BPIC 2017: Density Analysis of the Interaction With Clients

Elizaveta Povalyaeva[1], Ismail Khamitov[2], and Artyom Fomenko[3]

National Research University Higher School of Economics,
20 Myasnitskaya St., Moscow 101000, Russia
Faculty of Computer Science
[1]empovalyaeva@edu.hse.ru
[2]inkhamitov@edu.hse@hse.ru
[3]aofomenko@edu.hse.ru

Abstract. A clear vision and understanding of the embedded business processes is the basis for finding ways to re-think and optimize the company's operations. Process mining techniques make it possible to workout hidden relations by building real-life process models based on the logs of information systems. In this paper, we introduce our analysis of an event log with annual performance results for the process of credit approvals from Dutch Financial Institute. The intention for the analysis was to measure the effectiveness of system by analyzing its downtime and the density of interaction with clients. For in-depth analysis, we applied machine learning algorithms (Random Forest [6]) for finding hidden dependencies, pi-conformance analysis in Celonis [7], and built well-balanced models in ProM [9] and Disco [8]. The results of conducted analysis show that overall process is well-optimized, though some recommendations may be considered.

1 Introduction

Process mining as a combination of methods and techniques of business process optimization is a continuously developing sphere, the aim of which is to revise overtly or covertly functioning business processes through developing efficient patterns and receiving appropriate recommendations for the improvement of efficiency indicators of a company. CEO's, whose companies are on the stages of growth and maturity, need to clearly understand current intracorporate business processes and to find the ways to stabilize and improve them.

Before commencing the analysis of the process, indicated in the *BPI Challenge 2017* [1], we, as students, who had never had any experience with process mining, developed a plan of analysis for this process in accordance with Process Mining Manifesto [2].

1. **Deep dive into the data:** the goal of the first step was to obtain the clearest possible understanding of this process as well as developing and modeling probable general patterns of the process under consideration. According to Guiding Principle 1 [2], event log data were treated as First-Class Citizens:

it was necessary to understand the quality of received logs and understand the process without an opportunity to immerse into the field.
2. **Driven by Question:** then we started exploring the questions organizers had prepared. For each of them we had to determine relevant parts of the logs. The aim was to obtain a comprehensive answer to the question: the image of the process under consideration, the analysis of possible etiology, revealing potential consequences and, as a result, a list of concluding recommendations. We should note that we tried to obtain recommendations, which potentially would be useful for the company (Challenge 11 in [2]).

The list of questions, which we wanted to find answers to, is the following one:

1. **Analysis of the system's downtime.** What are the throughput times per part of the process, the difference between the time spent in the company's systems waiting for processing by a user and the time spent waiting on input from the applicant as this is currently unclear?
2. **Analysis of the consistence of cooperation with the client:**
   a) What is the influence on the frequency of incompleteness to the final outcome? The hypothesis here is that if applicants are confronted with more requests for completion, they are more likely to not accept the final offer.
   b) How many clients ask for more than one offer (where it matters if these offers are asked for in a single conversation or in multiple conversations)? How does the conversion compare between applicants for whom a single offer is made and applicants for whom multiple offers are made?

This document has the following structure. Chapter 2 contains a description of data and processes for analysis. Chapter 3 contains the analysis of the system's downtime. In Chapter 4 we described analysis of the consistence of cooperation with the client. In Chapter 5 we analyze offering scenarios. Chapter 6 concludes the paper with a list of recommendations for the company.

2 Overall understanding of the data and process.

A mandatory first step before conducting the analysis of business processes is an information collecting procedure about the company and its integrated processes. In general, this procedure includes: discussion with the client about business goals of the research and questions which need to be answered, obtaining process event logs (created by information companies upon functioning), and immersion into the research field by interviewing employees.

As part of the BPIC 2017 [1] researches were offered to conduct an analysis of the process of issuing credit offers in a financial institute. The data were two sets for the period from the beginning of 2016 to February 2, 2017:

– Application event log (AL) contains data describing the whole process under consideration from filling out a loan application to decision-making (approving or declining). It contains 1 202 267 unique event records and 31 509 unique cases.

There are events of three types:
A - Application state changes,
O - Offer state changes,
W - Workflow events.

– Offer event log (OL) is a part of the application event log which contains information about all offers to clients. It contains 193 846 unique event records and 42 995 unique cases (the financial institute had an opportunity to make more than one offer to the client).

There are events of just one type:
O - Offer state changes.

Based on the analysis of the logs, we developed a generalized structure of investigated process. We developed several models with the help of such instruments as Disco [8], ProM Framework [9] and Celonis [7]. The most suitable one was a BPMN 2.0 model which we developed with the help of Celonis [7] (Figure 1).

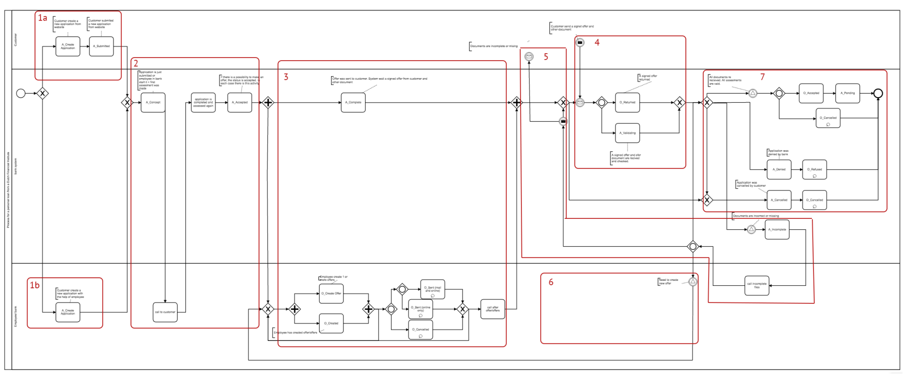

Fig. 1. BPMN 2.0 process model by Celonis

The following steps can be specified withis the process:

1. A new application is filed online through the web-site (Figure 1 1a), or by the bank's employee upon client's request (Figure 1 1b). Most applications are submitted through web-site (64 %).
2. The application undergoes a series o checks and corrections within the bank.
3. If the bank can make more than one offer to the client, employees compose one or several offers and send them to clients and notify them with a phone call.

4. The client chooses a credit offer and sends the necessary documents to the bank to confirm his application. These documents are checked and verified in the bank.
5. If the client hasn't sent all the documents or incorrect documents, the bank contacts the client and offers to resend the documents.
6. If the documents are unacceptable, the bank contacts the client and offers another credit option.
7. When the client has sent all necessary documents, the bank makes the decision (approves or declines the application).

To ensure correct understanding of the process we used PI-conformance (advanced conformance checking tool by Celonis based on Machine learning and AI). It determined the consistency of logs and the model. We managed to reach 99% of consistency (Figure 2).

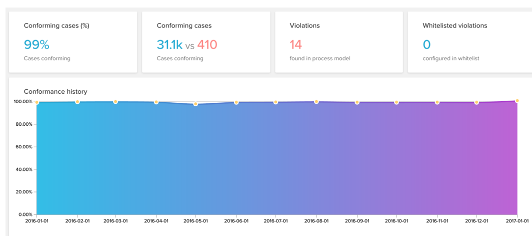

Fig. 2. Celonis PI-Conformance results

A description of event log activities is summarized in Tables 1, 2, and 3.

2.1 General description of process scenarios

All cases of the process under consideration can be divided into three categories:

1. **Successful completion:** process case includes "A_Pending" event;
2. **Denied by bank:** process case includes "A_Denied" event;
3. **Cancelled by client:** process case includes "A_Cancelled" event.

