

Process Mining and Intelligence Project

Emotion Based Music Selection

Ettore Ricci Francesco Boldrini Paolo Palumbo Zahra Omrani

January 27, 2025

Contents

1	BPMN modeling	2
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	4
1.2.5	Evaluate classifier performance	4
1.2.6	Configure systems	4
2	Data modeling	5
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	8
3	Task level modeling	9
3.1	Roles and salaries	9
3.2	Segregation system	11
3.2.1	Check data balancing	11
3.2.2	Check input coverage	12
3.2.3	Configure Segregation System	13
3.3	Development system	14
3.3.1	Set iteration number	14
3.3.2	Check learning report	15
3.3.3	Check validation report	16
3.3.4	Check test results	18
3.3.5	Configure Development System	19
3.4	Evaluation system	20
3.4.1	Evaluate classifier performance	20
3.4.2	Configure Evaluation System	21
3.5	Client-Side Systems	22
3.5.1	Configure Client-Side Systems	22
3.6	Production System	23
3.6.1	Configure Production Systems	23
3.7	Ingestion System	23
3.7.1	Configure Ingestion System	23
3.8	Preparation System	24

3.8.1 Configure Preparation System	24
4 Simulation	25
4.1 Collapsed workflow	25
5 Process mining	26

1 BPMN modeling

1.1 Process landscape

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

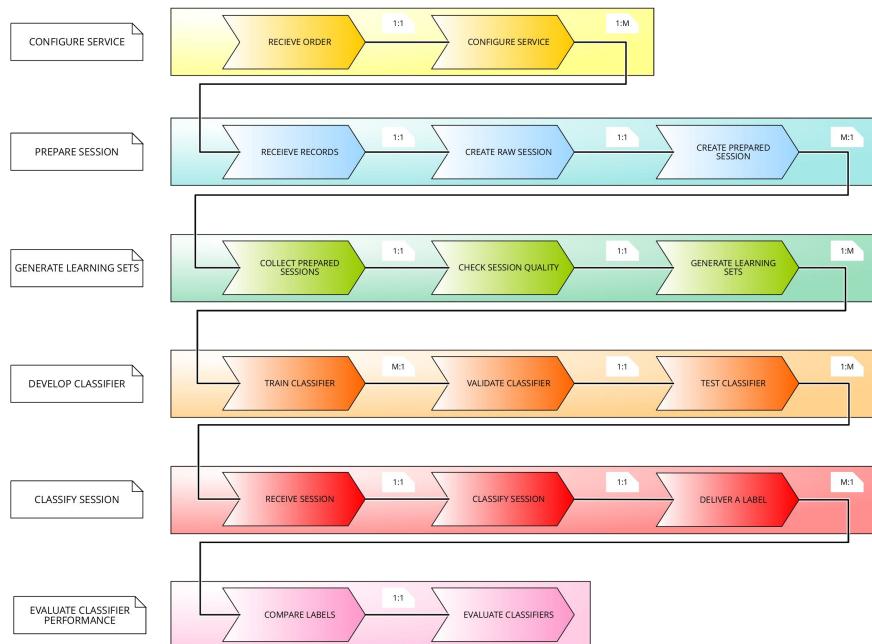


Figure 1: Process landscape

1.2 Process model

1.2.1 Prepare session

[*Ettore Ricci, Paolo Palumbo*]

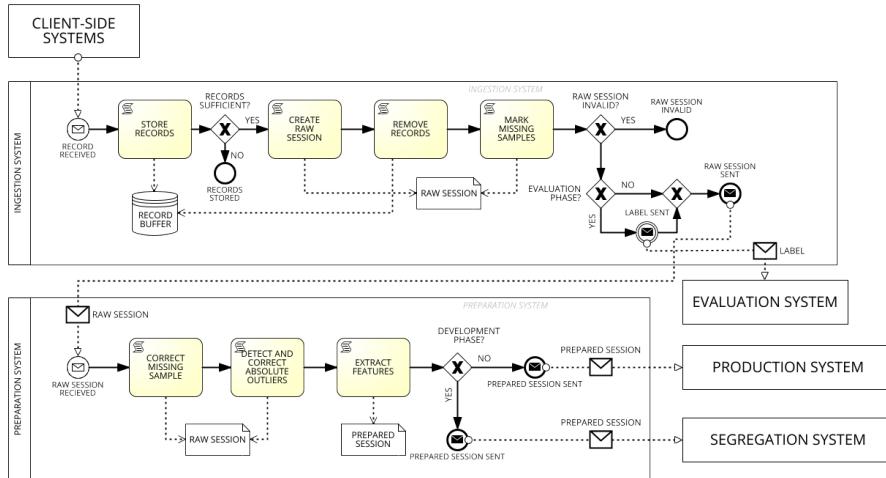


Figure 2: Business Diagram of the "Prepare session" process

1.2.2 Generate learning sets

[Ettore Ricci, Paolo Palumbo]

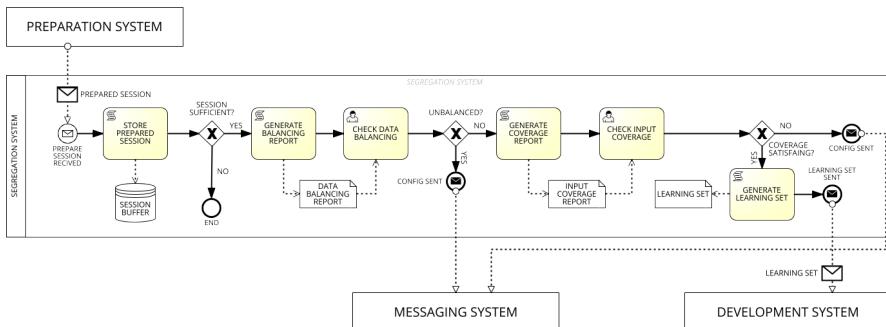


Figure 3: Business Diagram of the "Generate learning sets" process

1.2.3 Develop classifier

[Ettore Ricci, Paolo Palumbo]

