

# Process Mining and Intelligence Project

## Emotion Based Music Selection

Ettore Ricci

Francesco Boldrini

Paolo Palumbo

Zahra Omrani

January 29, 2025

## Contents

<b>1</b>	<b>BPMN modeling</b>	<b>2</b>
1.1	Process landscape . . . . .	2
1.2	Process model . . . . .	2
1.2.1	Prepare session . . . . .	2
1.2.2	Generate learning sets . . . . .	3
1.2.3	Develop classifier . . . . .	3
1.2.4	Classify session . . . . .	3
1.2.5	Evaluate classifier performance . . . . .	4
1.2.6	Configure systems . . . . .	4
<b>2</b>	<b>Data modeling</b>	<b>5</b>
2.1	Process model . . . . .	5
2.1.1	Prepare session . . . . .	5
2.1.2	Generate learning sets . . . . .	6
2.1.3	Develop classifier . . . . .	7
2.1.4	Classify session . . . . .	7
2.1.5	Evaluate classifier performance . . . . .	7
<b>3</b>	<b>Task level modeling</b>	<b>8</b>
3.1	Roles and salaries . . . . .	8
3.2	Segregation system . . . . .	10
3.2.1	Check data balancing . . . . .	10
3.2.2	Check input coverage . . . . .	11
3.2.3	Configure Segregation System . . . . .	12
3.3	Development system . . . . .	13
3.3.1	Set iteration number . . . . .	13
3.3.2	Check learning report . . . . .	14
3.3.3	Check validation report . . . . .	15
3.3.4	Check test results . . . . .	17
3.3.5	Configure Development System . . . . .	18
3.4	Evaluation system . . . . .	19
3.4.1	Evaluate classifier performance . . . . .	19
3.4.2	Configure Evaluation System . . . . .	20
3.5	Client-Side Systems . . . . .	21
3.5.1	Configure Client-Side Systems . . . . .	21
3.6	Production System . . . . .	22
3.6.1	Configure Production Systems . . . . .	22
3.7	Ingestion System . . . . .	22
3.7.1	Configure Ingestion System . . . . .	22
3.8	Preparation System . . . . .	23

3.8.1	Configure Preparation System	23
4	Simulation	24
4.1	Collapsed workflow	24
5	Process mining	25
5.1	Violations	27

## 1 BPMN modeling

### 1.1 Process landscape

[Ettore Ricci, Paolo Palumbo, Francesco Boldrini, Zahra Omrani]

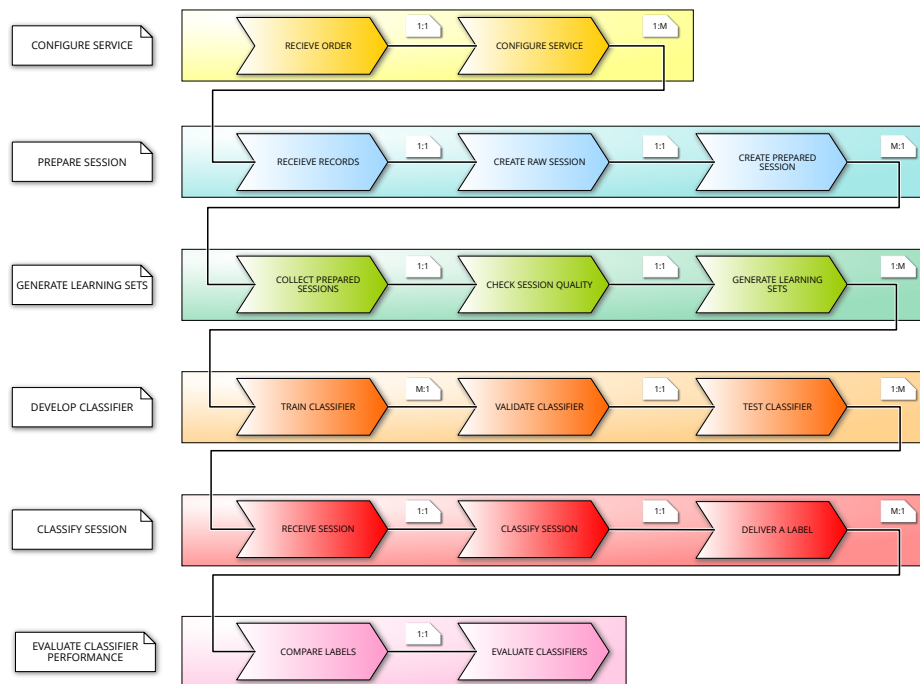
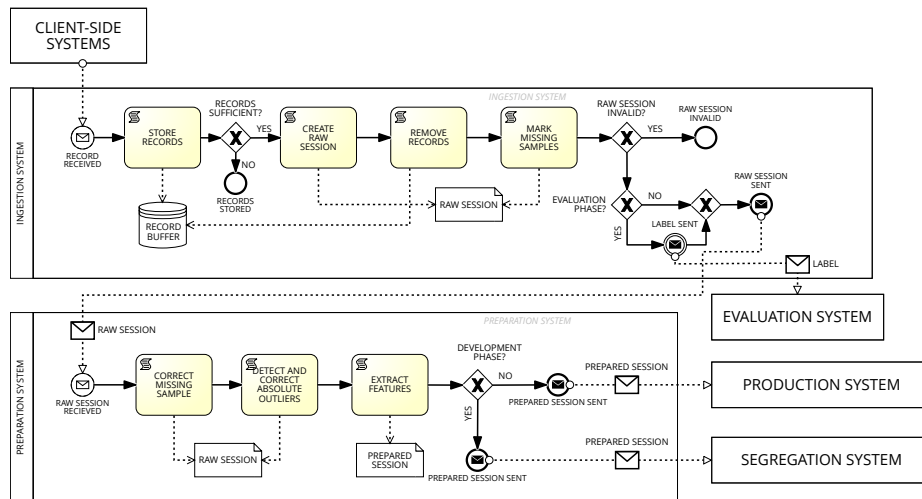


Figure 1: Process landscape

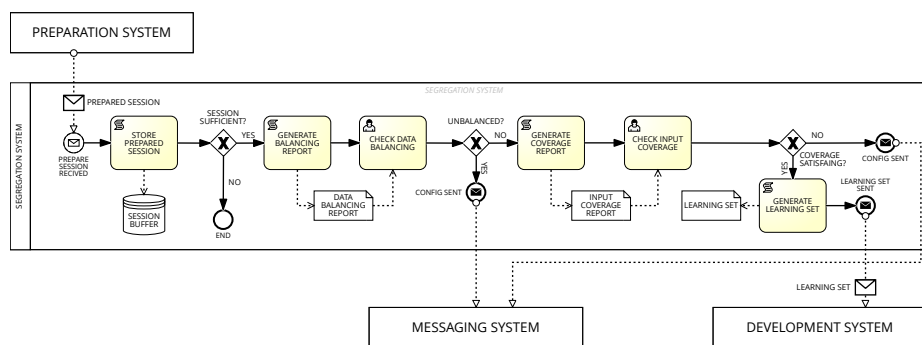
### 1.2 Process model

#### 1.2.1 Prepare session

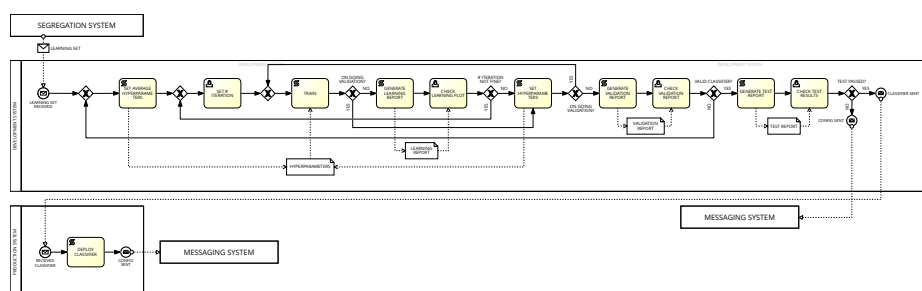
[Ettore Ricci, Paolo Palumbo]