Analyze chart showing outcome probability against case length with the help of machine learning (Random Forest [6]). This chart is shown in Figure 3.

Considering the supposition that a company, which provides the logs, is interested in the issue of increasing the ratio of successful results, analyze the "Pending" line.

1. **Period 1st-8th day:** gradual increase of probability of success of an application and gradual decrease of probability of declined application. At this stage filtering uninterested clients is of importance. The longer the client communicates with the company, the higher the probability of his (her) need for a loan.

Table 1. Application state changes activities

Activity	Value	Occurrence
A_Create Application	Creating an application. The application can be created by the client on the website of by a bank's employee.	100%
A_Submitted	Submitting application. Applicable for online-applications.	64%
A_Concept	First automatic check. If the application was submitted online, an employee calls the client to complete the application.	100%
A_Accepted	The application has been checked. The employee may make a credit offer. The employee creates 1 or more offers. **Note:** probability is 100%, which means that all clients receive at least one offer. Is it reasonable to make credit offers to all potential clients or would the bank introduces new verification techniques? There are several opportunities to decline irrelevant clients at early stages.	100%
A_Complete	Credit offers have been sent to the client. The bank anticipates required documents from the client (paylist, ID, etc.).	99%
A_Validating	Documents have been received. The bank checks and verifies the documents.	69%
A_Incomplete	The checking and verification process was completed with an error. Either the documents are incorrect or absent. The employee notifies the client that the latter must resend the documents.	47%
A_Pending	All documents have been received, checked, and verified successfully. An offer with a client has been signed.	54%
A_Denied	Bank denies the application.	11%
A_Cancelled	Client declined the credit offer, didn't send the documents or is out of reach for 30 days.	33%

2. **Period 9th-28th day:** stable high probability of success of applications. Here we can conclude that the process within the company is efficient, and clients, despite the period of consideration of their applications, retain interest in the success of their application.
3. **Period 29th-34th day:** rapid drop of probability of success. To analyze the etiology, we used the Disco tool. We set a filter for AL log, which showed cases, where events follow with a period of over 29 days, which amounted to 8344 cases (26% of the total).

Figure 4 shows an example of such case.

Having analyzed these cases we drew the conclusion that after a month of communicating with the client the company closes the application as cancelled by client. There were no additional attempts to contact the client in the logs.

Table 2. Offer state changes

Activity	Description	Occurrence
O_Create offer	Creating a credit offer.	100%
O_Created	Credit offer created.	100%
O_Sent (online only)	Credit offer sent online.	5%
O_Sent(mail and online)	Credit offer sent online and by mail.	98%
O_Returned	Client submitted documents for a selected credit offer.	69%
O_Accepted	Credit offer confirmed. Client's application passed all checks and verifications. Client selected this credit offer.	54%
O_Cancelled	Credit offer cancelled by client.	49%
O_Refused	Credit offer cancelled by bank.	11%

Table 3. Workflow activities

Activity	Brief description	Probability of occurrence
W_Handle leads	If an application is submitted on the website, the first workitem that's created is 'handle leads'. Now the application is assessed for the first time (automatically). If the assessment is positive, a new workitem 'complete application' is created. If the assessment is negative, the application is 'declined'. If the assessment can't be completed because of technical problems, the workitem is still 'handle leads' and a new assessment can be done manually.)	11%
W_Complete application	An employee completes processing in incoming application from a client.	94%
W_Shortend completion	The client has a certain profile that defines as a lower creditrisk. These applications are investigated less thorough then higher risk applications.	below 1%, 74 cases
W_call_after offers	Calling clients after developing and sending offers.	99%
W_Validate application	Calling clients after developing and sending offers.	69%
W_Call incomplete files	Calling clients about resending documents for a selected offer.	52%
W_Personal loan collection	Applications for a Personal Loan.	below 1%, 2 cases
W_Access Potential fraud	Estimation of potential fraud.	below 1%, 301 cases

Also, there are 333 successful cases, where an employee sent a new credit offer to clients after a month without contact.

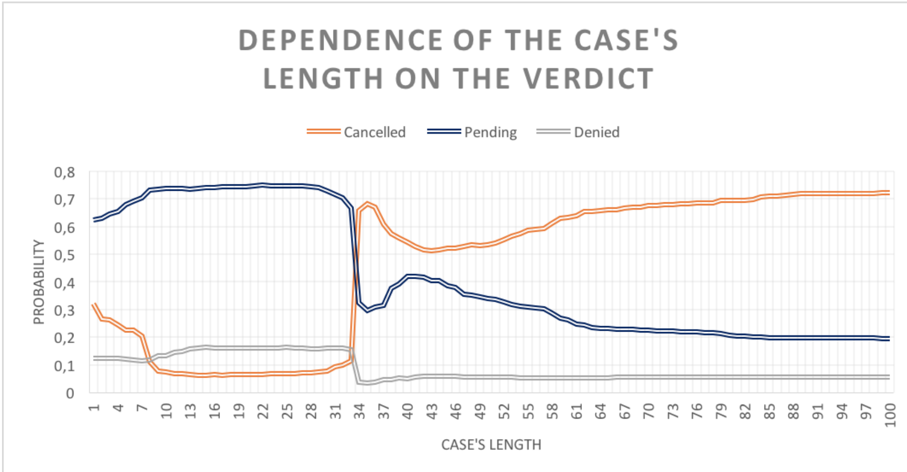

Fig. 3. Final probability decision by case length

	Activity	Resource	Date	Time	Duration
1	A_Create Application	User_1	01.01.2016	16:00:04	0 millis
2	A_Submitted	User_1	01.01.2016	16:00:04	0 millis
3	A_Concept	User_1	01.01.2016	16:01:44	0 millis
4	W_Complete application	User_19	02.01.2016	14:32:57	10 mins, 58 secs
5	A_Accepted	User_19	02.01.2016	14:40:45	0 millis
6	O_Create Offer	User_19	02.01.2016	14:42:23	0 millis
7	O_Created	User_19	02.01.2016	14:42:25	0 millis
8	O_Sent (mail and online)	User_19	02.01.2016	14:43:56	0 millis
9	W_Call after offers	User_19	02.01.2016	14:43:56	0 millis
10	A_Complete	User_19	02.01.2016	14:43:56	0 millis
11	A_Cancelled	User_1	02.02.2016	10:00:20	0 millis
12	O_Cancelled	User_1	02.02.2016	10:00:20	0 millis

Fig. 4. Case example with events follow with a period of over 29 days by Disco

Recommendation: a company should continue communicating with clients, who don't get in touch/don't send documents for a selected credit offer. The company should develop a strategy for additional notification of clients about their application and that the bank is anticipating their documents. We believe that it is possible to increase the probability of success (and the number of successful applications). The possible better situation is shown in Figure 5 as a blue line.

Our suggestion is *to make to additional calls*:

– in 15 days,
– in 30 days.

