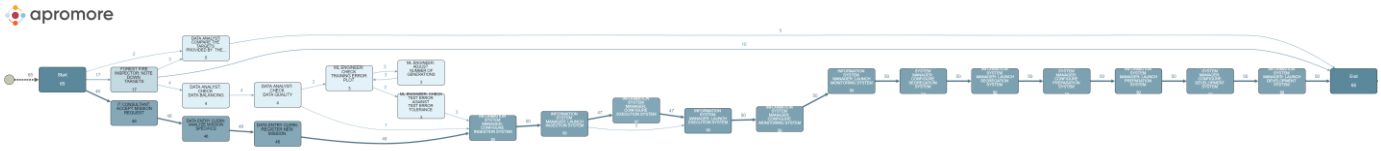
The global statistics are the following:

Metrics:

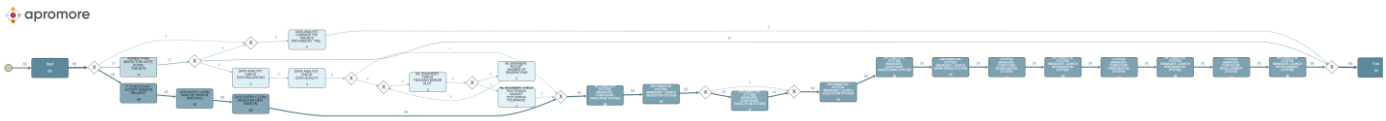
- **Fitness: 0.99757**
- **Simplicity**: 25 events + 10 gateways + 42 flows = 77;
- **Precision:** 0,76912
- **Generalization:** 0,88365

APROMORE WITH VIOLATIONS (Valerio, Elisa)

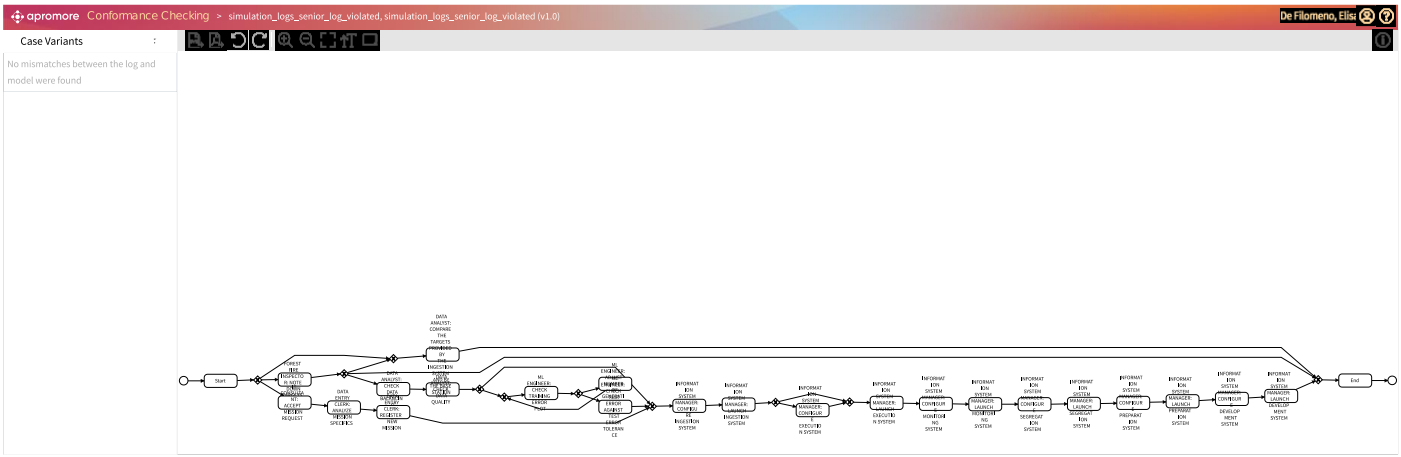
Mining a Transition Map from the log with violations in Apromore



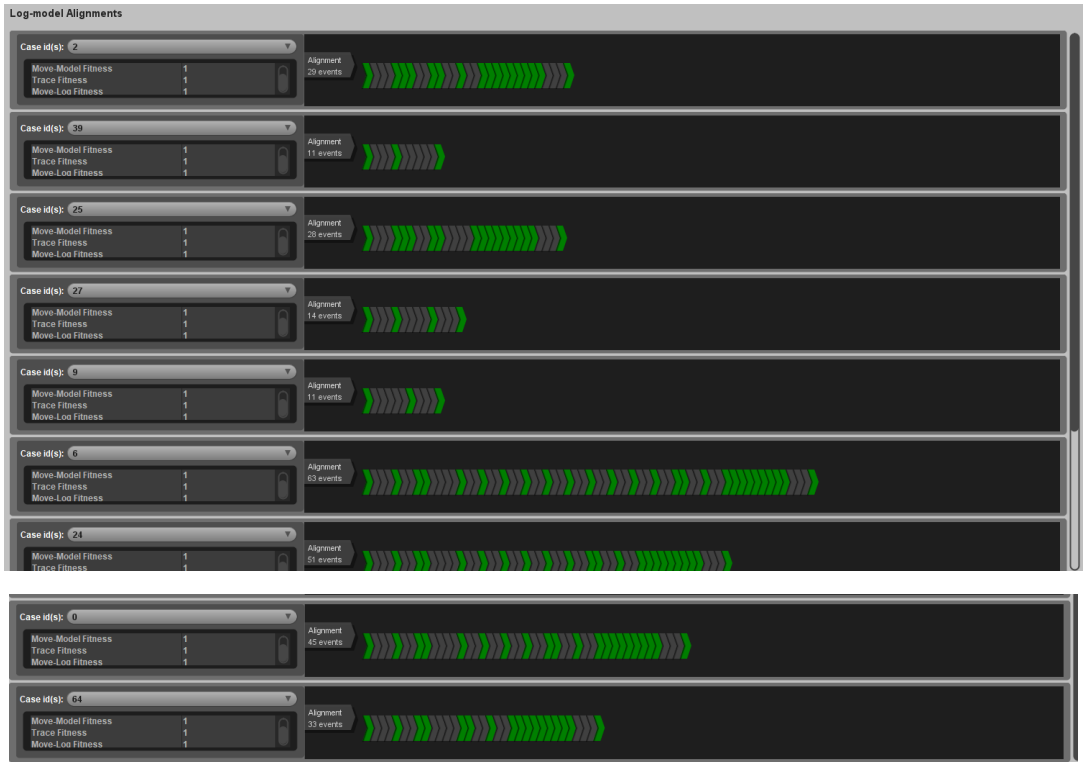
Mining a BPMN from the log with violations In Apromore



Conformance Checking Apromore



We used the bpmn generated by Apromore and replayed the log and computed the metrics as before:



We can see that with conformance checking of the violated model generated by Apromore on the violated log, no event is skipped from the log nor from the model.

Global statistics:

Metrics:

- **Fitness: 1**
- **Simplicity**: 24 events + 10 gateways + 43 flows = 77;
- **Precision:** 0,85394
- **Generalization:** 0,87469

Analyzing the metrics, we can say that, since the precision on both the model mined with Inductive Miner and with Apromore is lower than the ones obtained with the models mined with the non-violated log, that the modified models allow more extra behavior than the one observed in the violated log.