### 1.2.2 Generate learning sets



### 1.2.3 Develop classifier



### 1.2.4 Classify session

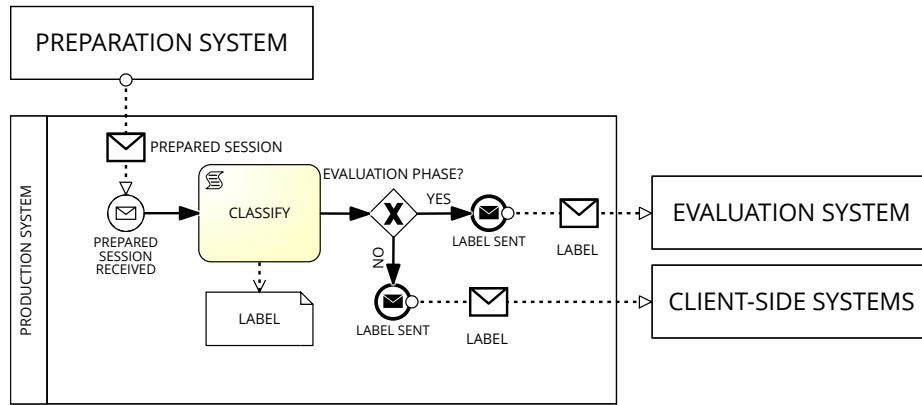


Figure 5: Business Diagram of the "Classify session" process

### 1.2.5 Evaluate classifier performance

[Ettore Ricci, Paolo Palumbo]

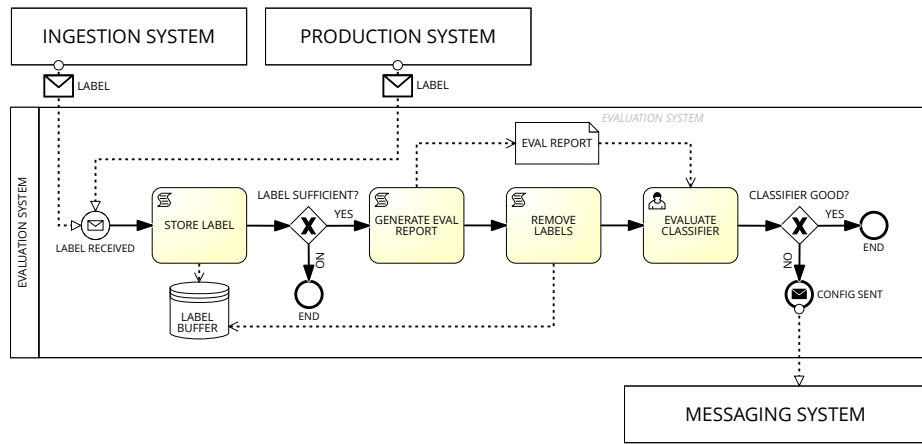


Figure 6: Business Diagram of the "Evaluate classifier performance" process

### 1.2.6 Configure systems

[Ettore Ricci, Paolo Palumbo]

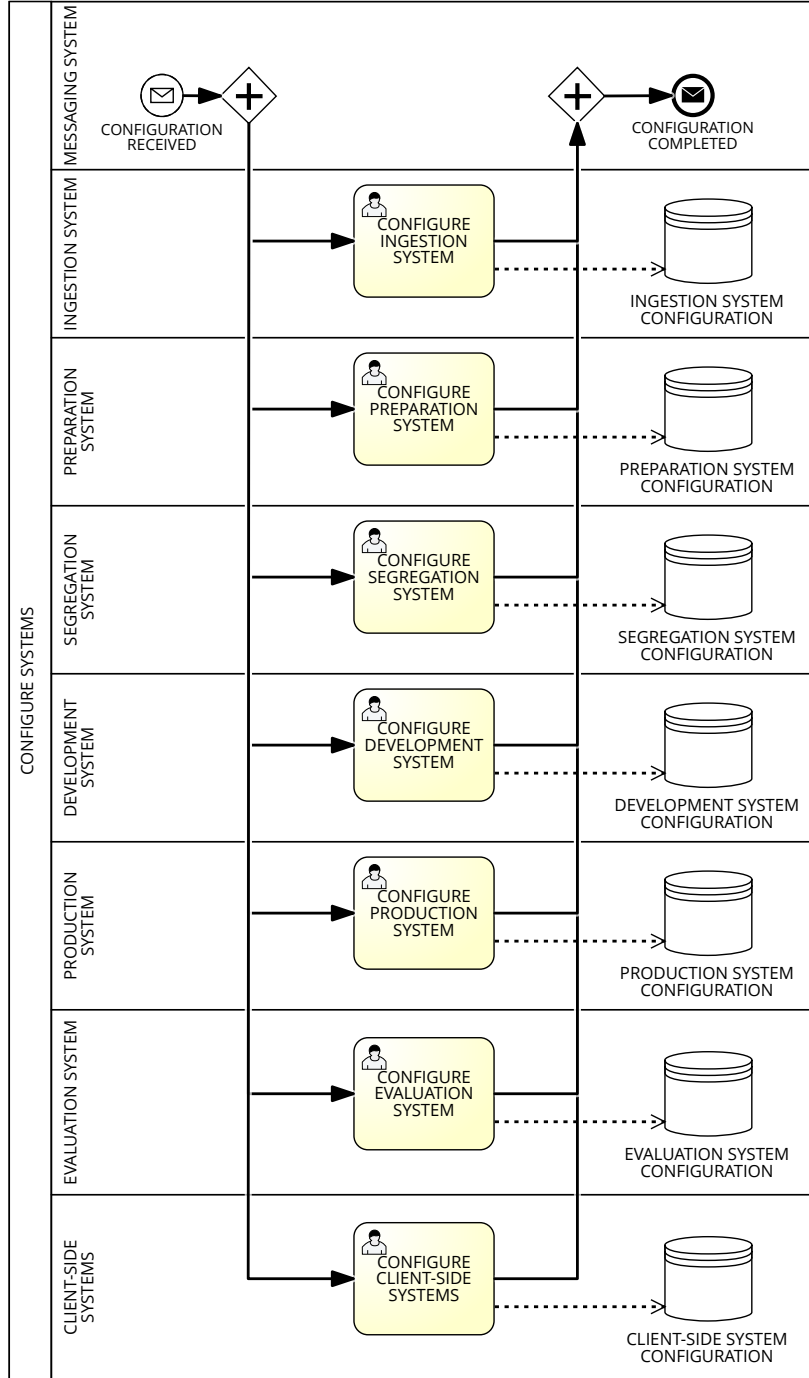


Figure 7: Business Diagram of the "Configure systems" process

## 2 Data modeling

### 2.1 Process model

#### 2.1.1 Prepare session

[Ettore Ricci]