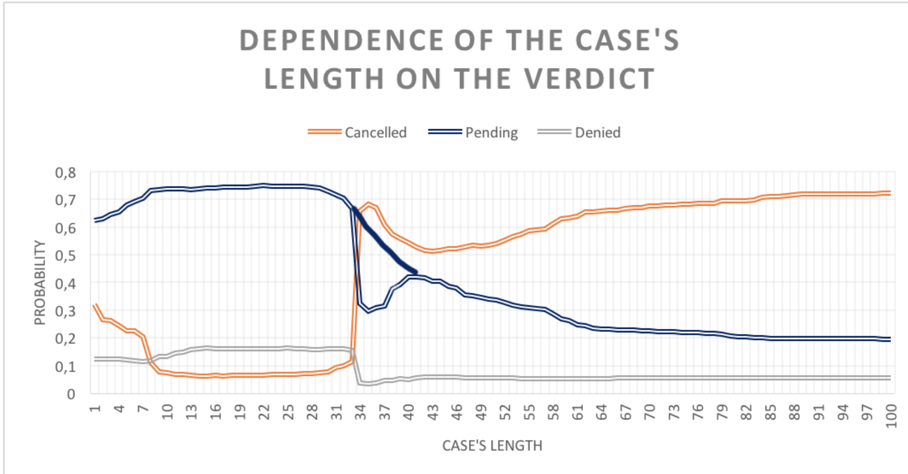

Fig. 5. Possible final probability decision by case length

2.2 Concept Drift Analysis

It was important to find out whether the process under consideration was stable. Here "stable" means the absence of cardinal structural changes in the process (reordering, adding, and deleting the events). For this purpose, we have used the Concept Drift plug-in for ProM (Figure 6). This plug-in calculates p-value based on the collection of case characteristics and application of statistical hypotheses [3].

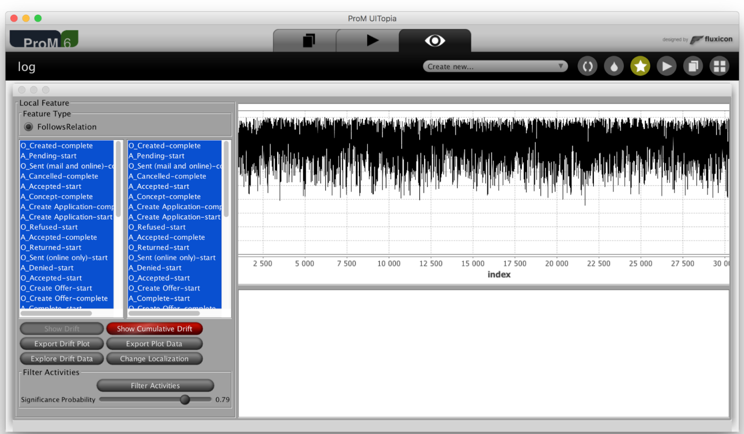

Fig. 6. Concept Drift plug-in's result for Application event log

Based on the obtained graph, we can conclude that the process was stable, without any structural changes.

We also considered the process logs in Disco and get a dependency diagram of the number of cases processed at moment in time (Figure 7).

Fig. 7. The Application event log in Disco

The log can be divided into two sub-logs:

– January – May, 2016;
– June, 2016 – February, 2017.

The reasons for this separation and the apparent of batch process in the second sub-log cannot be clarified without immersion in the company's domain. We can assume that the financial institute introduced some kind of mechanism to attract new clients in May-June, 2016.

Results: The analysis of changes showed that the process is stable and optimizable. More in-depth studies of the events dependence are required. Such studies were carried out in response to the questions posed by the financial institute.

3 Analyzing the intervals in system downtime

Question 1. *What are the throughput times per part of the process, in particular the difference between the time spent in the company's systems waiting for processing by a user and the time spent waiting on input from the applicant as this is currently unclear?*

3.1 Defining and examining the intervals in system downtime

The first step is to define the key elements of the process, which correspond to the time that the company spends in waiting of a client, and the time taken by internal processing of requests. These time expenditures can be interpreted as the "process lag by client's fault" and the "process lag by employees' fault".

The following intervals were taken into consideration:

1. Time intervals when clients wait for the financial institute response:
   a) between filling a request on site and acceptance of the request;
   b) between approval of a request and sending of an offer to client;

   c) time spent by employees on checking of a request and making a decision.
2. Time intervals when an institute waits for a client's response:
   a) between sending an offer to client and the beginning of checking the documents provided by client;
   b) between taking a decision "client should provide extra information (a request is not completed)" and the beginning of repeated checking the request after acquisition of missing documents from client.

3.2 Downtime by the institute's fault – a client is waiting

Let's analyze the time intervals that correspond to waiting times of institute's response by clients. First, we should underline an interval between events *A_Submitted* and *A_Accepted*. It can be interpreted as the time between filling a request on site and an acceptance of this request. It can be interpreted as the time between filling a request on site and accepatance of a request. It can be seen that between these events only *workflow* events and change of the request status (*A_Concept*) occur (Figure 8).

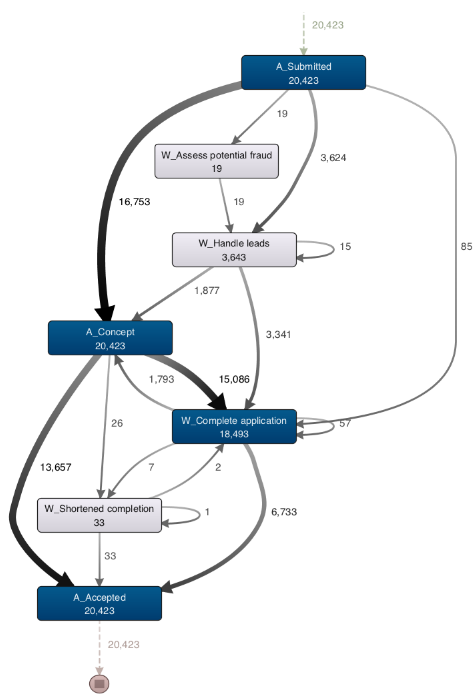

Fig. 8. The process scheme

For these intervals we built histograms that show the percentage of successfully completed requests, denied requests and requests, which were cancelled by client for different periods of waiting (Figure 9).

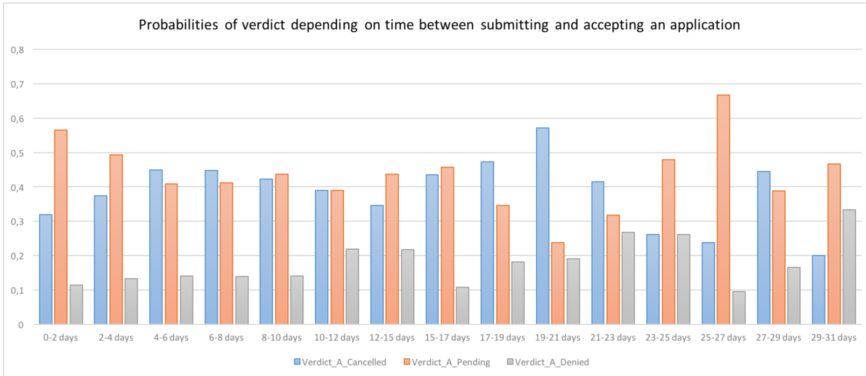

Fig. 9. Probabilities of verdict depending on time between submitting and accepting application

In this case, we should note that the number of requests by reason of denial or inactivity of a client does not have a tendency for increase with prolonged time for acceptance of request, which means the time should be spent on this in presence of reasons for more detailed examination of a request, as this will not lead to increased possibility of client outflow.

Analyze the interval between acceptance of a request and the first sending of an offer: $A\_Accepted \rightarrow O\_Sent$. This interval can be interpreted as a time of waiting by client of the first offer from institute (Figure 10).