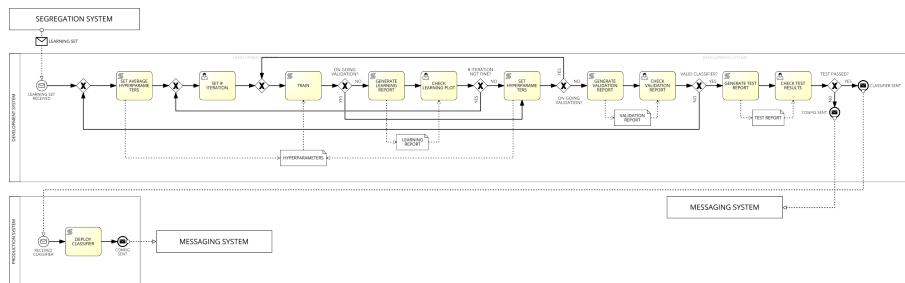


Figure 4: Business Diagram of the "Develop classifier" process

1.2.4 Classify session

[Ettore Ricci, Paolo Palumbo]

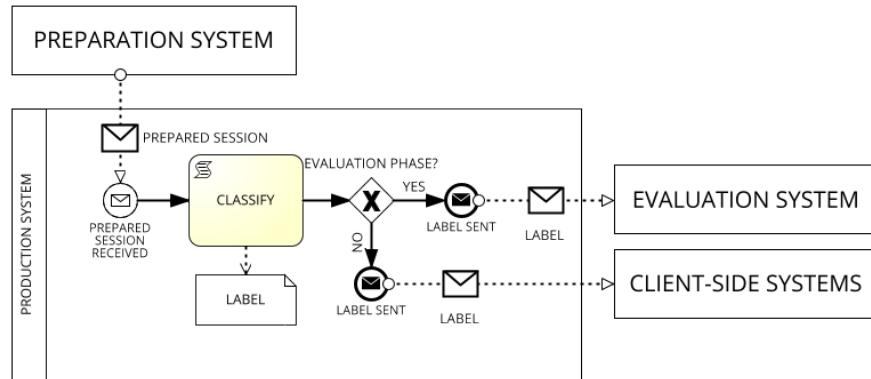


Figure 5: Business Diagram of the "Classify session" process

1.2.5 Evaluate classifier performance

[Ettore Ricci, Paolo Palumbo]

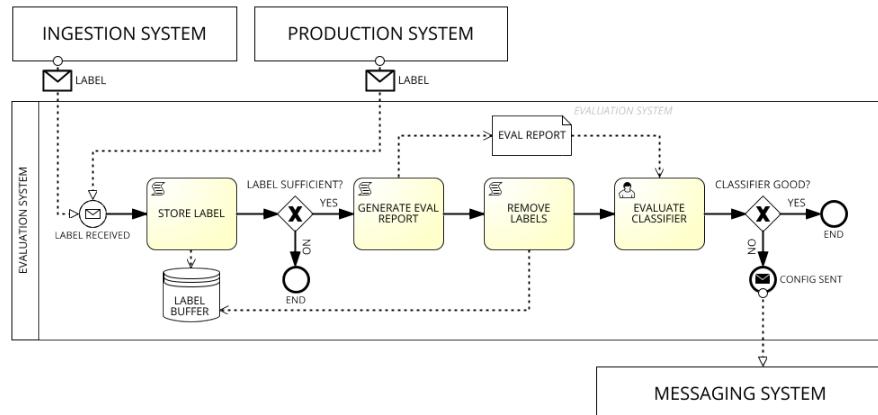


Figure 6: Business Diagram of the "Evaluate classifier performance" process

1.2.6 Configure systems

[Ettore Ricci, Paolo Palumbo]