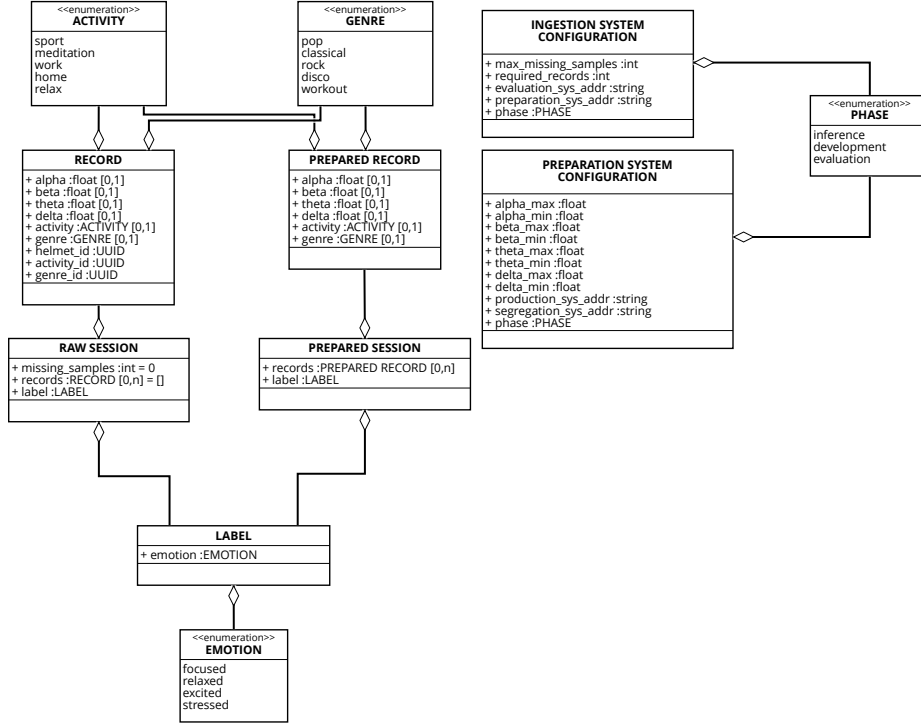


Figure 8: Data Model of the "Prepare session" process

## 2.1.2 Generate learning sets

[Paolo Palumbo]

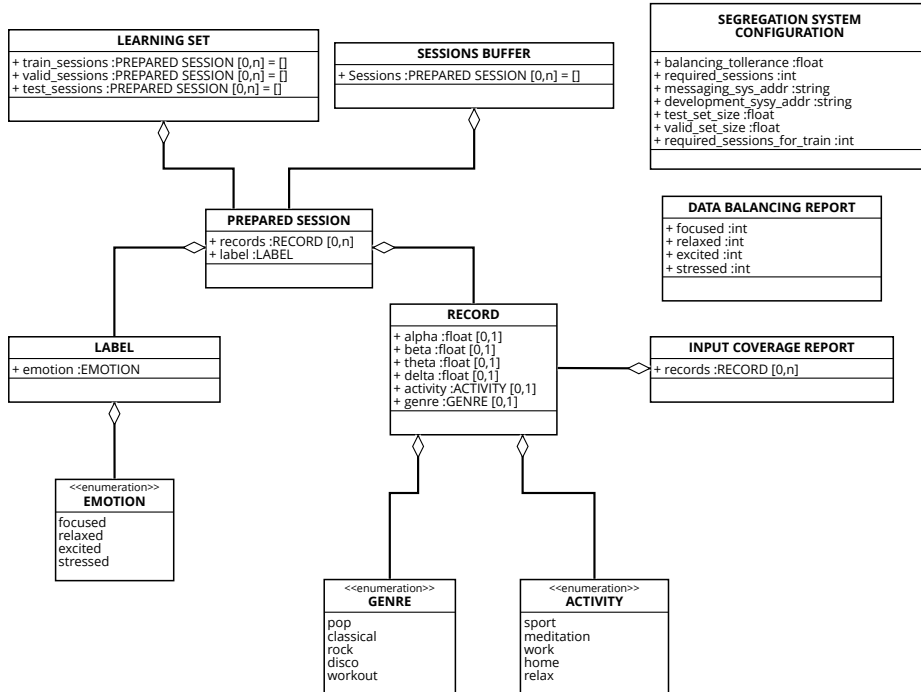


Figure 9: Data Model of the "Generate learning sets" process

### 2.1.3 Develop classifier

[Paolo Palumbo]

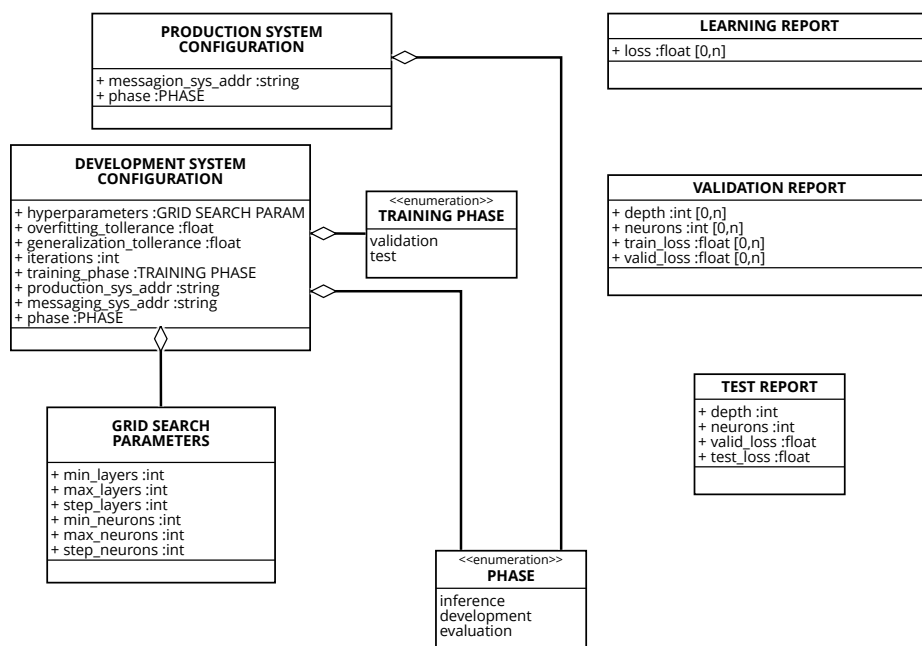


Figure 10: Data Model of the "Develop classifier" process

### 2.1.4 Classify session

[Francesco Boldrini]