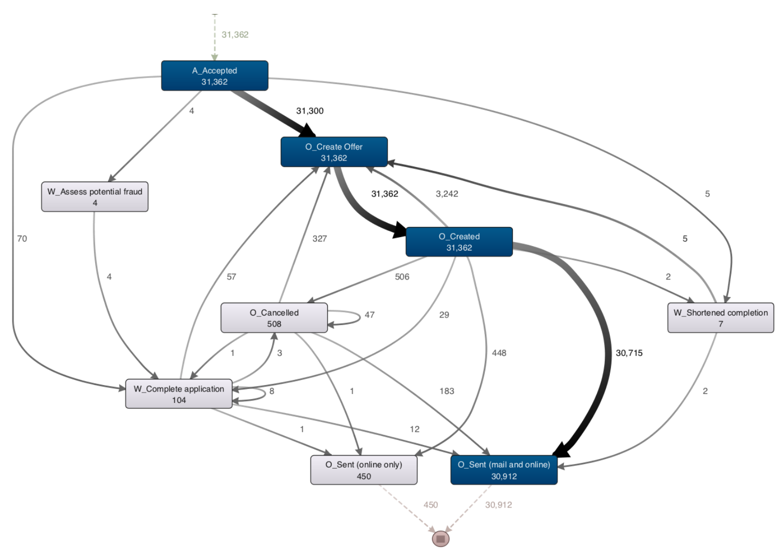

Fig. 10. Process from *A_Accepted* to *O_Sent*

Similar histograms can be build for this interval (Figure 11).

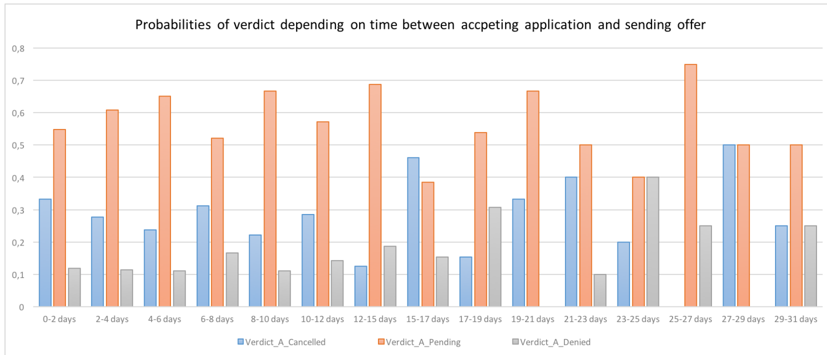

Fig. 11. Probabilities of verdict depending on time from accepting application to sending offer

In this case, we can see the growing possibility of client's cancellation when waiting time exceeds 2 weeks.

Moreover, an important period is the time spent by employees on checking of request and making a decision. These events correspond to intervals between events *A_Validating* $\rightarrow$ *W_Call_incomplete_files*, *A_Validating* $\rightarrow$ *A_Pending*, *A_Cancelled*, *A_Denied*. In such a case, the total time spent during all checks within single application should be accounted (Figure 12).

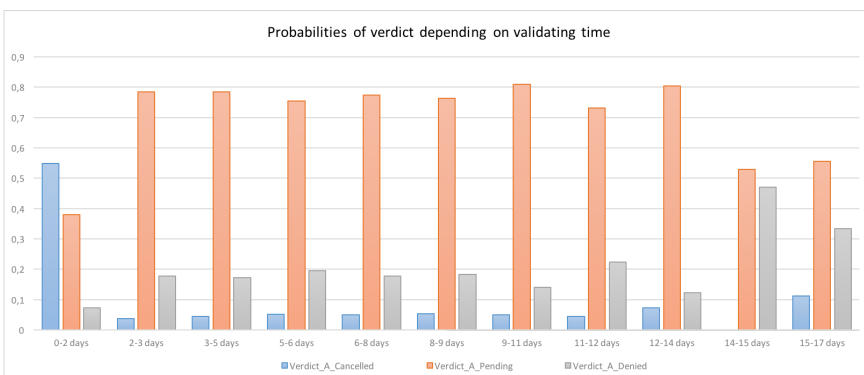

Fig. 12. Probabilities of verdict depending on validating time

These graphs show that the incrementation of the time spent by a financial institute for analyzing request progressively decrease the possibility of successful completion of a deal, while the possibility of client's cancellation increases.

Analysis of the log also shows that there are lots of cases, in which a request was marked "Cancelled" strictly a month after client's inactivity. This led us to a guess that financial institute considers a client having cancelled a request if he failed to respond within a month.

Besides, it should be also noted the possibility of request cancellation by institutes's initiative is most probable within the first 10 days of process (high index in the right part of the graph can be considered an outbreak).

3.3 Downtime by the client's fault – financial institute is waiting

In this sub-section, we will analyze time intervals when an institute is waiting for a client's response. The first one is the period between sending an offer to a client and the beginning of checking his response: *O_Sent* → *A_Validating*. For this interval, we draw a histogram, which is shown in Figure 13.

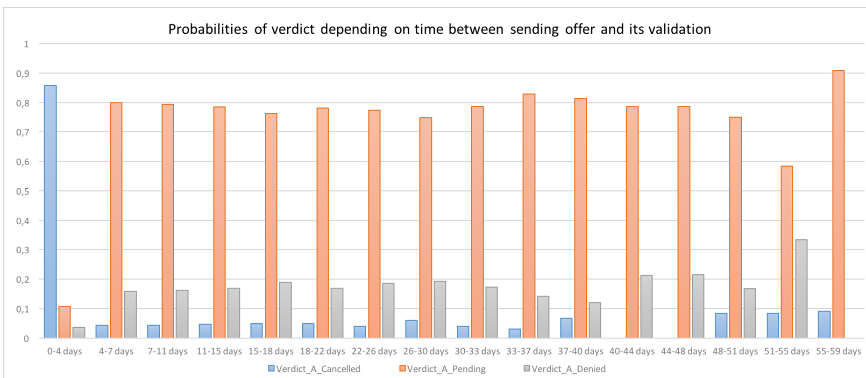

Fig. 13. Probabilities of verdict depending on time between sending offer to client and validating application

It should be mentioned that regardless of time of waiting of a client's response by institute, the possibility of successful completion remains high. However, 33% of all requests are marked as "Cancelled", and earlier we came to a guess that after a month of inactivity, the financial institute considers a client having denied the offer. In this case, we again come to a hypothesis that client should be notified of institute's interest in his decision and waiting period should be prolonged. Therefore, it is possible to transfer certain "absent-minded" people from "Cancelled" to "Pending" category.

Besides, the financial institute is waiting for a client's response after checking a request and finding any drawbacks in it – missing, incomplete or improperly completed documents, etc. Such periods correspond to events *A_Incomplete* → *A_Validating*, which start since a request was marked incomplete and finish at a repeated checking of a request. Analyze a histogram similar to the above (Figure 14).

In this case, we can note an absence of tendency for increase or decrease of possibility of successful or unsuccessful deal at prolonged time of waiting when

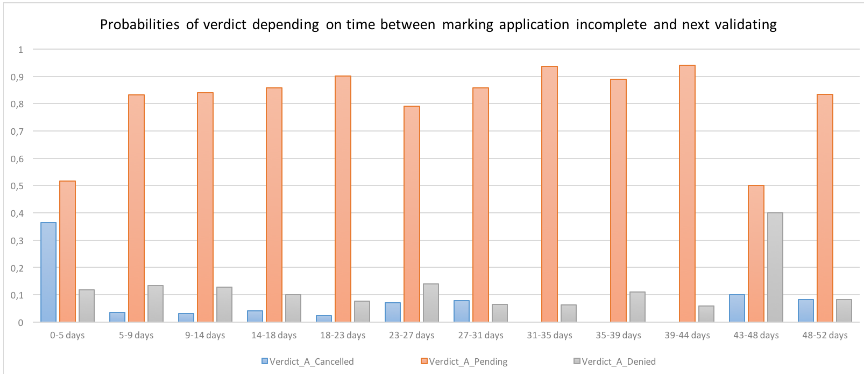

Fig. 14. Probabilities of verdict depending on time between marking application incomplete to next validating

a client provides the missing documents for a request, which can also serve as an argument in proposal not to consider clients who were inactive within a month as having denied, but remind them of a possibility of making a deal with financial institute.

