BPI CHALLENGE 2017:
A Goal-Oriented Approach to Loan Prioritization using Process Mining

Ilaria Boschetto, Federico Clerici, Raffaele D'Agostino, Gabriele Villa, Davide Volpi

Introduction

This report presents a comprehensive analysis of process improvement opportunities within a loan application process from a Dutch financial institution, utilizing the **BPI Challenge 2017** event log dataset. The study applies process mining techniques and predictive analytics to address operational inefficiencies and enhance process performance. The primary objective is to identify and evaluate automation opportunities that reduce processing time, optimize resource allocation, and improve overall loan application throughput.
Building on established process mining methodologies, this analysis integrates three key components presented sequentially:

1. **Machine learning-based loan prioritization** to predict applications with the highest probability of successful completion
2. **Automation potential assessment** to identify high-value activities suitable for automation
3. **A modified process model** incorporating goal-oriented improvements that align process redesign with strategic business objectives. The integration of these approaches enables data-driven decision-making and provides actionable insights for operational excellence in loan processing.

Main Contribution

Building upon the valuable analysis conducted by the winning teams of the BPI Challenge 2017, this project addresses two critical business questions for the financial institution:

- **BQ1**: **"How can we reduce the workload on human resources and speed up the process?"**
- **BQ2**: **"How can we identify and prioritize applications with the highest probability of success to optimize the process?"**

To answer these, we propose a combined approach of process automation and predictive prioritization.
Our main contributions are:

1. **Predictive Loan Prioritization (with Machine Learning):** We implemented a Machine Learning model specifically designed to predict the likelihood of the client to accept a loan offer. By analyzing early process features, the model assigns a success probability score to each running application.
    - <u>Business Impact:</u> This allows the bank to move from a First-In-First-Out (FIFO) approach to a value-based prioritization strategy. Resources can focus their efforts on applications with the highest probability of success, reducing time spent on cases destined to be declined or cancelled.
2. **Strategic Process Automation:** We developed a goal model to identify which activities are most suitable for automation based on three key dimensions: execution frequency, temporal stability (Relative IQR), and resource concentration.
    - <u>Business Impact:</u> By automating repetitive, high-volume tasks (such as W_Validate application or communication triggers), we free up human workforce capacity. This

"saved time" can then be reallocated to the high-priority cases identified by our Machine Learning model, creating a better cycle of efficiency and improved customer service.

Loan Applications Log

The analysis is based on the **BPI Challenge 2017 event log**, provided by a large Dutch financial institution, which contains 2 different event logs. Anyway, we decided to work only on the **Application log** in order to enhance that process. This real-life dataset records the execution of a loan application process during the year 2016, with case handling extending up to February 2017. The log contains 31,509 loan applications and comprises approximately 1.2 million events executed by 149 different resources.

Event Classifications

A distinctive feature of this log is its multi-object structure. As highlighted in the academic analysis, the events are categorized into three distinct prefixes, representing different perspectives of the process:

- **Application States (A-events):** These events denote the milestone states of the loan application itself (e.g., A_Submitted, A_Concept, A_Accepted, A_Validating, A_Pending). They represent the lifecycle of the case from submission to the final decision.
- **Offer States (O-events):** These events track the lifecycle of the specific credit offers created for the customer (e.g., O_Create Offer, O_Sent, O_Returned, O_Accepted). A single application may contain multiple offers if the customer negotiates terms, but only one can be ultimately selected.
- **Workflow Activities (W-events):** These represent the actual work items performed by bank resources (e.g., W_Handle leads, W_Validate application, W_Call incomplete files). These activities capture the operational effort and duration, including start, suspend, and complete timestamps.

Case Attributes

Each case (Application) is enriched with domain-specific attributes that are crucial for both process discovery and the machine learning prioritization model. Key attributes include:

- **Requested Amount:** The value of the loan requested by the applicant.
- **Loan Goal:** The purpose of the loan (e.g., Home improvement, Car, Existing loan takeover).
- **Application Type:** Indicates if it is a new credit or a limit raise.
- **Credit Score:** A numeric assessment of the applicant's creditworthiness (available at the offer level).

Process Context

Following the "AS-IS" process identification described in the academic winner's paper, the log captures a process divided into three logical phases: Receiving Application, Negotiating Offers, and Validating Documents. The process concludes with three possible final states: A_Pending (successful loan), A_Denied (rejected by the bank), or A_Cancelled (withdrawn by the customer or timed out). This structure serves as the foundation for the BPMN model and the predictive analysis presented in later sections.