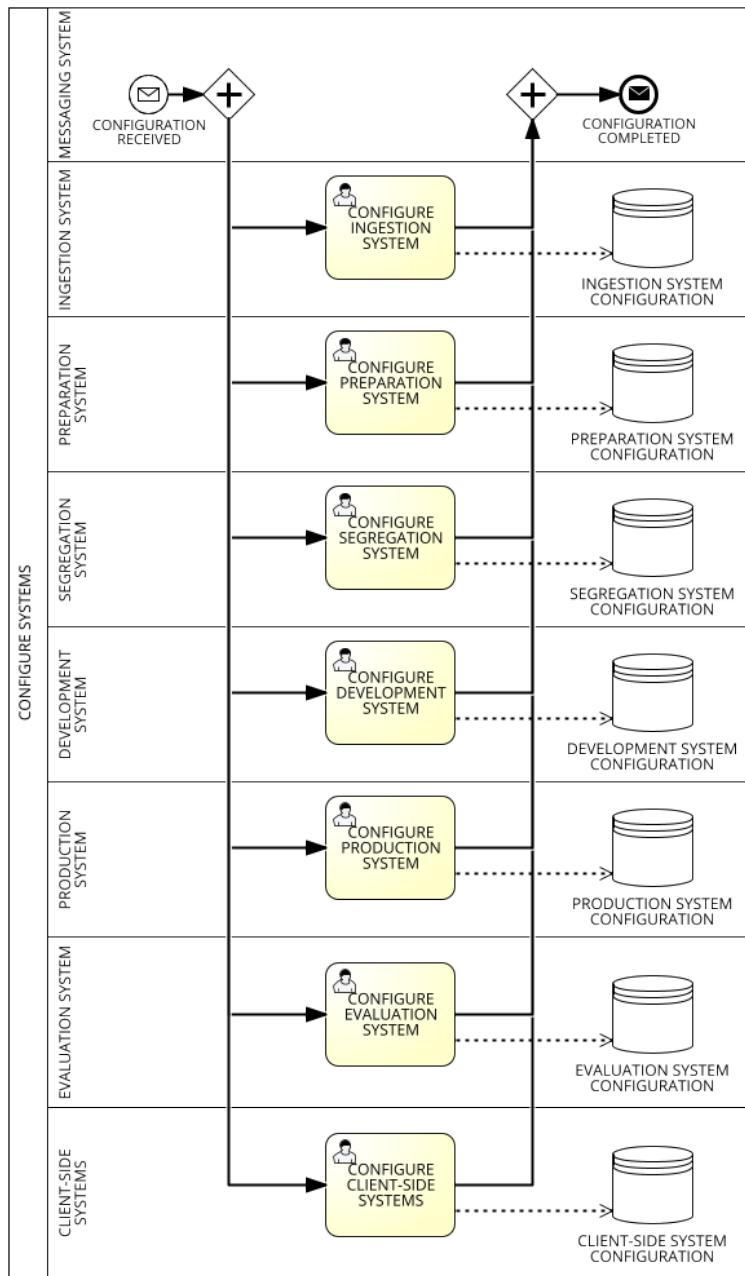


Figure 7: Business Diagram of the "Configure systems" process

2 Data modeling

2.1 Process model

2.1.1 Prepare session

[*Ettore Ricci*]

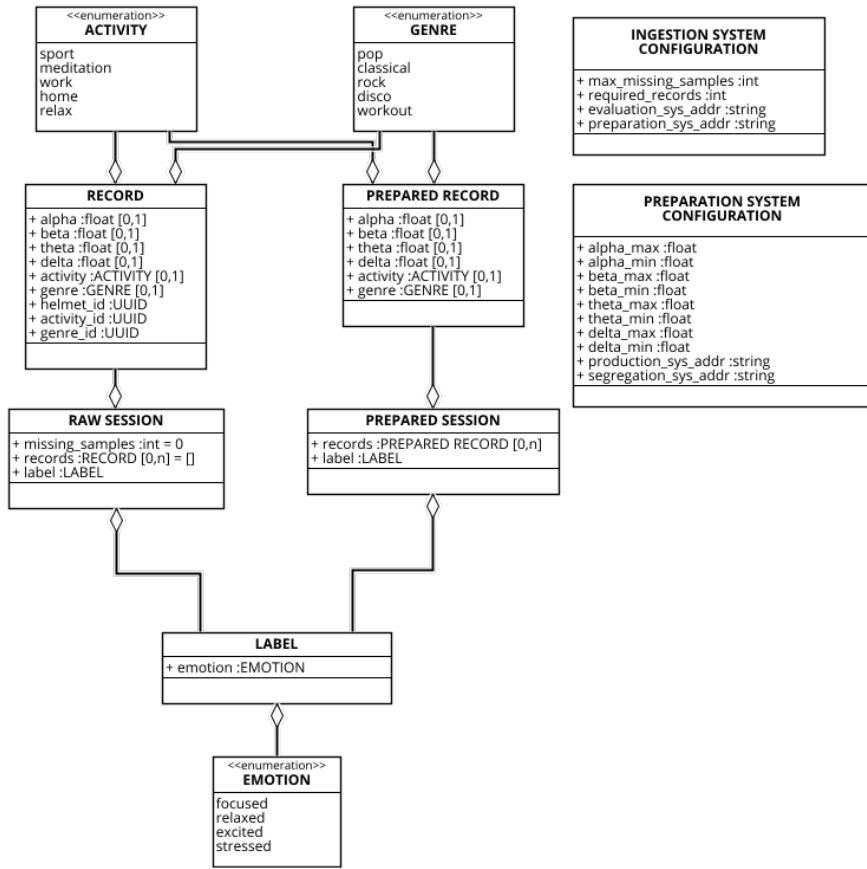


Figure 8: Data Model of the "Prepare session" process

2.1.2 Generate learning sets

[*Paolo Palumbo*]

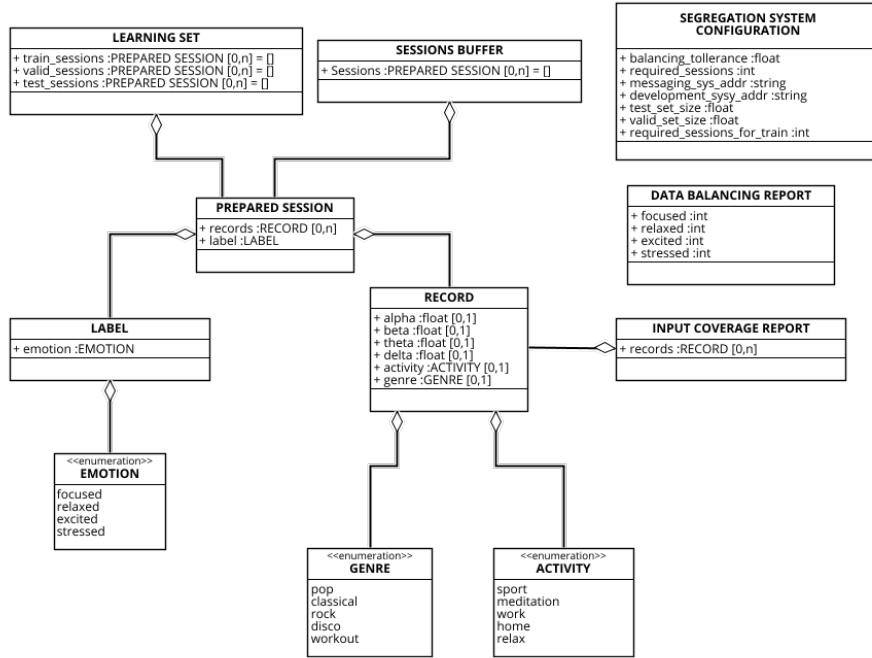


Figure 9: Data Model of the "Generate learning sets" process

2.1.3 Develop classifier

[*Paolo Palumbo*]