Moreover, we analyzed the total time of waiting by a client and the institute during all the process of their interaction (Figure 15 and 16):

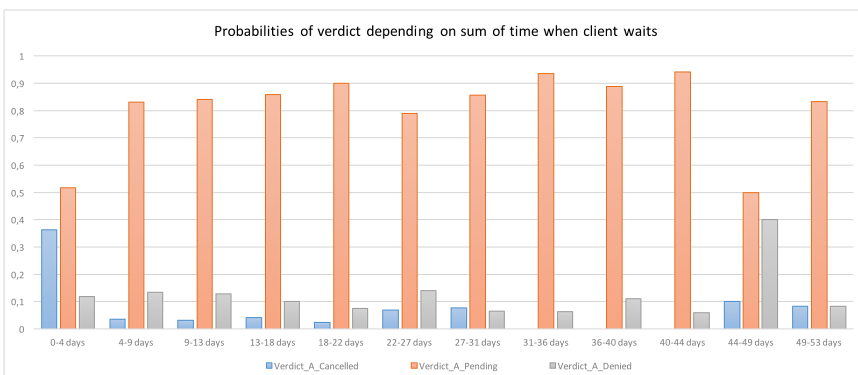

Fig. 15. Probabilities of verdict depending on time when client waits

The histogram, which shows the waiting time of a client (Figure 15), confirms the conclusions, which have been done when we investigated histograms of time spent by the institute for checking requests.

The histogram showing waiting time of the institute (Figure 16) leads to another interesting observation. A possibility of client's cancellation is most probable within the first two weeks, i.e. we can assume that clients who had been interacting with an institute for more than 2 weeks (and may be after the first

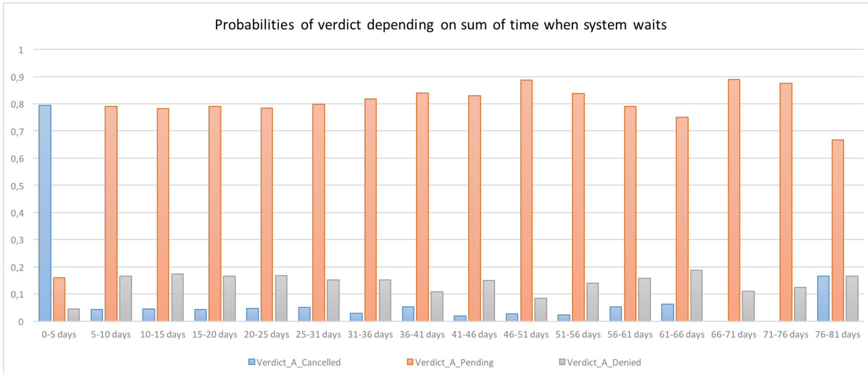

Fig. 16. Probabilities of verdict depending on time when system waits

week, as they become candidates for a successful client with higher probability) require attention, as they have potential for making a successful deal.

In the process of log analysis 147 similar requests/cases of loss of potentially successful deals were found: after processing of request by the institute the client offer is made up; then, the institute tries to send and offer to a client; however, the event of offer sending to a client does not come - an offer is marked as cancelled and processing of a request is suspended. These 147 cases present particular interest as the absence of additional attempts of sending or taking any measures. It can be assumed that for sending an offer to a client the financial institute lacks certain information on client (e.g., contacts). Process diagram for these cases is shown in Figure 17.

3.4 Recommendations

By analyzing the identified time intervals, we concluded that the institute should not consider a client having denied an offer a month later, but should proceed our interaction. The following conclusions also made:

- clients who continue to interact with the financial institute after 2 weeks have a greater potential for making a successful deal;
- there may exist a need in checking the validity of client data at early stages of a process, or extra attention should be paid to 147 cases when it was impossible to send established offers to clients.

Our recommendation is *to take measures for keeping the clients*:

- make a call/send a letter to a client after 15 and 30 days of absence of the client's response (to the same recommendation we came to in chapter 2.1);
- actively interact with clients, when a processing of request takes more than 10 days, while a client had not denied;
- to check contact information provided by a client before execution of any document, as this allows reducing number of cases when it was impossible to send established offers to clients.

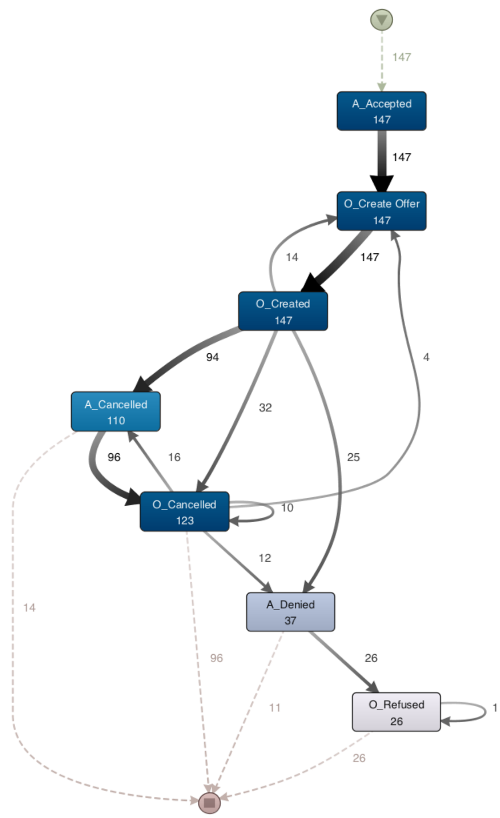

Fig. 17. Interesting cases with unsuccessful sending offer

4 Analysis of the density of interaction with client

Question 2. *What is the influence on the frequency of incompleteness to the outcome? The hypothesis here is that if applicants are confronted with more requests for completion, they are more likely to not accept the final offer.*

The answer is no, frequency and number of requests for completion does not influence the outcome. There is no evidence to reject null hypothesis that applicants that were called more frequently are less likely to accept final offer.

4.1 Method

For analysis purpose, each case was converted from standard XES format to the form of an attribute object. A total of 75 features were extracted from the cases, many of which were either not informative or served as target features (or features from which target feature can be explicitly derived).

The purpose of this analysis was to identify the effects of those attributes on the case endpoint. To do so, several random forests were constructed for the set of data on cases and their subsets. Random forest is a machine learning algorithm based on ensembling large amount of decision trees (100 in this case). It uses given features in attempt to estimate target feature value for each object. A Random forest algorithm was chosen due to it's useful ability to derive feature importances and work with incomplete data. Feature importances, derived from random forest regression were later used to filter out features and find most important features for close-up look. The figures below were plotted passing to random forest objects with only not empty and varying feature. Values plotted on graph is probability given by random forest based on this object. As more precise analysis, statistical significance tests were than performed on important features.

The program tool used for analysis is *Python 3.5* with *scikit-learn*, *scipy* and *pandas* packages.