Data Processing

Our analysis directly builds upon the foundational work established by the academic winners of the BPI Challenge 2017. Taking their rigorous approach as a starting point allowed us to skip initial discovery steps and immediately focus on improvement strategies.
When importing the dataset into Apromore, we configured a specific column mapping: *case:concept:name* was mapped as Case ID, *concept:name* as Activity, *org:resource* as Resource, and *time:timestamp* as Timestamp, while *Column1* was set to ignore attribute. All other columns retained their default mappings, ensuring the dataset was properly structured for process mining analysis.

Log Filtering and Case Definition

Following the methodology described by the academics, we first processed the raw event log to isolate complete cases. The original dataset contains “running” cases that had not reached a definitive conclusion by the end of the observation period. To ensure the reliability of our goal model and machine learning predictions, we applied a strict filter to retain only cases containing one of the following events:

- A_Pending (Loan approved and finalized)
- A_Cancelled (Application withdrawn by customer)
- A_Denied (Application rejected by the bank)

This filtering procedure removed 100 of the original 31,509 cases, resulting in a cleaner dataset that accurately reflects closed-loop process behaviors. This “cleaned” log serves as the training set of our predictive models, ensuring that our algorithm learns from definitive outcomes rather than incomplete intermediate states.

Activity Clustering

To simplify the complexity of the raw log and then produce a clean and useful BPMN model, we also adopted the activity clustering strategy proposed in the academic reference paper (Figure 1). Activities occurring in rapid succession (less than 2 minutes apart) and performed by the same resource were grouped into logical macro-activities. This abstraction is crucial for our Goal Model, as it allows us to map automation candidates to meaningful business tasks (e.g., Validating Application) rather than fragmented system logs.

Clustered activity	Activities of the cluster
A_Create Application	{A_Create Application; A_Submitted; A_Concept}
W_Complete application	{W_Complete application; A_Accepted; O_Create offer; O_Created; O_Sent; W_Call after offers; A_Complete}
W_Call incomplete files	{W_Call incomplete files; A_Incomplete}
W_Validate application	{W_Validate application; A_Validating; O_Returned}
A_Denied	{A_Denied; O_Refused}
A_Cancelled	{A_Cancelled; O_Cancelled}
A_Pending	{O_Accepted; A_Pending}

Figure 1: Clusters chosen in the academic paper

Starting BPMN model

Building upon the methodological framework established by the academic winners, we extracted the baseline BPMN model of the process using their activity clustering strategy. Following their approach, we grouped activities occurring in rapid succession, specifically those executed less than 2 minutes apart by the same resource, into logical macro-activities or clusters.

Following their procedure, we selectively retained certain milestone activities (O_Create Offer, A_Complete, W_Handle leads, and A_Concept) as standalone elements rather than merging them into clusters. This preserves critique decision points while ensuring that the model's fitness and precision metrics remain unaffected.
The resulting clustered BPMN model is shown in Figure 2:

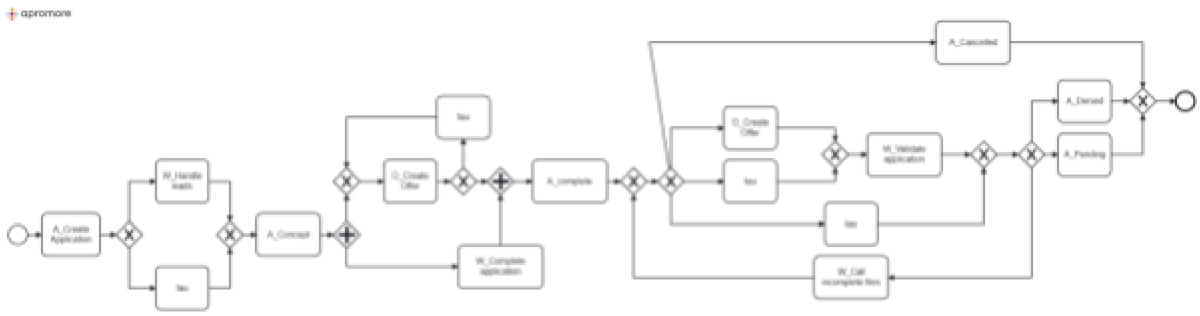

Figure 2: BPMN model discovered in the academical paper

In the next sections, we will explain in detail the instruments used to further optimise this already good process.