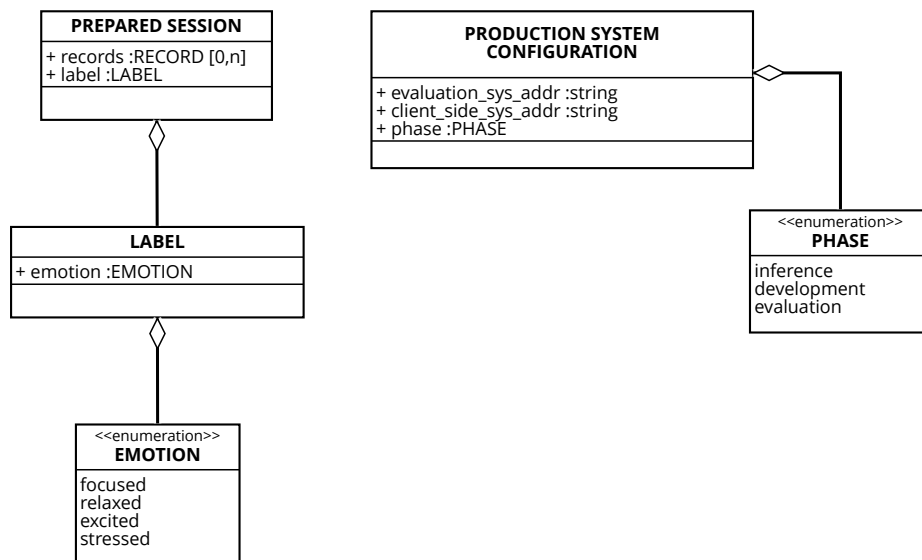


Figure 11: Data Model of the "Classify session" process

### 2.1.5 Evaluate classifier performance

[Zahra Omrani]

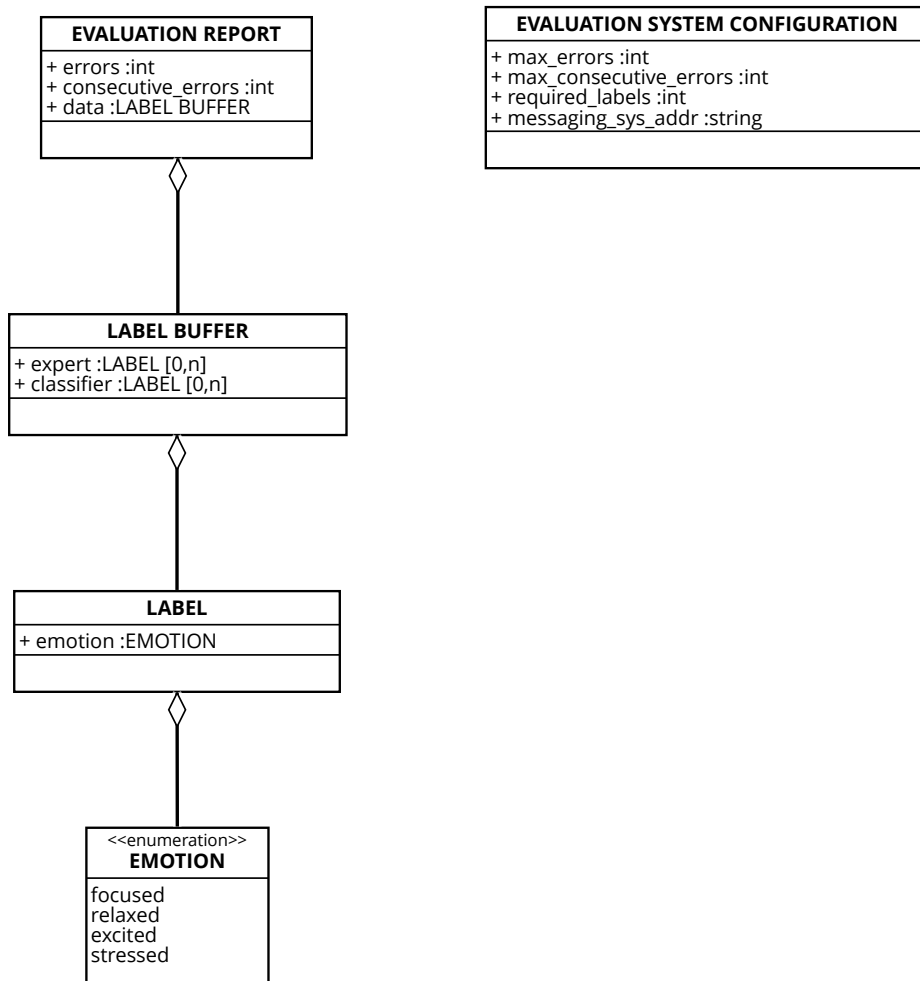


Figure 12: Data Model of the "Evaluate classifier performance" process

### 3 Task level modeling

#### 3.1 Roles and salaries

[Ettore Ricci, Paolo Palumbo]



Position	Description	Salary	Normalized Salary
Clerk	Handles administrative tasks, organizes documentation, and assists with data entry and labeling. Ensures smooth operations by coordinating communication and managing resources.	\$52,000.00	1.00
Data analyst	Prepares, analyzes, and visualizes data to extract insights. Collaborates on cleaning datasets, identifying trends, and supporting model validation.	\$60,000.00	1.15
ML engineer	Builds, tests, and deploys machine learning models, optimizing performance and scalability. Integrates AI solutions into production systems with a focus on efficiency.	\$130,000.00	2.50
Data scientist	Designs and experiments with AI models, applying advanced techniques to solve project challenges. Collaborates with experts to integrate domain knowledge and refine outputs.	\$123,000.00	2.37
Domain expert (Neurologist)	Provides medical expertise to guide AI development and validate results. Ensures solutions align with clinical standards and address neurological challenges.	\$267,000.00	5.13
<b>Minimum</b>		\$52,000.00	1.00

Table 1: Salary and normalized salary for each position

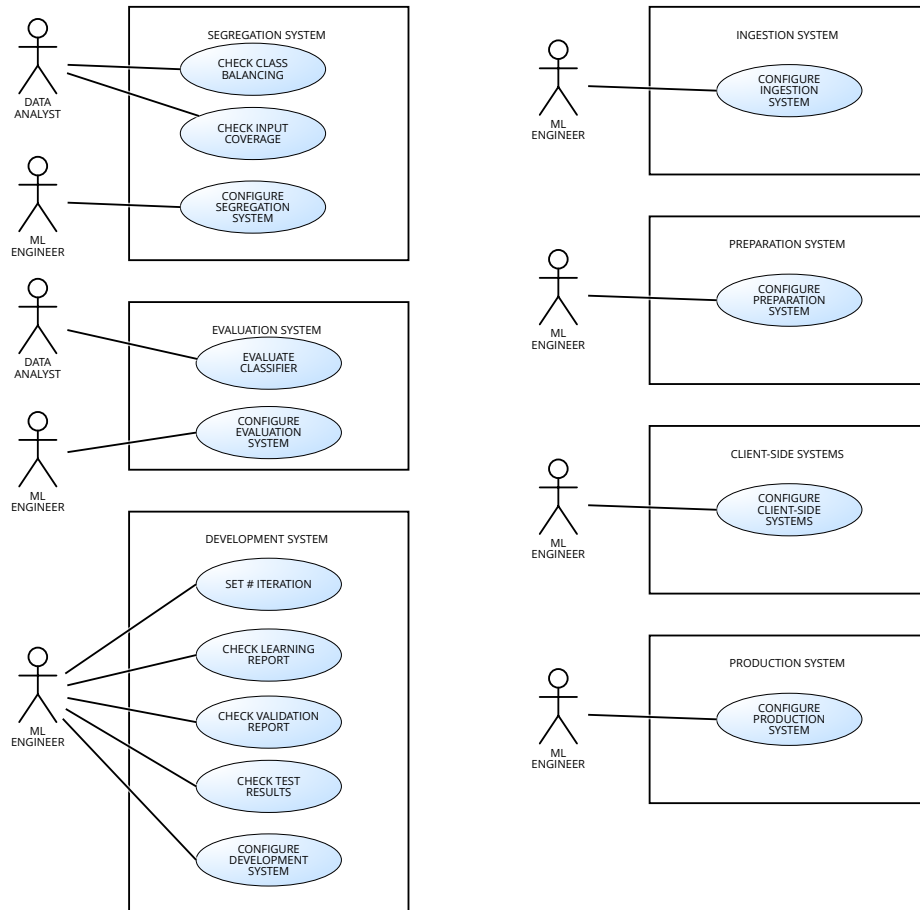


Figure 13: Use case diagram

## 3.2 Segregation system

### 3.2.1 Check data balancing

[Ettore Ricci, Paolo Palumbo]

The task is performed by a Data Analyst.