4.2 Splitting cases by endpoint

To study the effect of indicators on the case outcome, it was first necessary to split cases by the type of outcome. The following case outcomes were identified (Table 4.2).

Endpoint	Description	Occurrence
Cases in progress.	Do not have A_Cancelled, A_Pending, A_Denied in the list of events.	(98 cases - 0.3%)
Client did not contact after the application was submitted.	Such cases do not have O_Returned in the events.	(9741 cases - 31%)
Client declined the application after receiving the bank offers.	Cases with O_Returned, _Cancelled	(959 cases - 3%)
Client declined the application after request for completeness.	Cases with O_Returned, A_Cancelled, and W_Call_incomplete_files	(328 cases - 1%)
Bank declined the application	Cases with A_Denied	(3752 cases - 12%)
Client has passed all stages and received a loan.	Cases with A_Pending	(17228 cases - 55%)

Table 4. Cases splited by endpoints

4.3 Influence of number and frequency of calls for completion on the final outcome

To answer the Question 2, as stated above, a random forest algorithm was trained on cases with at least one request for completion. This algorithm was trained to identify cases with endpoint "Client declined the application after request for completeness". After this, a list of feature importances was analyzed and the least important features were removed. As the process has been repeated several times, the following most important features were revealed:

– Mean time interval between completion calls;
– The time interval between last O_Sent or W_Call event and O_Returned.;
– The time interval between application submission and first offer;
– Requested amount.

The feature "Number of completion calls" did not show any significance and was removed in the process of finding relevant features.

As shown in Table 5, Clients who were not contacted at least once every 5 days on completion stage 1.4 more often don't respond.

Table 5. Statistical tests for various significant features

Outcome	Cases filter	Samples count	% Positives 90 confidence	Fisher's Exact test
Client not responded	mean first withdrawal <5000	5813	18.4 - 19.2 - 20.1	2.1e-80
	mean first withdrawal >5000	14911	29.2 - 29.8 - 30.4	
Client not responded	mean first withdrawal = 0	7912	38.7 - 39.6 - 40.5	5.0e-56
	mean first withdrawal >0	23597	27.5 - 28.0 - 28.5	
Client not responded	First offer interval <1 day	16968	28.5 - 29.1 - 29.7	4.9e-14
	First offer interval >1 day	14541	32.4 - 33.0 - 33.7	
Client not responded	Number of terms <60	10220	35.0 - 35.8 - 36.6	1.5e-50
	Number of terms >60	18519	26.8 - 27.3 - 27.8	
Cancelled after offers	Client responded & Monthly cost <400	17676	3.9 - 4.1 - 4.4	8.4e-05
	Client responded & Monthly cost >500	2355	5.2 - 6.0 - 6.8	
Cancelled after offers	N of completion calls $\geq$ 2& Completion interval <5 days	4067	4.5 - 5.1 - 5.6	1.3e-03
	N of completion calls $\geq$ 2& Completion interval >5 days	1675	6.3 - 7.3 - 8.4	
Cancelled after offers	First offer interval <1 day	12031	3.7-4.0-4.3	3.1e-03
	First offer interval >1 day	9737	4.5 - 4.9 - 5.2	
Cancelled after offers	Client responded & Requested amount <40000	20078	3.9 - 4.1 - 4.4	5.1e-11
	Client responded & Requested amount >40000	1398	7.0 - 8.2 - 9.4	

4.4 Other revealed patterns

In this section, we will describe patterns that appeared to have statistical significance

Probability that the client will cancel application depending on first contact interval. The time interval between an application submission and a first offer has influence on cancellation rates with statistical significance. However, cancellation rates are only influenced by several percent (Figure 18).

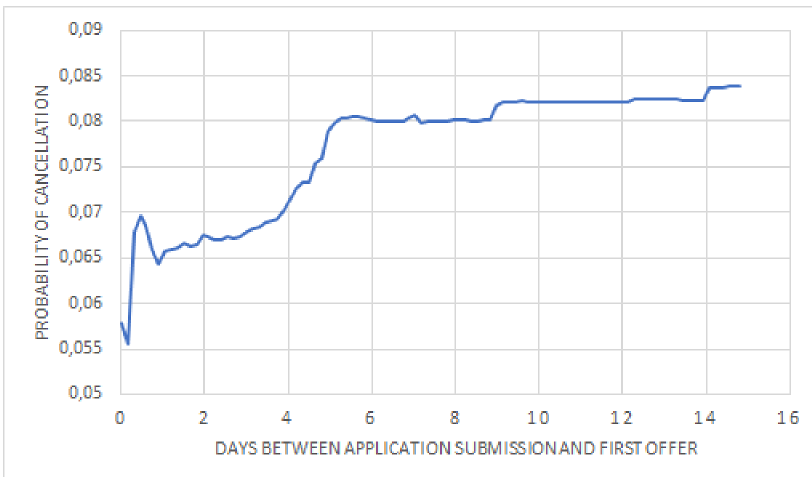

Fig. 18. Probability of cancellation after incompletion calls by first contact

Another statistically significant feature happened to be requested amount. It has statistical significance with dramatic increase in cancellation rates around 40 thousand (Figure 19).

The probability of client will not contact after the day of filing an application. The most important features influencing probability of not receiving response from client were discovered. Below are some conclusions made from Table 5.

– Clients, receiving offer with first withdrawal > €5,000 don't respond 1.5 times more often (Figure 21)
– Clients receiving offer with first withdrawal = 0 don't respond 1.4 times more often.
– Clients who were not contacted within first day don't respond 1.13 times more often.
– Clients who receive offer with number of terms < 60 don't respond 1.3 times more often.

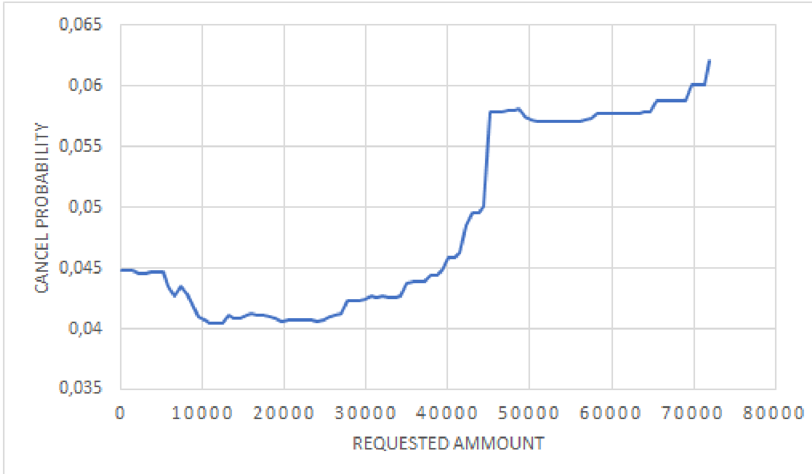

Fig. 19. Probability of cancellation after incompletion calls by first contact

The probability of client will refuse a loan offer. Here, case features influencing the probability of client cancellation after sending offers. Cases with users not responding after initial submission were not taken into consideration.

- Clients who receive offer with monthly cost >€500 respond 1.4 times less than ones who recieve offer with monthly cost <€400 (Figure 20)
- Clients who were not contacted at least once every 5 days on completion stage 1.4 more often don't respond.
- Clients who were not offered within 24 hours are 1.2 times more often cancel final offer.
- Clients who requested more than €40,000 are 2 times more often don't accept final offer.

Dependencies for the cases declined by the bank. The most significant indicator affecting the bank's deviation of an application is the requested amount. The bank does not like to approve the loans for less than 10 thousand c.u. As well, if the application is submitted on-line, in 13% of cases it will be declined by the bank, compared to 9% of declined cases for the applications submitted off-line.