Machine Learning for Loan Prioritization

The core motivation behind our machine learning initiative is financial and operational efficiency. In the loan application process, a significant amount of time and resources is spent on applications that have been technically "accepted" by the bank (contain the activity A_Accepted) but are ultimately not finalized. We computed that out of the 23,014 completed cases accepted by the bank, only 53.7% are finalized.
The professional winners recognized similar predictive opportunities and developed models achieving 89.8% accuracy (Random Forest) for predicting the final outcome A_Pending based on offer attributes. However, their analysis was retrospective, explaining what factors correlate with success, rather than prescriptive for resource allocation during process execution.
We extend this predictive analytics foundation by developing a post-acceptance prioritization model that operates at a strategic decision point: immediately after A_Accepted. Our model predicts the probability of the case attribute "Selected = True" using process behavior features (event count, resource diversity, temporal dynamics) rather than static offer attributes, enabling:

- **High-probability cases** to proceed immediately through the standard workflow for fast service
- **Low-probability cases** to be diverted to manual review (*Human Check*), where resources decide whether to continue or terminate early, preventing wasted validation effort

Methodology and Feature Extraction

Following the professional's approach, we chose a Random Forest algorithm for this classification task due to its robustness and ability to handle complex interactions between variables, while also enabling feature importance identification to understand which process attributes most influence application outcomes. A critical aspect of our approach was ensuring that the model relies only on information available up to the moment of acceptance. This "early prediction" capability is what makes the model actionable in a real-world scenario.

Preprocessing Pipeline

The raw application log underwent three distinct transformation steps:

1. Chronological Alignment: First, we sorted all events by Case Identifier and Timestamp. This step was essential to reconstruct the correct sequential flow of each application, ensuring that the temporal dynamics (e.g., duration between steps) were calculated accurately.
2. Lifecycle Filtering: To reduce noise, we filtered the dataset to retain only events with a complete lifecycle transition. Furthermore, we restricted the scope to valid application activities by keeping only those pertaining to the Application (A), Offer (O), and Workflow (W) phases, discarding irrelevant system logs.
3. Post-Acceptance Focus: Since our objective is post-acceptance prioritization, we filtered the dataset to include only loan applications that had successfully reached the “Accepted” state. We also removed cases with ambiguous final outcomes (where the target variable Selected was missing or inconsistent) to ensure a clean binary classification target.

Feature Engineering Methodology

Standard machine learning algorithms like Random Forest cannot directly interpret raw event sequences. Therefore, we performed feature aggregation to compress the history of each case, up to the point of acceptance, into a fixed set of predictive features.

In contrast to the professional winners’ approach, which used offer-level attributes (credit score, monthly cost, offered amount), we engineered process execution features from event log behavior. This choice enables earlier prediction (at A_Accepted, before final offer parameters) and captures implicit quality signals embedded in how the application has been processed, like its velocity, complexity, and resource utilization patterns.

These features capture three distinct dimensions of the process:

1) Process Volumetrics

These features measure the intensity and operational complexity of the work performed on an application:

- Event Count: The total number of recorded events for the case, indicating the overall volume of activity.
- Resource Diversity: The count of unique, distinct resources involved. A higher number serves as a proxy for “hand-off complexity,” suggesting a case that required input from multiple departments or specialists.
- Workflow Count: The specific count of workflow-related events (starting with ‘W_’), quantifying the amount of back-office processing required.

2) Temporal Dynamics

Time-based features were engineered to capture the speed and rhythm of interactions, which are often strong predictors of customer interest:

- Total Duration: The elapsed time (in hours) between the first event and the last recorded event up to acceptance.
- Processing Pace: The average time interval (in hours) between consecutive events, effectively measuring the “heartbeat” of the process.
- Business vs. Non-Business Activity: We calculated the number of events occurring during standard business hours (09:00–17:00) versus those occurring outside these hours or on weekends. This captures unusual processing patterns or urgent handling.

3) Case Attributes and Encoding

Finally, we extracted static attributes related to the loan request itself:

- Financial Context: The Requested Amount was retained as a numerical feature.
- Categorical Encoding: Attributes such as Loan Goal (e.g., car, home improvement) and Application Type were processed using One-Hot Encoding (dummy variables) to transform these categorical fields into a machine-readable binary format.

Feature Importance Analysis

The analysis of feature importance revealed which factors most strongly influence the likelihood of a loan being finalized. The model identified the following as the most critical drivers (see Figure 3 for more details):

1. Number of events: The total volume of activity recorded for a case.
2. Total duration (hours): How long the process has been running.
3. Number of resources: The count of unique employees involved.
4. Average time between events: The pace of the workflow.

Interestingly, these process-derived features (intensity and timing) proved to be highly predictive, alongside the requested loan amount. This suggests that the way an application is handled (e.g., speed, number of touchpoints) is just as important as the content of the application itself in determining its success.