Figure 14: "Check data balancing" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens "Check data balancing" form.	1	1	1.15	1.15
2 <b>SYSTEM</b> shows the report.				
3 <b>SYSTEM</b> shows a hint whether the data is balanced or not.				
4 <b>ACTOR</b> checks the hint to see if the data is balanced or not.	1	2	1.15	2.30
5.1 <b>IF</b> the data is balanced.	0.2			
5.1.1 <b>ACTOR</b> clicks "Balanced" button.	0.2	1	1.15	0.23
5.2 <b>ELSE</b>	0.8			
5.2.1 <b>ACTOR</b> clicks "Unbalanced" button.	0.8	1	1.15	0.92
7 <b>SYSTEM</b> shows a confirmation dialog.				
8 <b>ACTOR</b> closes the form.	1	1	1.15	1.15
Human task cost				5.74

Table 2: Detailed use case for "Check data balancing" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.2.2 Check input coverage

[Ettore Ricci, Paolo Palumbo]

The task is performed by a Data Analyst.



Figure 15: "Check input coverage" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens "Check input coverage" form.	1	1	1.15	1.15
2 <b>SYSTEM</b> shows a radar scatter plot of the input distribution.				
3 <b>FOR EACH</b> radius in the radar scatter plot:	6			
3.1 <b>ACTOR</b> checks if the distribution is uniform on the radius.	6	4	1.15	27.6
3.1.1 <b>IF</b> the distribution is not uniform as expected.	4			
3.1.1.1 <b>THEN</b> the input coverage is not satisfied.	4			
4.1 <b>IF</b> the input coverage is satisfied.	0.33			
4.1.1 <b>ACTOR</b> clicks "Accept" button.	0.33	1	1.15	0.38
4.2 <b>ELSE</b>	0.66			
4.2.1 <b>ACTOR</b> clicks "Reject" button.	0.66	1	1.15	0.76
5 <b>SYSTEM</b> shows a confirmation dialog.				
6 <b>ACTOR</b> closes the form.	1	1	1.15	1.15
Human task cost				31.04

Table 3: Detailed use case for "Check input coverage" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.2.3 Configure Segregation System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.

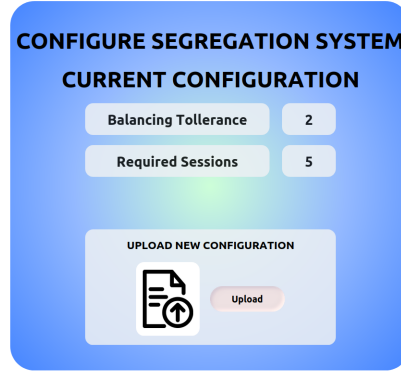


Figure 16: "Configure Segregation System" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens the "Configure Segregation System" form.	1	1	2.50	2.50
2 <b>SYSTEM</b> displays current configuration and "Upload" button.				
3 <b>ACTOR</b> checks parameters against previous iterations on file	1	3	2.50	7.50
4 <b>ACTOR</b> adjusts file based on current parameters	1	3	2.50	7.50
5 <b>ACTOR</b> pushes "Upload" button and uploads configuration file	1	1	2.50	2.50
6.1 <b>SYSTEM</b> IF config is correct and correctly formatted				
6.1.1 <b>SYSTEM</b> shows a confirmation message.				
6.2 <b>ELSE</b>				
6.2.1 <b>SYSTEM</b> shows error message and aborts				
7 <b>ACTOR</b> closes the form.	1	1	2.50	2.50
Human task cost				22.50

Table 4: Detailed use case for "Configure Segregation" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.3 Development system

#### 3.3.1 Set iteration number

[Zahra Omrani]

The task is performed by a ML engineer.

The mock-up form is titled "Set Iteration Number" and includes a close button (X) in the top right corner. It displays the "Current Iteration Number" as 10. Below this, there is a label "Enter New Iteration Number:" followed by an empty input field. A "Submit" button is positioned below the input field. At the bottom of the form, a message states "Iteration number updated successfully!".

Figure 17: "Set iteration number" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens "Set Iteration Number" form.	1	1	2.5	2.5
2 <b>SYSTEM</b> displays the current iteration number.				
3.1 <b>IF</b> it's the first configuration:				
3.1.1 <b>ACTOR</b> inputs the desired number of iterations based on task complexity and previous experience.	0.2	3	2.5	1.5
3.2 <b>ELSE</b> (subsequent configurations):				
3.2.1 <b>ACTOR</b> inputs the number based on the established learning curve.	0.8	1	2.5	2
4 <b>ACTOR</b> clicks "Submit" button to confirm the iteration number.	1	1	2.5	2.5
5 <b>SYSTEM</b> shows a confirmation dialog.				
6 <b>ACTOR</b> closes the form.	1	1	2.5	2.5
Human task cost				11

Table 5: Detailed use case for "Set iteration number" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.3.2 Check learning report

[Paolo Palumbo]

The task is performed by a ML engineer.



Figure 18: "Check learning report" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens "Check training report" form.	1	1	2.50	2.50
2 <b>SYSTEM</b> shows the training loss curve.				
3 <b>ACTOR</b> checks the learning curve.	1	3	2.50	7.50
3.1 <b>IF</b> the loss is flat for at least half of the iterations:	0.4			
3.1.1 <b>THEN ACTOR</b> clicks "Overfit" button.	0.4	1	2.50	1.00
3.2 <b>IF</b> the loss is not flat at the end of the iterations:	0.4			
3.2.1 <b>THEN ACTOR</b> clicks "Underfit" button.	0.4	1	2.50	1.00
3.3 <b>ELSE</b>	0.2			
3.3.1 <b>ACTOR</b> clicks "Approved" button.	0.2	1	2.50	0.50
4 <b>SYSTEM</b> shows a confirmation dialog.				
5 <b>ACTOR</b> closes the form.	1	1	2.50	2.50
Human task cost				15

Table 6: Detailed use case for "Check training report" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.3.3 Check validation report

[Ettore Ricci]

This task is performed by a ML engineer.