4.5 Recommendations

To minimize number of client cancellations we recommend the following:

1. **Send offers to client as soon as possible.** For all case endpoints, this have been shown to have the greatest affect on the cancellation rates. Sending offers to clients within 4 days may decrease cancellation rates by 5% up to 10%.

Fig. 20. Probability of cancellation by offered monthly cost

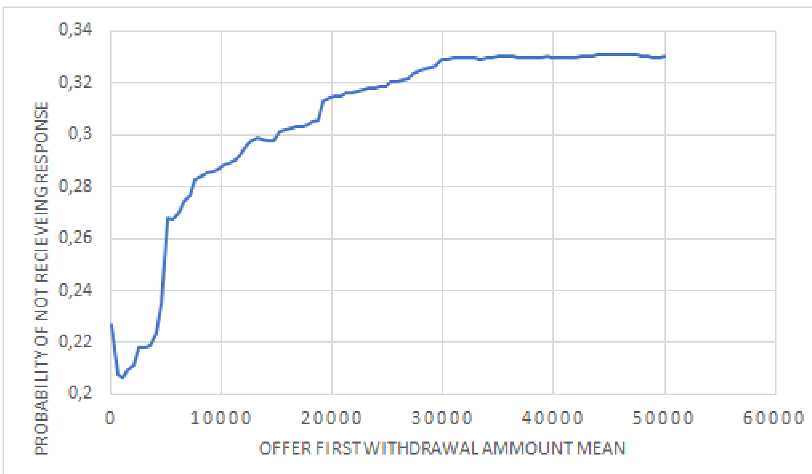

Fig. 21. Probability of not receiving response by first withdrawal

2. **Maximize number of terms.** By keeping offered number of terms above 60, total decrease of cancellation rates by 3% to 9% percent can be reached.
3. **Minimize first withdrawal amount.** By keeping offered first withdrawal amount below €5,000 a total decrease by 9% to 12% can be reached.
4. **Minimize monthly cost.** Offering to clients monthly cost below €400 cancellation rates can be brought down by up to 3%.
5. **Minimize time intervals between sending offers to client.** Keeping mean time of sending offers to client below 4 days may decrease cancellation rates by 2% to 5% percent.

To estimate impact of following these rules simultaneously, a more detailed knowledge of business process is required. However, cases, matching these conditions show 5% more *A_Pending*.

5 The effect of offering scenarios and number of offers on case endpoint

Question 3. *How many clients ask for more than one offer (where it matters if these offers are asked for in a single conversation or in multiple conversations)? How does the conversion compare between applicants for whom a single offer is made and applicants for whom multiple offers are made?*

Study all applications which contain more than one loan offer. With Celonis tool we can find cases which contain 2 or more credit offers (Figure 22). We get totally 8 559 cases of this type or 27% from total number of cases.

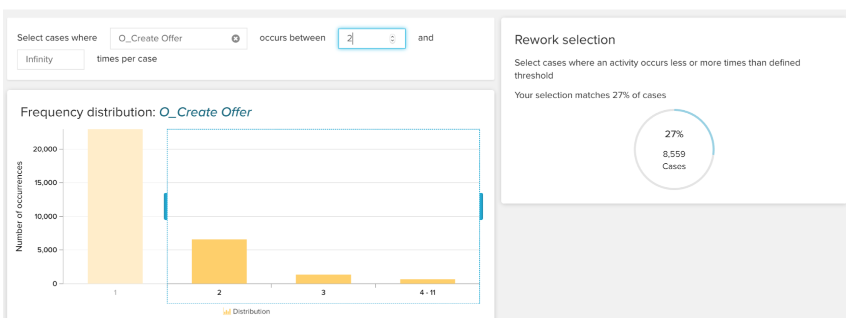

Fig. 22. Proportion of cases which contain 2 or more credit offers

Analyze the correlation of results for the application cases, when:

- single credit offer has been proposed to the client;
- more than one offer has been proposed to the client.

We can assume that when we propose more credit offers, we increase the probability of obtaining a successful result (increasing the number of cases with the final “A_Pending” event). We will assume that “multiple conversations” characterizes the different scenarios of the second and subsequent offers. Three such scenarios can be distinguished for the cases, where more than one loan offer has been proposed. Examine these scenarios on the example of the most popular case of issuing two loan offers on one application (21%/27%) (Figure 25).

1. Scenario **Create - Create:** preparation and issuance of two loan offers at once.

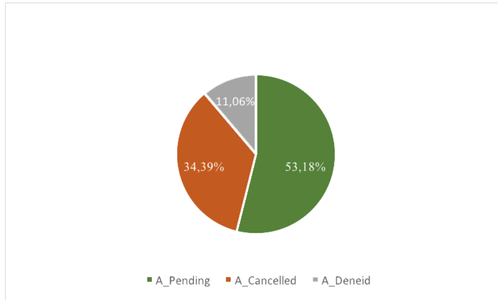

Fig. 23. Case with one offer

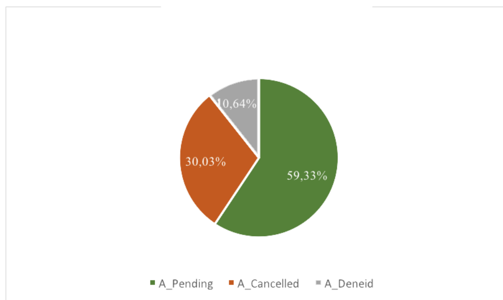

Fig. 24. Case with more than one offer

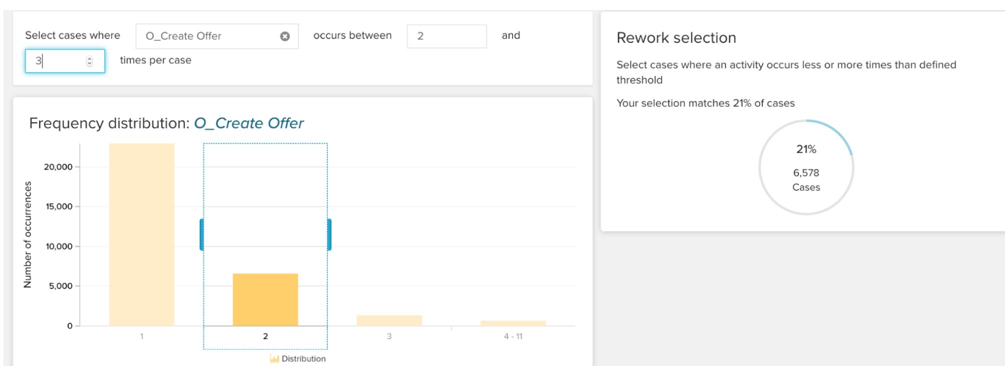

Fig. 25. Proportion of cases which contain 2 credit offers

2. Scenario **Create - Call - Create:** issuance of two loan offers in succession. Loan offer - call to the client - loan offer. It is likely that the client expresses a desire to get another loan offer.
3. Scenario **Incomplete - Create:** issuance of the second loan offer comes after receiving the necessary documents from the client on the first loan offer. It is likely that the first loan offer does not match the client's documents and the bank is forced to offer another one.

The issuance of three or more loan offers come under the same scenarios, or by the combinations thereof.