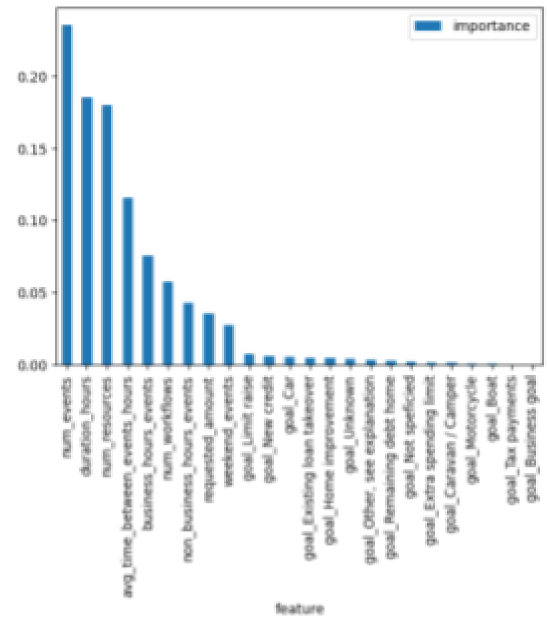

Figure 3: Random Forest features importance

Model Performance

The results of the RF model on the test set were highly encouraging, demonstrating the model's reliability:

- **Accuracy**: 0.91. This means that in 91% of cases, the model correctly predicted whether the client would finalize the loan or not.
- **Precision**: 0.92. When the model predicts a case is "high priority" (likely to be finalized), it is correct 92% of the time. This is crucial for avoiding false alarms that would waste resource time.

- **Recall**: 0.92. The model successfully identifies 92% of all actual positive cases, meaning it misses very few high-value opportunities.
- **F1 Score**: 0.92. This harmonic mean confirms that the model maintains a strong balance between precision and recall, without sacrificing one for the other.

These metrics confirm that the model is a viable tool for operational decision-making, capable of distinguishing high-potential loans with a high degree of confidence.

Automations

Selection Methodology

To identify repetitive activities suitable for automation within the BPI Challenge 2017 loan application process, we developed a multi-dimensional analytical framework. Rather than relying solely on frequency, we employed a weighted scoring system that combines quantitative process metrics with qualitative domain expertise.

Step 1: Filtering and Scope Definition

The analysis began by filtering out activities that were already automated in the “AS-IS” process (specifically W_Handle leads, A_Submitted, and A_Concept) and excluding events performed by the known automated resource User 1, identified by academic winners. To ensure operational relevance, we restricted the analysis to high-volume activities, defined as those with more than 1,000 occurrences in the dataset.

Step 2: Framework Dimension Analysis

We evaluated the remaining 21 candidate activities using three complementary dimensions, weighted according to their impact on automation success using the SMART direct rating method [1]:

1. **Frequency Analysis (30% weight):** We prioritized high-volume activities (e.g., W_Validate application with >208,000 occurrences) as they offer the highest Return on Investment (ROI)[1] potential.
2. **Temporal Variance Analysis (20% weight):** Automation requires predictability. We measured the stability of execution times using the Relative IQR (IQR/Median). Activities with low Relative IQR (e.g., A_Complete, O_Created) exhibit consistent, machine-like execution patterns suitable for automation. Conversely, activities with high variance imply complex, case-dependent handling that is difficult to standardize.
3. **Resource Concentration (15% weight):** We calculated Gini coefficients to identify tasks performed by a small, specialized group of employees. High concentration often signals the existence of strict, standardized procedures, making these activities easier to define and translate into automated rules compared to tasks distributed generically across the entire workforce.
4. **Domain Knowledge (35% weight):** This critical qualitative dimension incorporated findings from previous BPI winners (e.g., bottlenecks, importance in the process) and regulatory considerations. It acted as a filter to distinguish between tasks that can be automated technically and those that should be automated strategically (e.g., preserving human touch in sensitive customer interactions).

The final score is calculated as:

$$\text{Feasibility\_Score} = (\text{Resource\_Score} \times 0.15) + (\text{Frequency\_Score} \times 0.30)$$

---

[1] *The ROI in this context denotes the value gained from process mining interventions, measured primarily through throughput time reduction and process efficiency enhancement, as opposed to traditional financial Return Of Investment metric.*

+ (Predictability_Score × 0.20) + (Domain_Knowledge_Score × 0.35)

Where:

- Resource_Score = Gini_Coefficient × 10
- Frequency_Score = (Activity_Frequency / Max_Frequency) × 10
- Predictability_Score = ((Max_Relative_IQR - Relative_IQR) / (Max_Relative_IQR - Min_Relative_IQR)) × 10
- Domain knowledge is between 0 and 10

Step 3: Categorization

Based on their composite scores, activities were classified into four strategic categories (see Figure 4 for more information):

- **High Automation Potential**: comprises 7 activities accounting for the majority of high-frequency events, including communication triggers (O_Sent mail and online, O_Sent online only), offer management activities (O_Create Offer, O_Created), document validation (W_Validate application), status tracking (A_Incomplete), and fraud assessment (W_Assess potential fraud).
- **Requires Human Action**: includes 2 activities inherently requiring human interaction: W_Call after offers and W_Call incomplete files.
- **Requires Human Judgment**: contains 2 high-stakes decision points (A_Denied, A_Cancelled) where human expertise and discretion are essential.
- **To Be Evaluated**: encompasses 10 activities requiring further regulatory or technical feasibility studies, including various acceptance/approval activities (A_Accepted, A_Pending, A_Approved, A_Declined, A_Registered), workflow completion (W_Complete application), and offer outcomes (O_Returned, O_Accepted, O_Cancelled, O_Declined).