	ID	Depth	Neurons	Train MSE	Valid MSE	Delta MSE
1	954	3	4000	0.13	0.14	0.01
2	321	4	3000	0.23	0.24	0.01
3	5	3	1000	0.35	0.35	0.00
4	764	2	2000	0.24	0.45	0.21
5	202	3	2500	0.20	0.47	0.27
Reject						
Overfitting Tolerance:						0.10

Figure 19: "Check validation report" mock-up form



Step	O	CL	S	SC
1 <b>ACTOR</b> opens "Check validation report" form.	1	1	2.5	2.5
2 <b>SYSTEM</b> shows the best 5 models sorted by increasing Validation Loss.				
3 <b>FOR EACH</b> model in the list:	5			
3.1 <b>IF</b> the model Validation Loss minus the Training Loss is less than the Overfitting Tolerance and the Best Model is not selected.	1	2	2.5	5
3.1.1 <b>THEN</b> select the model as the Best Model.	1	1	2.5	2.5
4 <b>FOR EACH</b> model in the list:	4			
4.1 <b>IF</b> the model is not the Best Model and the Validation Loss minus the Training Loss is less than the Overfitting Tolerance and the Second Best Model is not selected.	1	2	2.5	5
4.1.1 <b>THEN</b> select the model as the Second Best Model.	1	1	2.5	2.5
5.1 <b>IF</b> the Best Model is not selected.	0.05	1	2.5	0.125
5.1.1 <b>ACTOR</b> clicks "Reject" button.	0.05	1	2.5	0.125
5.2 <b>ELSE IF</b> the Second Best Model is not selected or the Validation Loss of the Second Best Model is one order of magnitude greater than the Validation Loss of the Best Model.	0.3	3	2.5	2.25
5.2.1 <b>ACTOR</b> clicks on the Best Model.	0.3	1	2.5	0.75
5.3 <b>ELSE</b>	0.65	3	2.5	4.875
5.3.1 <b>ACTOR</b> clicks on the least complex model among the Best Model and the Second Best Model.	0.65	3	2.5	4.875
6 <b>SYSTEM</b> shows a confirmation dialog.				
7 <b>ACTOR</b> closes the form.	1	1	2.5	2.5
Human task cost				32.91

Table 7: Detailed use case for "Check validation report" task  
O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.3.4 Check test results

[Ettore Ricci]

This task is performed by a ML engineer.

ID	Depth	Neurons	Valid MSE	Test MSE	Delta MSE
954	3	4000	0.14	0.15	0.01

Overfitting Tolerance:

0.10

Accept

Reject

Figure 20: "Check test results" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens "Check test results" form.	1	1	2.5	2.5
2 <b>SYSTEM</b> shows the test results.				
3 <b>ACTOR</b> checks if the difference between the test results and the validation results is within overfitting tolerance.	1	2	2.5	5
4.1 <b>IF</b> the test results is not satisfactory.	0.01			
4.1.1 <b>ACTOR</b> clicks "Reject" button.	0.01	1	2.5	0.025
4.2 <b>ELSE</b>	0.99			
4.2.1 <b>ACTOR</b> clicks "Approve" button.	0.99	1	2.5	2.475
5 <b>SYSTEM</b> shows a confirmation dialog.				
6 <b>ACTOR</b> closes the form.	1	1	2.5	2.5
Human task cost				12.5

Table 8: Detailed use case for "Check test results" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.3.5 Configure Development System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.

The mock-up form is titled "CONFIGURE DEVELOPMENT SYSTEM" and "CURRENT CONFIGURATION". It features a central blue gradient bar. On the left, there are two input fields: "Overfitting Tolerance" with value "1" and "Generalization Tolerance" with value "1". Below these is a section titled "UPLOAD NEW CONFIGURATION" containing a document icon with an upload arrow and an "Upload" button. On the right, there are six input fields: "Min Neurons" (1), "Max Neurons" (200), "Step Neurons" (1), "Min layers" (10), "Max Layers" (10), and "Step Layers" (1).

Figure 21: "Configure Development System" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens the "Configure Development System" form.	1	1	2.50	2.50
2 <b>SYSTEM</b> displays current configuration and "Upload" button.				
3 <b>ACTOR</b> checks parameters against previous iterations on file	1	3	2.50	7.50
4 <b>ACTOR</b> adjusts file based on current parameters	1	3	2.50	7.50
5 <b>ACTOR</b> pushes "Upload" button and uploads configuration file	1	1	2.50	2.50
6.1 <b>SYSTEM</b> IF config is correct and correctly formatted				
6.1.1 <b>SYSTEM</b> shows a confirmation message.				
6.2 <b>ELSE</b>				
6.2.1 <b>SYSTEM</b> shows error message and aborts				
7 <b>ACTOR</b> closes the form.	1	1	2.50	2.50
Human task cost				22.50

Table 9: Detailed use case for "Configure Development" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.4 Evaluation system

#### 3.4.1 Evaluate classifier performance

[Zahra Omrani]

This task is performed by a Data Analyst.

×

Evaluate Classifier Performance

Session ID	Expert Label	Classifier Label	Error
0	1	2	Yes
1	1	3	Yes
2	2	1	Yes
3	3	3	No

Max number of errors tolerated (th1): 4  
Max number of consecutive error tolerated (th 2) :2

th1 satisfied  
3 < 4

th 2 exceeded  
3 > 2

Pass

Fail

Figure 22: "Evaluate Classifier Performance" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens the "Evaluate Classifier Performance" form.	1	1	1.15	1.15
2 <b>SYSTEM</b> displays a table of sessions with Expert Label (ground truth) and Classifier Label (predicted label). The difference between the labels (if any) represents an error.				
3.1 <b>ACTOR</b> checks the total errors threshold color.	1	2	1.15	2.30
3.2 <b>ACTOR</b> checks the consecutive errors threshold color	1	2	1.15	2.30
3.3 <b>IF</b> at least one threshold is red				
3.3.1 <b>ACTOR</b> clicks the "Fail" button.	0.4	1	1.15	0.46
3.4 <b>ELSE</b>				
3.4.1 <b>ACTOR</b> clicks the "Pass" button.	0.6	1	1.15	0.65
4 <b>SYSTEM</b> shows a confirmation dialog.				
5 <b>ACTOR</b> closes the form.	1	1	1.15	1.15
Human task cost				8.05

Table 10: Detailed use case for "Evaluate Classifier Performance" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.4.2 Configure Evaluation System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.

The mock-up form is titled "CONFIGURE EVALUATION SYSTEM" and "CURRENT CONFIGURATION". It features three input fields with values: "Max Errors" set to 2, "Max Consecutive Errors" set to 5, and "Required Labels" set to 5. Below these is a section titled "UPLOAD NEW CONFIGURATION" which includes a document icon with a circular arrow and an "Upload" button.

Figure 23: "Configure Evaluation System" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens the "Configure Evaluation System" form.	1	1	2.50	2.50
2 <b>SYSTEM</b> displays current configuration and "Upload" button.				
3 <b>ACTOR</b> checks parameters against previous iterations on file	1	3	2.50	7.50
4 <b>ACTOR</b> adjusts file based on current parameters	1	3	2.50	7.50
5 <b>ACTOR</b> pushes "Upload" button and uploads configuration file	1	1	2.50	2.50
6.1 <b>SYSTEM</b> IF config is correct and correctly formatted				
6.1.1 <b>SYSTEM</b> shows a confirmation message.				
6.2 <b>ELSE</b>				
6.2.1 <b>SYSTEM</b> shows error message and aborts				
7 <b>ACTOR</b> closes the form.	1	1	2.50	2.50
Human task cost				22.50

Table 11: Detailed use case for "Configure Evaluation" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.5 Client-Side Systems

#### 3.5.1 Configure Client-Side Systems

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.