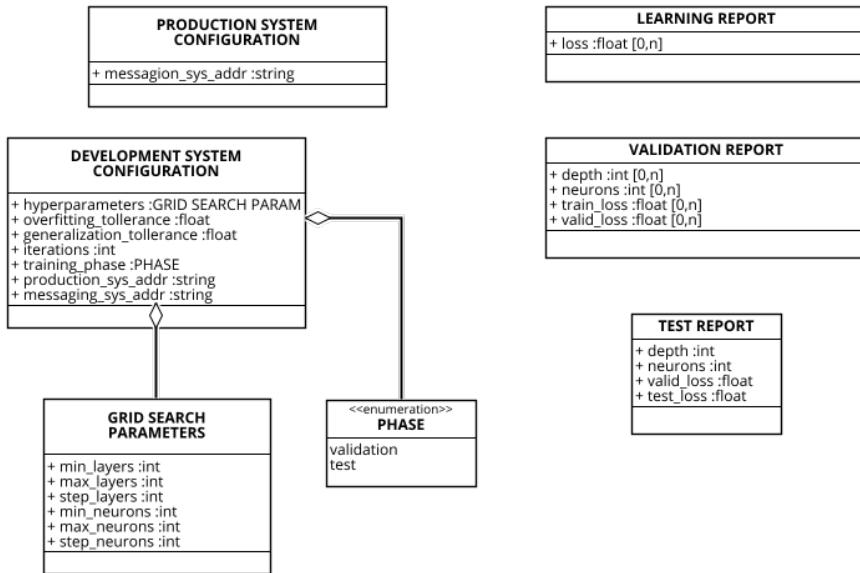


Figure 10: Data Model of the "Develop classifier" process

2.1.4 Classify session

[*Francesco Boldrini*]

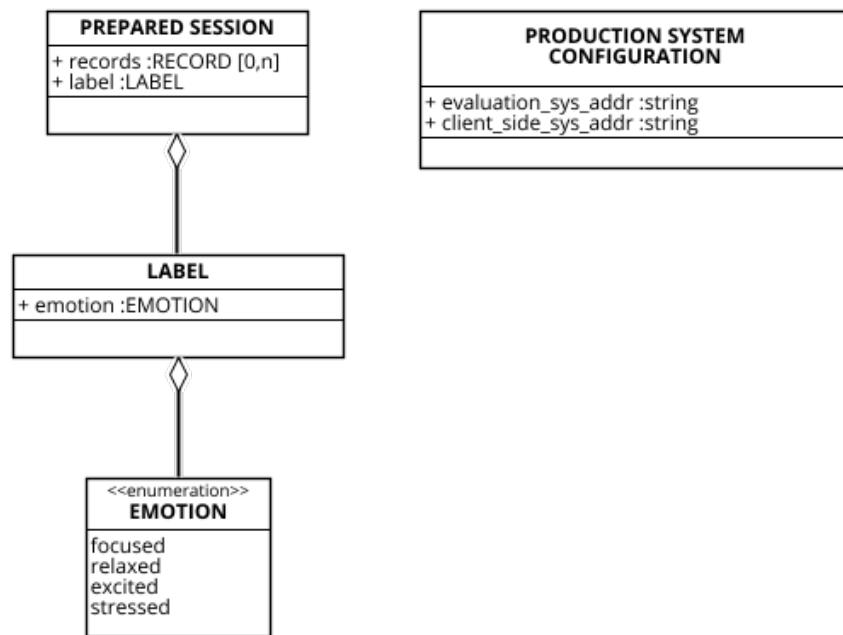


Figure 11: Data Model of the "Classify session" process

2.1.5 Evaluate classifier performance

[*Zahra Omrani*]