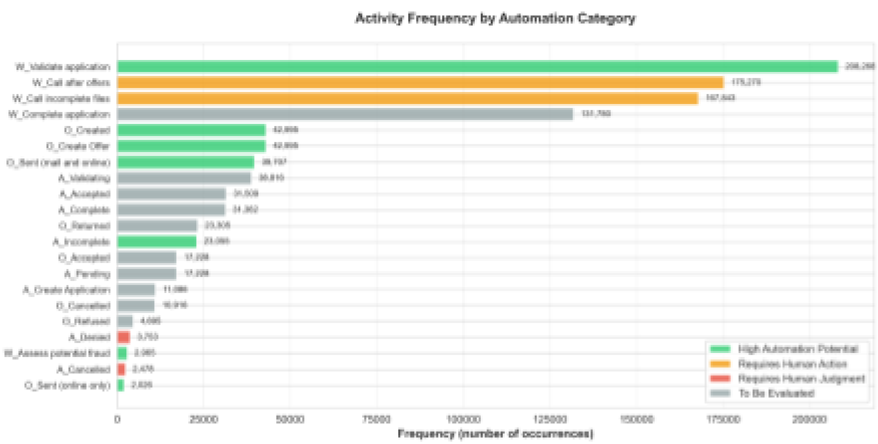

Figure 4: Frequency of the different activities (different colors indicate different categories)

Figure 4 shows that some high-frequency activities are not suitable for automation due to their nature (e.g., call for incomplete files or call after offer). Conversely, some automation-suitable activities are expected to be highly impactful due to their high frequency, while others exhibit medium to low frequency and therefore lower expected impact. Additionally, certain activities may reduce process times but require further investigation regarding practical implementation feasibility.

Assumptions and Constraints

Our estimation of automation potential and ROI relies on several key statistical and operational assumptions:

Statistical Metrics

To assess automation feasibility, we measured execution time predictability using Relative IQR (Interquartile Range divided by Median). This metric quantifies temporal consistency while remaining robust to the extreme outliers characteristic of real-world process data. The quartile-based Relative IQR provided robust discrimination between stable and unstable activities.

Operational Parameters

Automation Efficiency (90%): We assumed that for “High Potential” activities, automation is not 100% autonomous. We modeled a 90% efficiency rate, meaning automation handles the vast majority of the workload, but 10% of the time is reserved for human exception handling and oversight.

Start Timestamp Estimation

The original event log lacks the start time for all activities. To accurately measure execution duration, we employed the “Estimate Start Timestamp” plugin in Apromore. This technique infers the likely start time of an event based on the completion time of the preceding event and the resource’s workload, providing a more realistic basis for our ROI calculations than raw timestamps alone.

Regulatory Considerations (EU AI Act):

Our automation strategy adheres to the EU AI Act’s risk-based framework, which directly impacts implementation approaches. Creditworthiness assessment activities (specifically O_Create Offer) are classified as high-risk, requiring Human-in-the-Loop (HITL) oversight and explainability mechanisms. We address this through a hybrid model: rule-based offer generation handles 90% of routine cases automatically, while human reviewers approve edge cases and complex scenarios, ensuring compliance while capturing efficiency gains. For instance, W_Assess potential fraud qualifies for more permissive treatment as a decision-support tool rather than an autonomous decision-maker. By automating initial fraud screening with specialist escalation for high-risk flags, we reduce dependence on scarce specialist resources while maintaining full audit trails and transparency. This compliance-forward approach not only mitigates regulatory risk but positions the bank as a responsible AI adopter, building customer trust through transparent, explainable lending processes that satisfy regulatory expectations while delivering operational benefits.

Automation suggestions

We select the following activities based on the 5 highest feasibility coefficients.
To operationalize the improvement opportunities, we propose a phased automation framework that prioritizes document digitization, rule-based decisioning, and exception handling. The framework addresses the previously identified activities while accounting for cases that require human judgment and intervention.