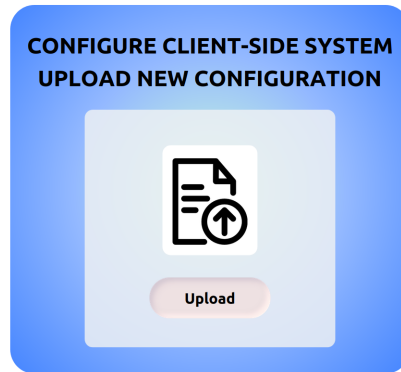


Figure 24: "Configure Client-Side Systems" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens the "Configure Client-Side System" form.	1	1	2.50	
2 <b>SYSTEM</b> displays the "Upload" button.				
3 <b>ACTOR</b> push the "Upload" button and upload the configuration file.	1	1	2.50	2.50
4 <b>SYSTEM</b> shows a confirmation message.				
5 <b>ACTOR</b> closes the form.	1	1	2.50	2.50
Human task cost				7.50

Table 12: Detailed use case for "Configure Client-Side Systems" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.6 Production System

#### 3.6.1 Configure Production Systems

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.

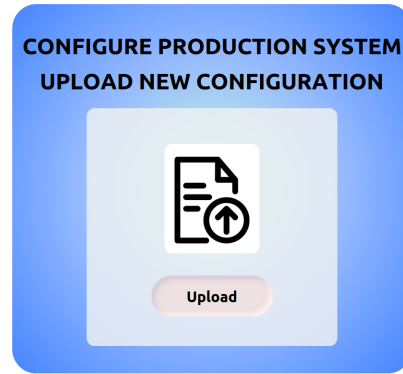


Figure 25: "Configure Production System" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens the "Configure Production System" form.	1	1	2.50	2.50
2 <b>SYSTEM</b> displays the "Upload" button.				
3 <b>ACTOR</b> push the "Upload" button and upload the configuration file.	1	1	2.50	2.50
4 <b>SYSTEM</b> shows a confirmation message.				
5 <b>ACTOR</b> closes the form.	1	1	2.50	2.50
Human task cost				7.50

Table 13: Detailed use case for "Configure Production" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.7 Ingestion System

#### 3.7.1 Configure Ingestion System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.

Figure 26: "Configure Ingestion System" mock-up form

Step	O	CL	S	SC
1 <b>ACTOR</b> opens the "Configure Ingestion System" form.			2.50	
2 <b>SYSTEM</b> displays current configuration and "Upload" button.				
3 <b>ACTOR</b> checks parameters against previous iterations on file	1	3	2.50	7.50
4 <b>ACTOR</b> adjusts file based on current parameters	1	3	2.50	7.50
5 <b>ACTOR</b> pushes "Upload" button and uploads configuration file	1	1	2.50	2.50
6.1 <b>SYSTEM</b> IF config is correct and correctly formatted				
6.1.1 <b>SYSTEM</b> shows a confirmation message.				
6.2 <b>ELSE</b>				
6.2.1 <b>SYSTEM</b> shows error message and aborts				
7 <b>ACTOR</b> closes the form.	1	1	2.50	2.50
Human task cost				22.50

Table 14: Detailed use case for "Configure Ingestion" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

### 3.8 Preparation System

#### 3.8.1 Configure Preparation System

[*Francesco Boldrini, Zahra Omrani*]

This task is performed by a ML Engineer.

Figure 27: "Configure Preparation System" mock-up form

Step	O	CL	S	SC
<b>1 ACTOR</b> opens the "Configure Preparation System" form.			2.50	
<b>2 SYSTEM</b> displays current configuration and "Upload" button.				
<b>3 ACTOR</b> checks parameters against previous iterations on file	1	3	2.50	7.50
<b>4 ACTOR</b> adjusts file based on current parameters	1	3	2.50	7.50
<b>5 ACTOR</b> pushes "Upload" button and uploads configuration file	1	1	2.50	2.50
<b>6.1 SYSTEM</b> IF config is correct and correctly formatted				
<b>6.1.1 SYSTEM</b> shows a confirmation message.				
<b>6.2 ELSE</b>				
<b>6.2.1 SYSTEM</b> shows error message and aborts				
<b>7 ACTOR</b> closes the form.	1	1	2.50	2.50
Human task cost				22.50

Table 15: Detailed use case for "Configure Preparation" task

O - Occurrence, CL - Cognitive Level, S - Normalized Salary, SC - Step Cost

## 4 Simulation

### 4.1 Collapsed workflow

[Ettore Ricci, Paolo Palumbo, Francesco Boldrini]



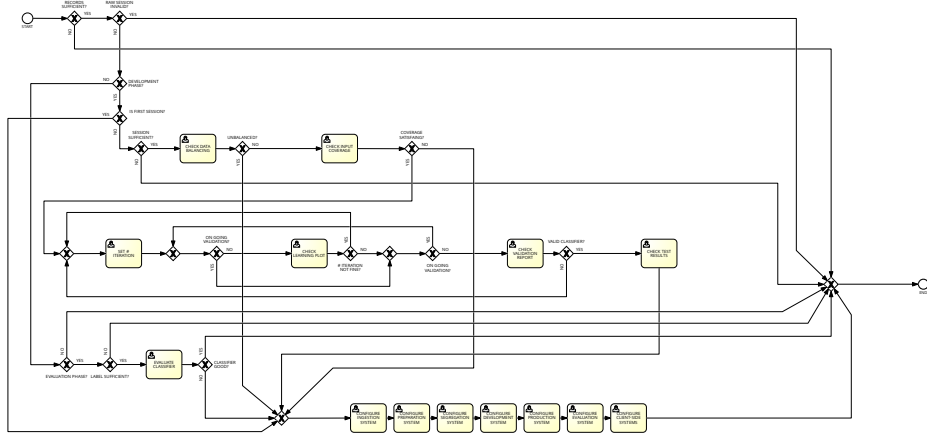


Figure 28: Collapsed workflow

## 5 Process mining

We mined the logs generated by the simulation of the collapsed workflow.

We modified the simulation configuration to make the 100 tokens flow through every path of the workflow. The most important gateways that we changed are listed in the following table.