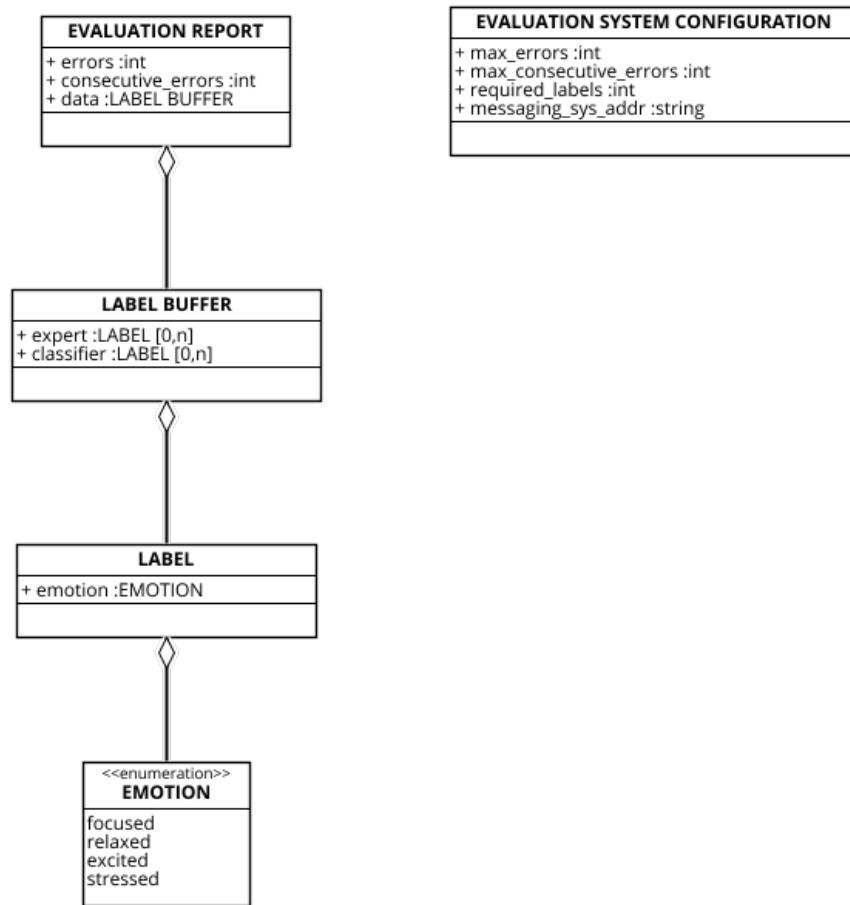


Figure 12: Data Model of the "Evaluate classifier performance" process

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
Minimum		\$52,000.00	1.00

Table 1: Salary and normalized salary for each position

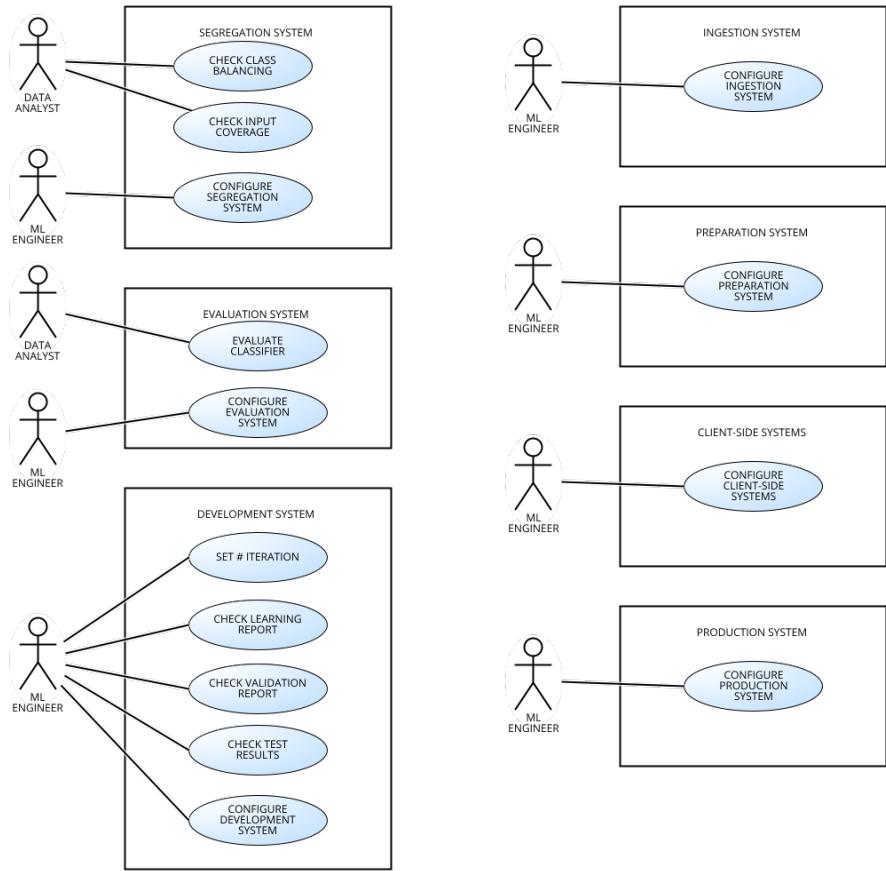


Figure 13: Use case diagram

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

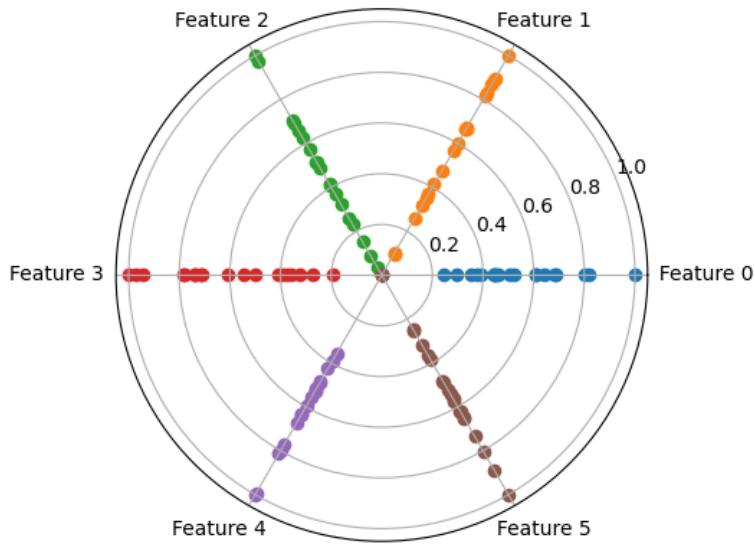
Step	O	CL	S	SC
1 ACTOR opens "Check data balancing" form.	1	1	1.15	1.15
2 SYSTEM shows the report.				
3 SYSTEM shows a hint whether the data is balanced or not.				
4 ACTOR checks the hint to see if the data is balanced or not.	1	2	1.15	2.30
5.1 IF the data is balanced.	0.2			
5.1.1 ACTOR clicks "Balanced"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



Accept

Reject

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

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

The form is titled "CONFIGURE SEGREGATION SYSTEM". It has a blue header bar with the title. Below it is a white card-like area. The top part of the card is titled "CURRENT CONFIGURATION" and contains two buttons: one for "Balancing Tollerance" set to 2 and another for "Required Sessions" set to 5. Below this is another card titled "UPLOAD NEW CONFIGURATION" containing a file icon and a pink "Upload" button.