- <u>W_validate application</u> (Score: 9.4/10): Currently, 48% of applications require incomplete file calls after the validation of the application, consuming significant employee time. By implementing Optical Character Recognition (OCR) techniques and business rules for document validation, we can save time and go directly to the problem source by performing the call, reducing time, resources, and improving the customer experience.
- <u>W_Complete application</u> (Score: 6.3/10): This represents a hybrid opportunity that requires an additional evaluation. We suggest starting with a feasibility assessment: split into (i) data completion tasks vs (ii) expert judgment tasks, then automate only (i) via guided workflows and pre-filled suggestions.
- <u>A_Incomplete</u> (Score: 6.0/10): Auto-detect incompleteness from validation outcomes and trigger an automated “missing info” workflow.
- <u>O_Sent</u> (Score: 5.8/10): Fully automate communication triggers (event-based notifications when status changes, standardized templates, delivery tracking).

- <u>O_Created</u> (Score: 5.5/10): Prefer rule-based offer construction (eligibility + pricing/limits) via decision tables. We suggest escalating edge cases to manual approval.

A notable observation is that activities occurring after W_Validate application exhibit lower average scores, even among cases that initially scored highest at the validation stage. This score degradation could compromise the automation's effectiveness and warrants careful monitoring during an eventual production deployment to ensure prediction reliability remains acceptable.

Final BPMN model and simulation

Based on our process mining analysis, we developed the following BPMN model (Figure 5) that represents the modified process according to our improvement suggestions:

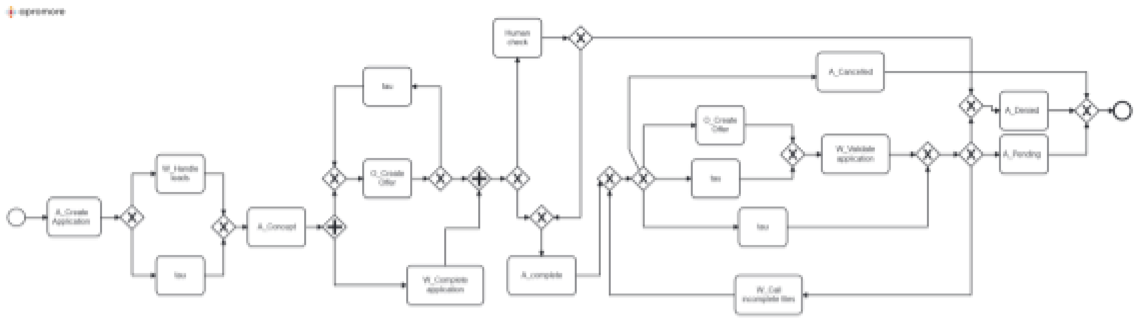

Figure 5: The BPMN model for the optimized process

The machine learning prediction is included in the process as an XOR block before the A_complete activity, and a new *Human Check* activity is created, where negatively classified applications are checked by hand. The outcome of the human check is again represented by a XOR split: an application can be reintroduced in the standard cycle with some probability, or can be directly denied and stopped.

In order to run the simulations, we had to decide the parameters of the new activities, such as how to allocate resources, the time distributions, and the probability of splitting in the XOR.

We set the first new XOR probability (machine learning tagging) as 50%-50%, since the target in our dataset is almost balanced: approximately 50% of the applications are selected, and 50% are not.

We set the XOR probability of the human check 92%-8%: with probability 92%, a case is directly denied, and with probability 8,% it is sent back to the standard cycle. These numbers come from the performance of our ML model: the model has both precision and recall of ~92%. This means that the probability of having a False Negative, knowing that a case is classified as negative, is the following:

$$P(FN \mid Neg) \ = \ Prec \ * \ ((1 - Recall)/Recall) \ = \ 8\%$$

While our modified model introduces additional structural complexity compared to the baseline AS-IS model, this trade-off is intentional and justified by our optimization objectives. Our priority is to deliver actionable process improvements through automation and predictive prioritization, which necessitate incorporating new decision points (ML classification, human review) and process flows.

Simulation Parameters and Time Distribution Setup

We defined the estimated time distributions for each activity in the simulation, explicitly accounting for the impact of the proposed automations. For all non-automated activities, we retained the original distributions and parameters as estimated from the simulation of the clustered BPMN model derived from the event log, ensuring consistency with the AS-IS scenario.

Modeling Automation Impact

For automated activities, we assumed a 90% reduction in execution time for simplicity. Our automation analysis was conducted on granular activities, while the simulation model utilizes clustered activities we previously described, aggregating activities with different durations. To address this problem, we applied a proportional adjustment logic: we calculated the new duration for clustered activities by subtracting 90% of the specific automated sub-activity's original duration from the cluster's total average time. Example: If a clustered activity has an average duration of 30 hours, and the specific sub-activity targeted for automation originally contributed 10 hours on average, the new simulated duration is calculated as: (30-10*0.9)h = 21h.
This approach ensures a fair and conservative estimation of time savings, reflecting that only specific parts of a cluster are optimized.