Analyze the correlation of results for the applications consideration regarding three scenarios, when issuing two, three, four and five loan offers (Figure 26, 27, 28, 29):

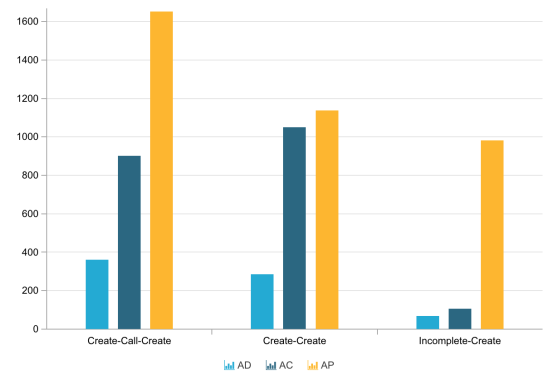

Fig. 26. 2 offers to client

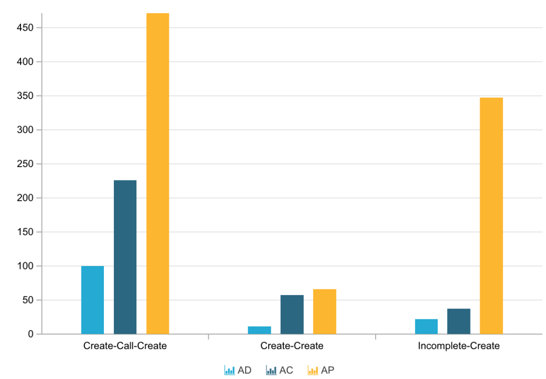

Fig. 27. 3 offers to client

Note:

– AC - cases containing A_Cancelled event
– AD - cases containing A_Denied event
– AP - cases containing A_Pending event

Scenario **Create-Create** raises questions about the desirability of offering two or more loan offers at once. If initially, when filling in the application, a

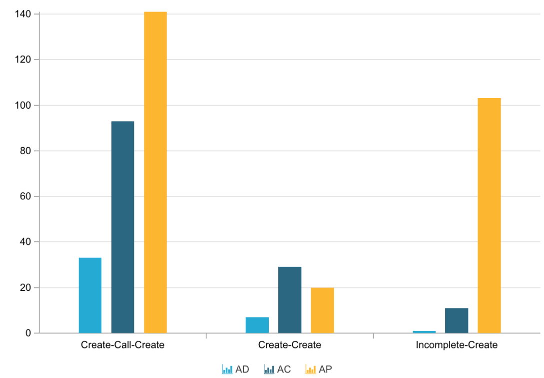

Fig. 28. 4 offers to client

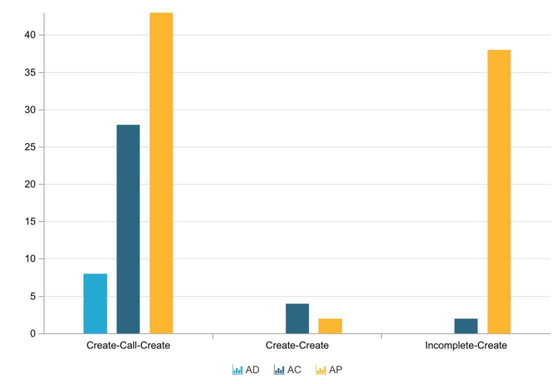

Fig. 29. 5 offers to client

client for a loan does not indicate the receipt of two or more loan offers himself, whether it's necessary to push him to multiple offers at once.

On the contrary, the provision of next offer after the next call (**Create-Call-Create**) has a relatively constant ratio of successful / unsuccessful situations.

Scenario **Incomplete - Create** is the most stable of all. This is due to the fact that the client is interested in granting the required amount of money and is ready to have to consider a new offer.

Table 6 contains all possible variants of the proposed loan offers on a case. Dependency diagram of the number of offers on a case against the final result - (Figure 30).

For cases with 1, 2, 3, 4, 5, 6 offers there is a gradual increase in the ratio of the number of successful applications and a decrease in the ratio of the number of applications canceled by the client. For the case with 7 offers, the ratio of

Table 6. Distribution of cases with different number of offers

Number of loan offers	offer _amount1	offer _amount2	offer _amount3	offer _amount4	offer _amount5	offer _amount = 6	offer _amount >6
Ends with A_Cancelled	7875	2058	320	133	34	4	7
Ends with A_Denied	2847	717	133	41	8	2	5
Ends with A_Pending	12178	3775	884	264	83	23	21

Fig. 30. The normalized dependency of the number of offers on a case to the final result

successful applications do not exceed the ratio of successful applications, when issuing 1 loan offer.

Recommendations: When proposing several loan offers, there is no negative trend of reducing the ratio of the number of successful applications to the number of canceled and rejected applications. Company should propose more offers to the client, if necessary. When issuing several loan offers, it is preferable to use such Scenarios as Create-Call-Create and Incomplete-Create.

6 Conclusion

Analyzing the process proposed in BPIC 17, we have found several dependencies that answer the questions posed. Also, during our analysis a number of observations were made, from which some recommendations were worked out.

The general recommendation will be to focus on the process of interaction with company customers: maximizing the necessary contacts and minimizing waiting times, as well as optimizing certain conditions for transactions with customers, which will help make the institute's offers more attractive. This will allow the company to get more successful applications and reduce the number of lost customers.

The following observations and recommendations were obtained:

- For customers who do not answer for a long time, it is necessary to implement the retention mechanisms: for example 2 calls (15 and 30 days after).
- With clients, who interacts with institute more than 10 days, company can communicate more intensive to increase probability of successful deal.
- There is need to improve process of checking clients data. It can decrease number of cases when offer can not be delivered to client.
- Time interval from application submission to first offer have been shown to have great influence on refusal rates, even on final stages of requesting incomplete information.
- Monthly cost and first withdrawal amount influence probability of cancellation after receiving offers significantly. Clients prefer smaller first withdrawal and monthly cost.
- There was a weak correlation between the number of proposed offers and the successful outcome. Company should propose more offers to the client, if necessary. Moreover, it is preferable to offer each next offer in accordance with the client's preferences.
- Logging workflow events requires improvement. Some events have a duration of 0 seconds, which is hardly correct. Proper logging would help improve further analysis.

7 Acknowledgement

The authors express their gratitude to the laboratory of Process-Aware Information Systems (PAIS Lab) [5] and, in particular, to Alexey A. Mitsyuk and prof. Irina A. Lomazova for the tactful guidance. Also we would like to express our gratitude to Jerome Geyer-Klingeberg (Celonis) for the opportunity to use the tool for research within BPIC 17 [1], by means of which an important part of the analysis was carried out.

References

1. TUE, https://www.win.tue.nl/bpi/doku.php?id=2017:challenge
2. van der Aalst, W. M. P. et al., Process Mining Manifesto, in Business Process Management
3. Jagadeesh Chandra Bose, R.P., Process Mining in the Large, Preprocessing, Discovery, Techniques, SIKS (2012) Workshops 2011, Lecture Notes in Business Information Processing, Vol. 99, SpringerVerlag, 2011
4. van der Aalst, Wil M.P., Process Mining, Discovery, Conformance and Enhancement of Business Processes, Springer Verlag, Berlin-Heidelberg (2011)
5. Laboratory of Process-Aware Information Systems (PAIS Lab), https://pais.hse.ru/en/ Business Processes, Springer Verlag, Berlin-Heidelberg (2011)
6. Wikipedia, https://en.wikipedia.org/wiki/Random_forest
7. Celonis, http://www.celonis.com/
8. Disco, http://www.fluxicon.com/products/
9. ProM, http://www.promtools.org/