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

Step	O	CL	S	SC
1 ACTOR opens the "Configure Segregation System" form.	1	1	2.50	2.50
2 SYSTEM displays current configuration and "Upload" button.				
3 ACTOR checks parameters against previous iterations on file	1	3	2.50	7.50
4 ACTOR adjusts file based on current parameters	1	3	2.50	7.50
5 ACTOR pushes "Upload" button and uploads configuration file	1	1	2.50	2.50
6.1 SYSTEM IF config is correct and correctly formatted				
6.1.1 SYSTEM shows a confirmation message.				
6.2 ELSE				
6.2.1 SYSTEM shows error message and aborts				
7 ACTOR 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 form has a header "Set Iteration Number" with a close button (X). It contains a "Current Iteration Number" field showing "10" and an "Enter New Iteration Number:" field with an empty input box. Below these is a "Submit" button. At the bottom, a message box displays "Iteration number updated successfully!".

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

Step	O	CL	S	SC
1 ACTOR opens "Set Iteration Number" form.	1	1	2.5	2.5
2 SYSTEM displays the current iteration number.				
3.1 IF it's the first configuration:				
3.1.1 ACTOR inputs the desired number of iterations based on task complexity and previous experience.	0.2	3	2.5	1.5
3.2 ELSE (subsequent configurations):				
3.2.1 ACTOR inputs the number based on the established learning curve.	0.8	1	2.5	2
4 ACTOR clicks "Submit" button to confirm the iteration number.	1	1	2.5	2.5
5 SYSTEM shows a confirmation dialog.				
6 ACTOR 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 ACTOR opens "Check training report" form.	1	1	2.50	2.50
2 SYSTEM shows the training loss curve.				
3 ACTOR checks the learning curve.	1	3	2.50	7.50
3.1 IF the loss is flat for at least half of the iterations:	0.4			
3.1.1 THEN ACTOR clicks "Overfit" button.	0.4	1	2.50	1.00
3.2 IF the loss is not flat at the end of the iterations:	0.4			
3.2.1 THEN ACTOR clicks "Underfit" button.	0.4	1	2.50	1.00
3.3 ELSE	0.2			
3.3.1 ACTOR clicks "Approved" button.	0.2	1	2.50	0.50
4 SYSTEM shows a confirmation dialog.				
5 ACTOR 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 ACTOR opens "Check validation report" form.	1	1	2.5	2.5
2 SYSTEM shows the best 5 models sorted by increasing Validation Loss.				
3 FOR EACH model in the list:	5			
3.1 IF 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 THEN select the model as the Best Model.	1	1	2.5	2.5
4 FOR EACH model in the list:	4			
4.1 IF 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 THEN select the model as the Second Best Model.	1	1	2.5	2.5
5.1 IF the Best Model is not selected.	0.05	1	2.5	0.125
5.1.1 ACTOR clicks "Reject" button.	0.05	1	2.5	0.125
5.2 ELSE IF 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 ACTOR clicks on the Best Model.	0.3	1	2.5	0.75
5.3 ELSE	0.65	3	2.5	4.875
5.3.1 ACTOR clicks on the least complex model among the Best Model and the Second Best Model.	0.65	3	2.5	4.875
6 SYSTEM shows a confirmation dialog.				
7 ACTOR closes the form.	1	1	2.5	2.5
	Human task cost			

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

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". It features a central blue box labeled "CURRENT CONFIGURATION" containing various configuration parameters. The parameters and their values are:

- Overfitting Tolerance: 1
- Generalization Tolerance: 1
- Min Neurons: 1
- Max Neurons: 200
- Step Neurons: 1
- Min layers: 10
- Max Layers: 10
- Step Layers: 1

Below the configuration box is a section titled "UPLOAD NEW CONFIGURATION" featuring a file icon and a "Upload" button.

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

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

The form is titled "Evaluate Classifier Performance". It contains a table with four columns: Session ID, Expert Label, Classifier Label, and Error. The data is as follows:

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

Below the table, there are two status indicators:

- A green box labeled "th1 satisfied 3 < 4"
- A red box labeled "th 2 exceeded 3 > 2"

At the bottom, there are two buttons: "Pass" and "Fail".

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

Step		O	CL	S	SC
1 ACTOR opens the "Evaluate Classifier Performance" form.		1	1	1.15	1.15
2 SYSTEM 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 ACTOR checks the total errors threshold color.		1	2	1.15	2.30
3.2 ACTOR checks the consecutive errors threshold color		1	2	1.15	2.30
3.3 IF at least one threshold is red					
3.3.1 ACTOR clicks the "Fail" button.		0.4	1	1.15	0.46
3.4 ELSE					
3.4.1 ACTOR clicks the "Pass" button.		0.6	1	1.15	0.65
4 SYSTEM shows a confirmation dialog.					
5 ACTOR 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". It features a central blue box labeled "CURRENT CONFIGURATION" containing three settings: "Max Errors" (value 2), "Max Consecutive Errors" (value 5), and "Required Labels" (value 5). Below this is a section titled "UPLOAD NEW CONFIGURATION" with a file icon and a "Upload" button.

