**BPI2017 Challenge: Business process mining–A Loan process application Introduction- roei michael , avichay hayne and asaf koren**

Process mining, the practice of extracting knowledge from event logs recorded by systems, plays a pivotal role in comprehending and optimizing complex operational processes across various fields, including banking. This project, in particular, focuses on the application of process mining to the BPI2017 Challenge dataset. This dataset, made available by the BPI Challenge 2017, comprises various application traces within a banking environment. The project aims to glean insights, identify patterns, and propose enhancements that could potentially revolutionize the operational workflows of financial institutions.

The BPI2017 dataset originates from the system logs of a financial institution. Through a meticulous examination of this data, we aim to map/discover all process flows and investigate any possible inefficiencies. By focusing on the frequency of events, we anticipate identifying points of improvements, thus, enhancing the overall efficiency and effectiveness of banking operations. The project also entails searching for behavior patterns that might enable the institution to perform more in-depth analysis, suggesting changes, improvements, corrections, and learning from its processes.

In the contemporary economic environment, financial institutes face stiff competition, especially from emerging Financial Technology (FinTech) firms. One strategy to withstand this competition is by enhancing customer experience through the application of digitalization and automation techniques that streamline loan processes. Consequently, this project's goal is not only to analyze the loan application process, but also to focus on improving the customer experience and potentially increase revenue.

By addressing these aspects, our project seeks to answer critical questions raised by the process owners. These include queries related to throughput times for various parts of the process, the influence of the frequency of incompleteness on the final outcome, the number of customers requesting more than one offer, and the comparison of conversion between applicants who receive a single offer and those who receive multiple offers.

By intertwining process mining with the banking sector, we hope to unlock profound insights that could help financial institutions improve their operational processes, customer experience, and ultimately, their bottom line. Through this project, we wish to demonstrate the untapped potential and wide applicability of process mining in contemporary financial practices.

**Preprocessing**

The preprocessing phase aimed to prepare the dataset for a smooth and effective process mining activity. It was performed in a stepwise manner, systematically reducing the size and complexity of the dataset, while ensuring to maintain and highlight the most critical and informative parts of the process.

Trace Frequency Filter:

Our first step was to filter out all traces that occurred only once in the log file. The assumption behind this action was that infrequent traces, specifically those that occurred only a single time, were less likely to be representative of the common process pathways in our application system and hence, could be removed without significant loss of information.

This can be expressed as follows:

Let T be a set of all traces in the log file and n(t) be the number of occurrences of a specific trace t in T. Then, the reduced set of traces T' after this preprocessing step is given by:

Event Redundancy Filter:

We noticed through a visual inspection of the event log that some events seemed to contribute little to no new information about the processes. These included duplicated events and others that were not essential to the process. We removed such events, simplifying the event logs while still maintaining the overall process information.

Creation of Offer Dataset:

In order to enrich our understanding of the processes and enhance our analysis, we created a separate dataset focused on the offers part of the applications. This included important information such as the loan goal, requested amount, offered amount, and credit score. This dataset, while being valuable on its own, also served as an additional dimension to our primary event log data, allowing us to derive deeper insights about the process.

Additional Filtration Methods:

Finally, we devised and implemented several other filtration methods that were more oriented towards improving the visual analysis and exploration of the dataset. These included:

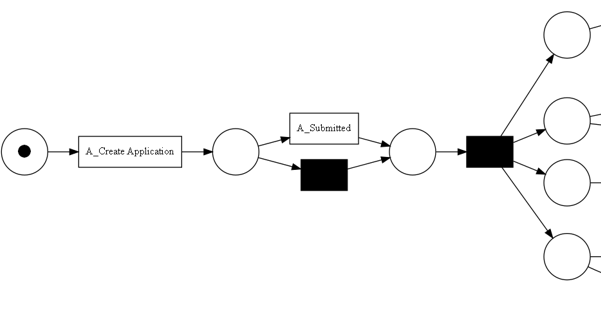
* **Top X Trace Filter**: This filter reduced the dataset to include only the top X most frequent traces, thereby focusing on the most common process paths.
* **Application Outcome Filter**: This filter selectively included only those traces that ended in either 'O\_accepted' or 'O\_rejected', i.e., the processes which resulted in a definitive outcome for the application.
* **Workflow Segment Filter**: This filter included only the events that occurred between 'W\_complete' and 'A\_accepted', thereby focusing on the main workflow of the application process.

These filters enabled us to investigate specific aspects of the dataset, focusing on the areas of most interest and thereby enhancing our overall analysis.

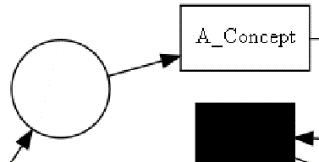
Overall, the preprocessing stage involved careful and systematic curation of the dataset, rendering it more manageable and comprehensible, while still capturing the essential dynamics of the process flows.

**Process Discovery Results**

For the process discovery we have used a plethora of algorithms such as alpha miner, inductive miner, heuristics miner and ILP miner. After our filtration and preprocessing for the data and using the miners on the data while using different filters each time we have reached the conclusion that the inductive miner was the best possible choice for our choice of filters and the best fit for our data. in this section we will dissect the Petri net that we had received and analyze it to our full capacity.



As we can see in this part of the Petri net each application has to go through the process of actually creating that application and after that we can see that some of the applications are going through submission and some don't the reason to that is that some of the applications are submitted through the website instead of an actual banker and because of that there will be a waiting time for the process to actually be handled by a real banker instead of it being instantaneously when doing it the conventional way. After the submission process there are four conditions that need to be met. and in the Petri net we can see the Petri net separates into four branches, the first branch is a transition that is called a\_concept, that is a first assessment which is executed automatically the moment an application finishes submission.



**Results**

**Preprocessing Results**

Discuss the results of your preprocessing. Did it achieve the goals you set? Were there any unexpected results or challenges?

The preprocessing stage yielded a more concise, yet comprehensive dataset suitable for effective process mining.

**Process Discovery Results**

Describe the visual models you generated and any significant insights they provided. Include diagrams where possible.

The process discovery phase led to the generation of insightful Petri nets and other models that effectively illustrate the process pathways.

**Analysis Results**

Discuss the correlations, repeating patterns, and potential areas for improvement you identified.

Our analysis unearthed notable correlations and repeating patterns within the process, hinting towards potential areas for process optimization.

**Discussion**

This is where you interpret your results. Discuss what they mean in a broader context and how they could be applied to improve banking services.

Our findings provide valuable insights into the application process in a banking context. The recurring patterns and correlations identified can be instrumental in refining the process for increased efficiency.

**Conclusion**

Summarize the project, its findings, and their implications. This is also a good place to suggest future work or next steps based on your project.

Through effective process mining of the BPI2017 Challenge dataset, we were able to derive actionable insights to enhance the banking application process.

**References**

List all the resources you used or referred to in your project.