Human Check and Resource Configurations

The newly introduced *Human Check* activity was modeled using a Normal distribution (μ=4h, σ=0.3h). This parametrization was chosen to realistically represent the cognitive complexity and variability inherent in manually reviewing potentially rejected applications.
We consider that every resource has a working schedule of 9-17 (Monday-Friday), and that the Tau activities are executed by a dummy resource which takes 0 time, so they do not affect the simulation time.
Regarding resource performance, we introduced a seniority differentiation factor. We applied a 20% efficiency penalty (handicap in Apromore) to 70% of the resource pool, simulating the presence of junior staff members who typically operate at a slower pace compared to senior experts.

Final comparison with the academic starting BPMN model

Overall Process Analysis

The comparison between the Academic and our model demonstrates a reduction in process lead times, confirming the capability of the ML implementation to identify unfavorable applications and discard them before continuing the process, and the effectiveness of the automations proposed.

- <u>Mean Case Duration</u> (Figure 6): The overall average duration decreased from 2.41 weeks to 1.59 weeks (-34%). This improvement is likely driven by the early rejection of unsuitable applications, preventing them from clogging the workflow, and by the automated activities.
- <u>Overall Resource Usage</u> (Figure 7): At the same time, the Resources workload is decreased, with a reduction of both Average Resource Duration and Average Resource Frequency.

Figure 6: Statistics comparison between the model we proposed and the academic one

Figure 7: Comparison of resource usage between the model we proposed and the academic one

The reduction in average case duration may come from the improvement of cases with extended durations, which can benefit the most from our process modifications. Moreover, we can clearly see that in both situations, the median time is also lower.

Resource Efficiency in Standard Cycle.

Here we are comparing the resource usage of applications of the academic model against applications of our model, which are kept in the standard cycle by the ML learning model.
We can clearly see that the Workload on resources significantly decreases, especially with regard to Average Resource Frequency, dropping from ~75 to ~37 (Figure 8).
This confirms the usefulness of the prediction of the ML model in making the work lighter for resources when talking about the standard cycle, i.e., the most important branch of the process.

Figure 8: Resource comparison between the academic model and our model (the latter only considering activities where Human Check doesn't happen)

Resource Efficiency and Case Duration in Valid Applications (Pending)

The introduction of ML filtering and automations has substantially optimized resource utilization for Pending applications (which are the "good" applications from the point of view of the business), reducing both the frequency of interventions and the total time dedicated to each case.

- Reduced Workload per Case (Figure 9): The Average Resource Duration (total time resources spend actively working on a case) decreased significantly for successful/processed applications (Pending), dropping from 1.56 days to 1.08 days (-27%). This indicates that automation effectively handles routine tasks, reducing the manual effort required per application.
- Lower Interaction Frequency: The Average Resource Frequency also declined (from ~25 to ~13 for Pending cases), suggesting a reduction in process "ping-pong" and rework for successful applications. Operators touch each case fewer times to achieve the same outcome, streamlining the workflow.
- Lower Average and Median Case Duration (Figure 10): Both Average and Median case duration significantly dropped for cases containing A_Pending. This means that automation and prioritisation successfully help speed up the process for favorable cases.

Figure 9: Resource comparison between our model and the academic one considering just the cases where the application gets accepted

Figure 10: Statistics comparison between our model and the academic one considering just cases where the application gets accepted

Comparison between cases with and without "Human check" in Our Model

Here, we also show a comparison between cases that are sent to the human check (cases predicted to result in offer rejection) and cases that follow the standard process (cases predicted to result in offer acceptance).

- Resources Analysis (Figure 11): Both the resource frequencies and time are slightly lower for the cases that pass under the Human Check. This is due to the fact that a lot of cases are directly denied (even though this takes quite a lot of time w.r.t. Other activities), and only a few of them are re-inserted in the standard cycle. The higher frequency and time in the "no human check" cases are instead due to the ping pong of offers in the process, and cannot be avoided by this mechanism. Anyway, we find that the process is well balanced and resources are not overcharged in one branch of the process nor the other.
- Case duration analysis (Figure 12): The Average and Median case durations for the "human check" cases are both lower than the standard cycle. This is expected because we want the human check to easily get rid of unfavorable applications.