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

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

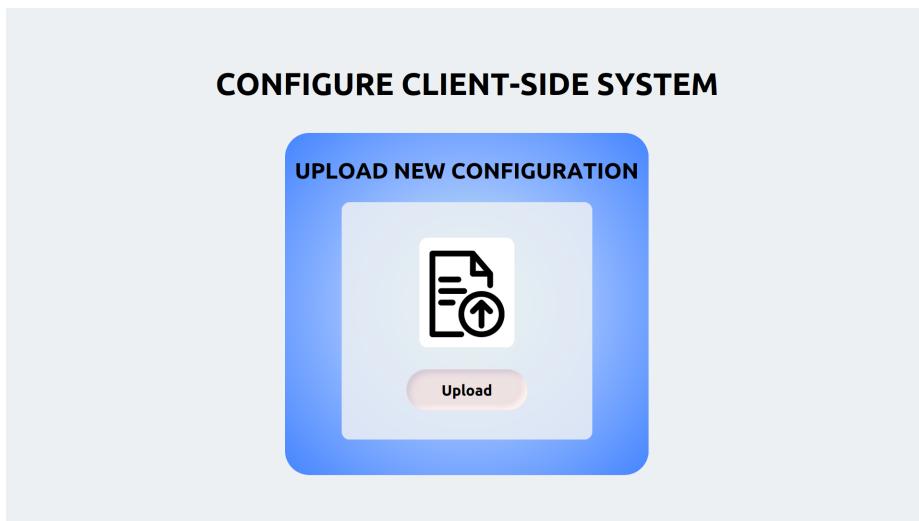


Figure 24: "Configure Client-Side Systems" mock-up form

Step	O	CL	S	SC
1 ACTOR opens the "Configure Client-Side System" form.	1	1	2.50	
2 SYSTEM displays the "Upload" button.				
3 ACTOR push the "Upload" button and upload the configuration file.	1	1	2.50	2.50
4 SYSTEM shows a confirmation message.				
5 ACTOR 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.

The mock-up form is titled "CONFIGURE PRODUCTION SYSTEM" in bold capital letters at the top center. Below the title is a blue rectangular area with rounded corners. Inside this area, the text "UPLOAD NEW CONFIGURATION" is centered at the top. Below this text is a white square containing a black icon of a document with an upward arrow. At the bottom of the blue area is a pink rounded rectangle containing the word "Upload" in white capital letters.

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

Step	O	CL	S	SC
1 ACTOR opens the "Configure Production System" form.	1	1	2.50	2.50
2 SYSTEM displays the "Upload" button.				
3 ACTOR push the "Upload" button and upload the configuration file.	1	1	2.50	2.50
4 SYSTEM shows a confirmation message.				
5 ACTOR 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.

The form is titled "CONFIGURE INGESTION SYSTEM". It has a blue header and a white body. In the header, there is a title "CURRENT CONFIGURATION" and two input fields: "Max Missing Samples" with value "2" and "Required Records" with value "5". Below this is a section titled "UPLOAD NEW CONFIGURATION" containing a file icon and a "Upload" button.