Gateway	Yes	No
RAW SESSION INVALID	5%	95%
RECORD SUFFICIENT	95%	5%
SESSION SUFFICIENT	95%	5%
IS FIRST SESSION	20%	80%
COVERAGE SATISFYING	70%	30%
DEVELOPMENT PHASE	70%	30%

Table 16: Gateways configuration

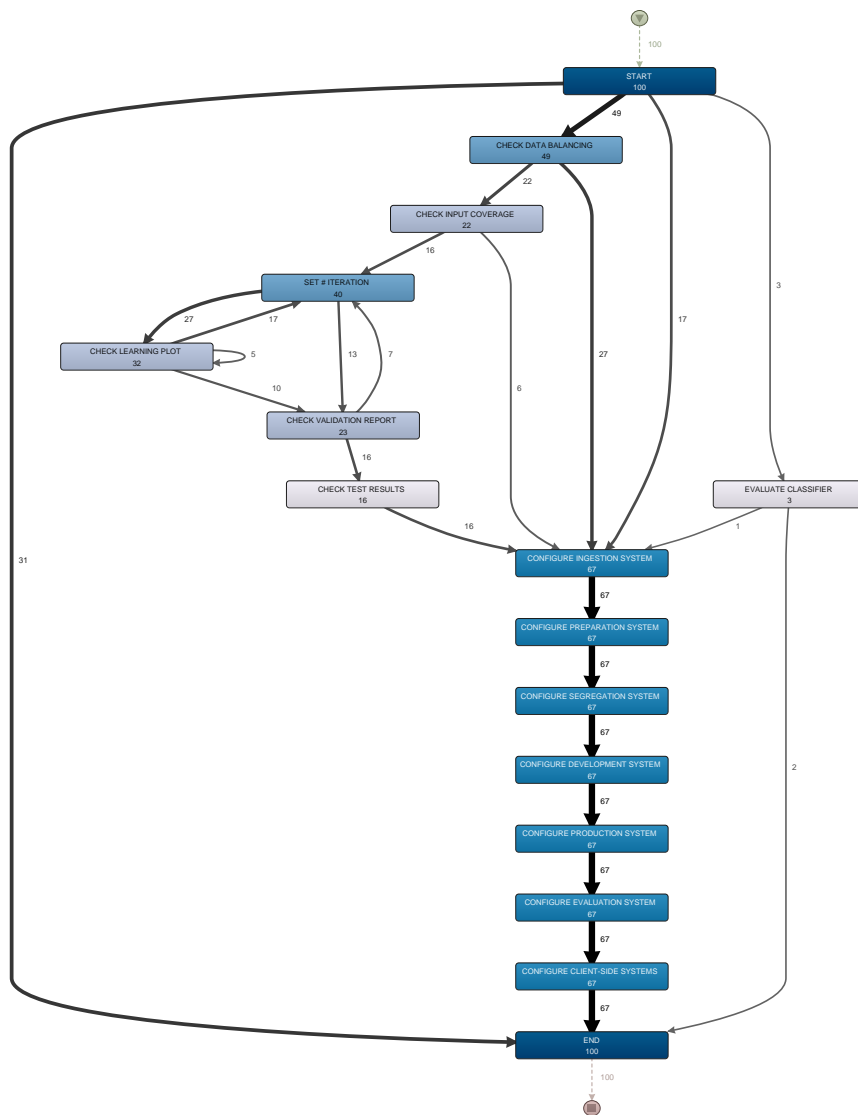


Figure 29: Disco analysis

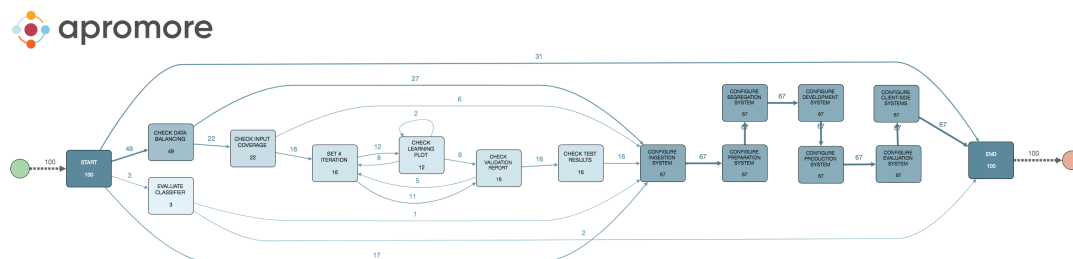


Figure 30: Apromore analysis

As we can see, the two transition maps mined from Disco and from Apromore are identical. The only difference stays in the frequencies because in Disco the frequencies are calculated as the total number of times a transition is executed, even on the same token; while in Apromore the frequencies are calculated as the number of individual tokens that execute a transition. This behavior can be changed with a setting in both tools.

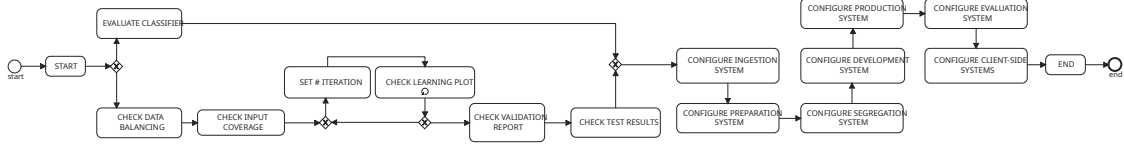


Figure 31: ProM mined BPMN model

We mined the logs using the "Heuristics Miner ProM6" mining algorithm.

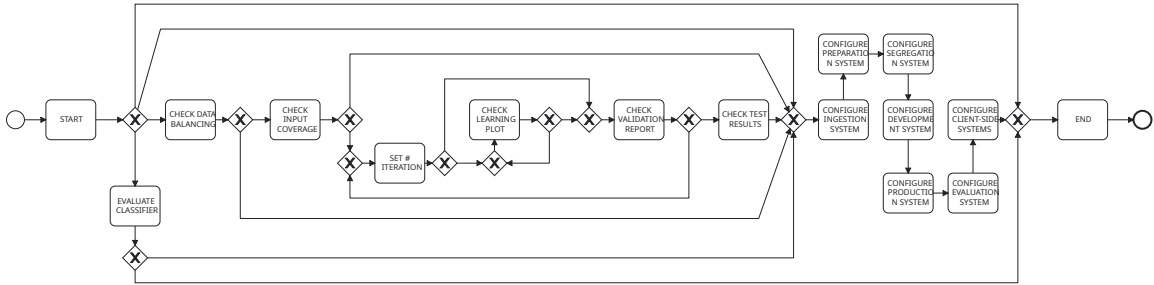


Figure 32: Apromore mined BPMN model

The BPMN model mined from Apromore is more detailed and covers more cases than the one mined from ProM. The key differences between the ProM model and the Apromore one are that the ProM model is missing the paths that skip the training and the configuration as well as one of the two paths that skip only the training. Furthermore, the training loop is much simpler in the ProM model, as it is missing every path that restarts the training after "CHECK VALIDATION REPORT".

Tool	Trace	Generalization	Precision	Simplicity
Apromore	0.4203	0.9872	0.7566	62
ProM	0.2917	0.9871	0.9871	39

Table 17: Comparison of the process mining tools

## 5.1 Violations

We modified the logs to introduce 3 violations in the workflow. The violations are the following:

1. Skipping the dataset creation ("CHECK DATA BALANCING" and "CHECK INPUT COVERAGE") using data from another user.
2. Skipping "SET # ITERATIONS" and "CHECK LEARNING PLOT" because we are using early stopping.
3. Skipping "CHECK DATA BALANCING" because we are using a resampling technique.

Each violation is introduced 3 times in the logs.

CaseID	Violation	Fitness ProM	Fitness Apromore
10	1	0.91	0.87
20	1	0.85	0.84
47	1	0.86	0.86
53	2	0.91	0.93
63	2	0.84	0.82
88	2	0.91	0.93
6	3	0.91	0.93
72	3	0.91	0.85
81	3	0.94	0.87

Table 18: Cases, violations and fitness on models generated by ProM and Apromore