Figure 11: Resource comparison between the cases where the activity Human Check happens and those where it doesn't happen

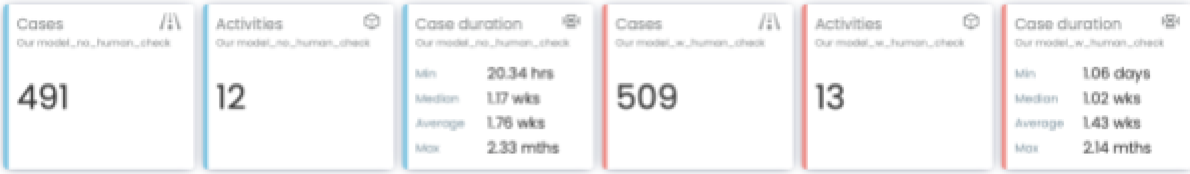

Figure 12: Statistics comparison of our model between cases where Human Check doesn't happen versus those where it happens

Discussion and conclusions

This project addressed the two main business questions by combining Machine Learning with process analysis and automation. Here is how our approach contributes to answer these questions:

1. BQ1: How can we prioritize the best applications?

We answered this by using a Machine Learning model. By predicting the outcome of an offer early in the process, we stopped treating every case the same way. Instead of letting likely-to-fail applications stay in the system for weeks, our model identifies and rejects them immediately. This solves the prioritization problem because it allows the business and resources to focus on the "good" applications (those likely to succeed), which no longer have to wait in line behind hopeless cases.

2. BQ2: How can we reduce human workload and speed up the process?

We addressed this through two mechanisms: process automation and intelligent prioritization. Automating high-frequency activities reduces manual workload directly, while ML-based early filtering shifts the workload distribution by preventing low-probability cases from entering the standard cycle. Resources no longer waste time processing applications destined for rejection, making the entire process faster and lighter. Efficiency in processing successful applications (A_Pending) is significantly higher in both time and resources.

Limitations and further improvements

For simulation purposes, we assumed 90% automation efficiency for selected activities, representing an optimistic upper-bound estimate. Future work should validate this assumption through sensitivity analysis using varying efficiency thresholds (e.g., 50%, 75%) to assess the robustness of expected benefits. Additionally, pilot testing would be essential to verify technical feasibility and real-world performance before full-scale deployment. This would also enable practitioners to calibrate the analysis framework weights according to their specific requirements.

Furthermore, our analysis assumes sufficient data quality and document digitization rates, particularly for high-impact activities like the W_Validate application, which requires structured, machine-readable inputs for effective automated validation. Additionally, the two automated activities O_Created and A_incomplete have little impact on the total execution time, so automating them does not significantly improve the average time.

Moreover, the newly introduced "Human check" activity requires further validation. We assumed this review step could be integrated without introducing delays or bias, but its practical implementation depends on domain-specific requirements and resource availability. Future work should collaborate with domain experts to design realistic staffing models, define review criteria, assess potential fairness implications, and ensure compliance with current regulations.

Lastly, the resource usage and case duration comparison between the two newly added branches (cases with and without Human check) shows that the number of resources can be tuned according to the business needs, in order to slow down the less useful branch of the process.

In conclusion, a cost-benefit analysis would strengthen these findings by quantifying the economic impact of proposed interventions. Scenario-based simulations incorporating implementation costs, resource savings, and payback periods would provide concrete ROI estimates to support investment decisions.

Bibliography

[1] W. Edwards and F. H. Barron, “SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement,” Organizational Behavior and Human Decision Processes, vol. 60, no. 3, pp. 306–325, Dec. 1994.

Individual Contributions

Ilaria Boschetto:

- Designed and implemented the BPMN process models in Apromore
- Contributed to the Data Processing section
- Prepared presentation slides and visualizations
- Reviewed the report

Federico Clerici:

- Proposed and developed the automation framework
- Developed the *Automation* section
- Contributed to the checking of the ML model
- Synthesized and compared insights from BPI Challenge 2017 winning solutions with our analytical findings
- Contributed to limitations and further works
- Reviewed the report

Raffaele D’Agostino:

- Defined the idea of the academic vs modified BPMN model and the business questions to answer
- Contributed to the checking of BPMN models correctness and simulations
- Introduction section of the report
- Contributed to the data description section
- Analyzed and discussed the results of the simulations
- Contributed to limitations and further works
- Reviewed the report

Gabriele Villa:

- Implemented the BPMN model on Apromore
- Tested the BPMN model performance and compared it with the original model
- Ran the simulations with the optimal parameters
- Contributed to the following sections of the report: *Main Contributions, Loan Applications Log,* and *Data Processing*
- Curated the layout of the report and reviewed it

Davide Volpi:

- Proposed and developed the ML framework
- Developed the *Machine Learning for Loan Prioritization* section
- Contribute to the development and checking of the BPMN models and simulations
- Reviewed the report