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

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

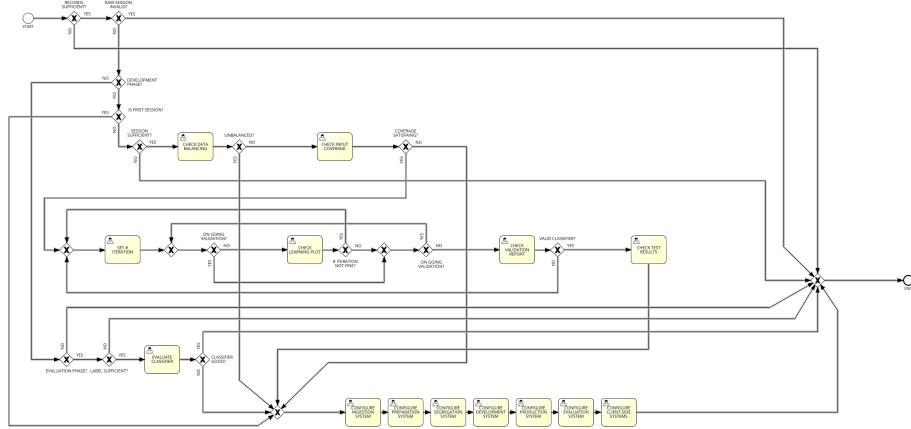


Figure 28: Collapsed workflow

5 Process mining

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

We modified the simulation configuration

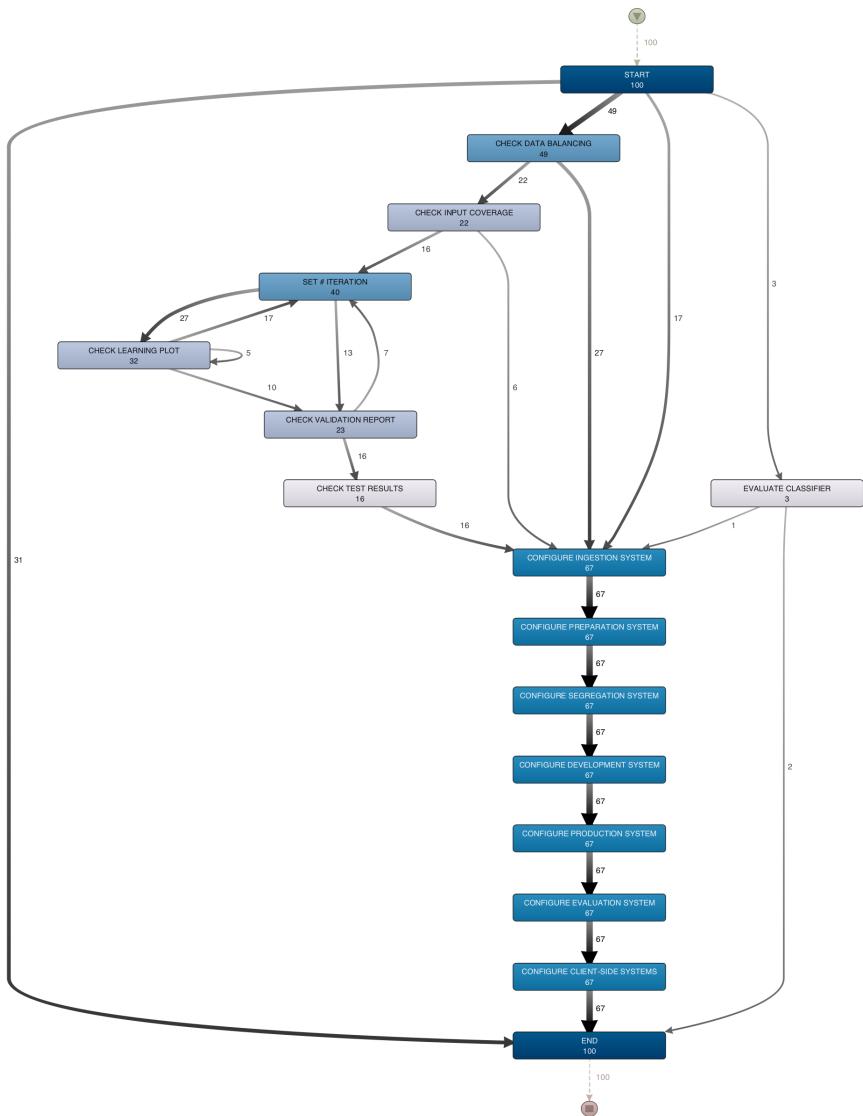


Figure 29: Disco analysis

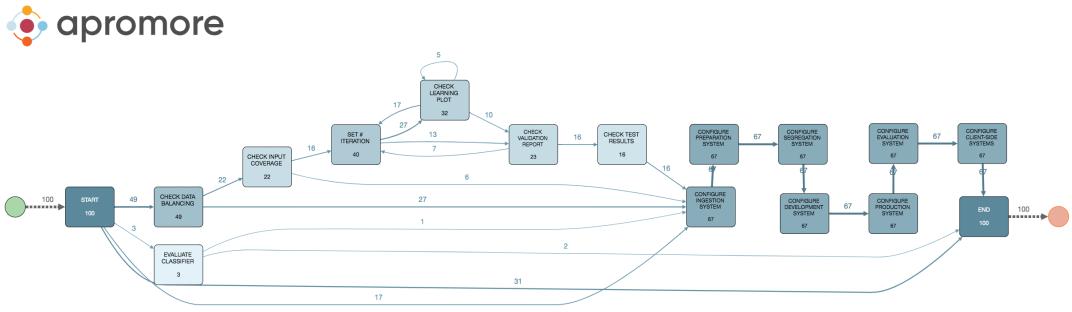


Figure 30: Apromore analysis

As we can see, the two transition maps mined from Disco and from Apromore are identical.

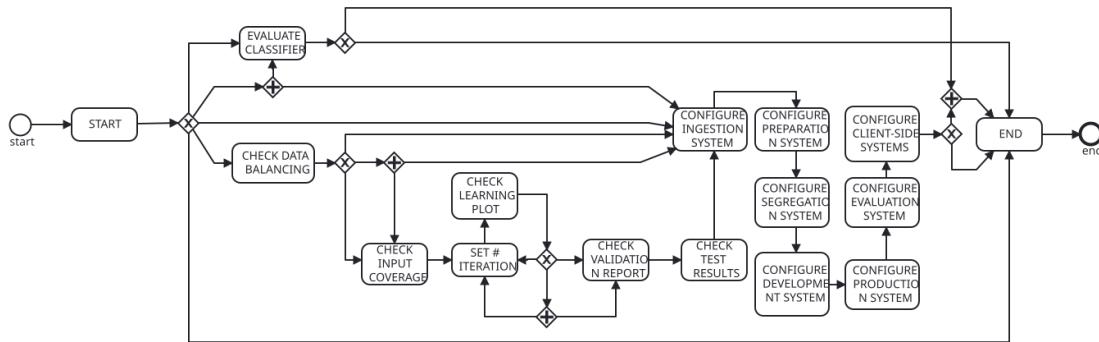


Figure 31: ProM mined BPMN model

The BPMN model mined from ProM is not compliant with the BPMN 2.0 specification (section 10.5). Furthermore, the model contains parallel gateways that are never present in the collapsed model.

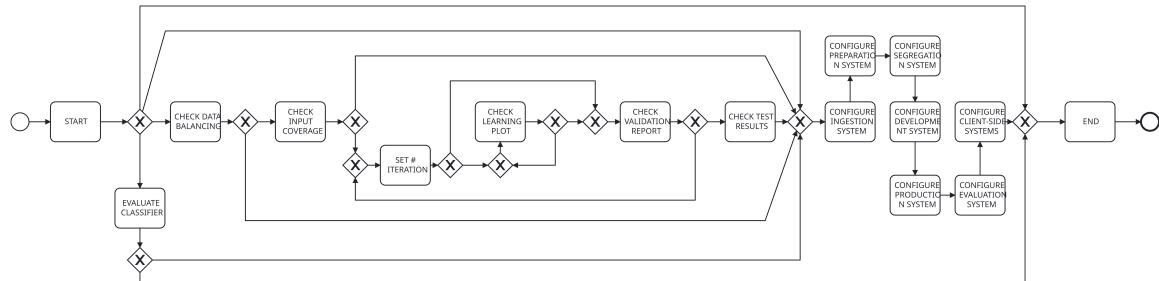


Figure 32: Apromore mined BPMN model

The BPMN model mined from Apromore is much more sensible compared to the one from ProM, also it is compliant with the BPMN 2.0 specification.

Tool	Trace	Generalization	Precision	Simplicity
Apromore	0.4203	0.9872	0.7566	
ProM	0.6666	0	0	

Table 16: Comparison of the process mining tools