BAKERY PROJECT REPORT

Marco Lassandro 945907

1 SCENARIO

Startup Bakery is an Italian startup studio, i.e. a "generator of startups", a company that builds new enterprises in series following a "parallel entrepreneurship" approach.

The startup studio model aims at significantly reducing the risk of failure, typical of newly created enterprises while enhancing the quality of the new startups. The key idea is implementing a well-established receipt that makes the most of the lean approach and appropriately uses the latest Cloud and AI technologies. Startup Bakery has developed a business process to guide and monitor the execution of the activities required to launch a startup, they directly work on the project and finance part of the investments. The final target is to sell the startup created to a corporation so for this reason they aim to accelerate the growth of the startup in the market.

The creation of a new startup follows four main phases:

- **Pick Phase**: in this phase the company **identify the best business idea** that can fit with the market's needs. Such phase **has a duration up to 6 months** where they **analyze different business idea** with **the help of a business analyst** and an **advisory committee**, they **evaluate** each business idea **by using different metrics**.
  Once the business idea is picked they make a so called **problem solution fit**, where there is an **initial validation on the market**, and if such validation has a positive outcome the business idea becomes a startup.

- **Mix Phase**: this phase **has a duration of 6 months** where there are **different players involved** in the validation and creation of the startup. Some activities are: the engagement of co-founder, found the startup, register the brand, create a prototype of the software, get enough pre-orders (so they see if there is enough interest on the target) and look for external investments.
  This is **the hardest phase** because **they need to lunch the startup in a limited time-frame** so the operations need to run quite smoothly, every piece of the puzzle needs to fit.

- **Bake Phase**: The startup goes in the **phase of development** of the software, they **hire different individuals** for both the IT and the marketing domain. They are focused in the creation of startups with limited team and resources, such research begins from the Mix Phase and end in the Bake phase.

- **Exit/Serve phase**: They **provide the startup to the corporation**, and **they work with them** in order **to facilitate the integration** of the company in their processes.
  As soon as they put the startup inside the corporation, **there could be integration issues**, because **corporation have slow processes** about bureaucracy. The processes of the created startup need to match with the ones of the corporation, they have to be sure the startup is very solid and quite organized.

2 GOALS

The most delicate phase of the Bakery Startup company, in the creation of new startups, **is the Mix Phase**. Here the **activities are strictly dependent** each other and **just a reiteration of an activity** can **increase significantly the risk of failure** of the process because of the limited time-frame imposed (6 months). **The company provided** the **process model** of the **Mix Phase**, that is **specified by a graphical representation** through the Business Process Model and Notation (**BPMN**). Moreover, **some activities** of the business process **are annotated** with related **time constraints**.
With the use of the BPMN model and the known time constraints the goals are:

- **GOAL 1**: Have an evaluation of the risk of failure of the process.

- **GOAL 2**: Reduce the risk of failure of the Mix phase's business process.

3 KNOWLEDGE UPLIFT MODEL

KUM for GOAL 1: Have an evaluation of the risk of failure of the process.

STEPS	INPUT	ANALYTICS/MODEL	TYPE	OUTPUT
STEP 1	BPMN, Known time constraints	Deduction of the time constraint for each activity in the BPMN	Predictive Knowledge	Time constraints of each activity
STEP 2	Mix Phase Logs, Time constraints	Time divergence between the time execution of each activity with its time constraint	Descriptive Knowledge	Time divergences
STEP 3	Mix Phase Logs, Time constraints	Frequency of timely activities, Frequency of not-timely activities	Descriptive Knowledge	Frequency of of timely and not timely activities
STEP 4	Frequency of of timely and not timely activities	Percentage of timely activities	Descriptive Knowledge	Percentage of timely-activities
STEP 5	Mix Phase Logs	Detection of repeated activities	Descriptive Knowledge	Execution time of repeated activities
STEP 6	Target Mix Phase duration	Definition of a tolerance with respect to the target Mix Phase duration	Prescriptive Knowledge	Mix Phase Tolerance
STEP 7	Time divergences, Execution time of repeated activities, Mix Phase Tolerance, target Mix Phase duration	Evaluation of the wasted time with respect to the tolerance	Descriptive Knowledge	Measure of the risk of Failure

KUM for GOAL 2: Reduce the risk of failure of the Mix phase's business process.

STEPS	INPUT	ANALYTICS/MODEL	TYPE	OUTPUT
STEP 1	BPMN, Known time constraints	Deduction of the time constraint for each activity in the BPMN	Predictive Knowledge	Time constraints of each activity
STEP 2	BPMN, Time constraints	Detection of the most time-consuming activities	Descriptive Knowledge	Most time-consuming activities
STEP 3	Most time-consuming activities	Identification of the pre-requisites to accomplish the most time-consuming activities	Descriptive Knowledge	pre-requisites
STEP 4	BPMN, logs	Detection of the most frequent loops	Descriptive knowledge	Top frequent loops
STEP 5	pre-requisites	Identification of pre-requisites that can be accomplished in the the Pick Phase process	Predictive knowledge	Selected pre-requisites.
STEP 6	Top frequent loops	Identification of activities in the loop that can be prepared in the Pick Phase	Predictive knowledge	Selected activities.
STEP 7	Selected pre-requistes, Selected activities	Business process modification of the Mix and Pick phase	Prescriptive knowledge	Modified BMPNs, modified time constraints.
STEP 8	Modified BMPNs, modified time constraints, logs of previous created startups	Evaluation of the new business process over previous logs	Descriptive knowledge	Validation of the modified BMPNs.

4 DATASET AND MATERIALS

In order to analyse the behaviours of the BPMN of the Mix Phase synthetic logs were generated by using the **PM4PY library**.
The original **BPMN has been modified** in order to convert it into a valid Petri Net, from which by a playout of the network the logs are generated. What It has been modified is:

- The introduction of **parallel gateways** in order to obtain, were needed, a Petri Net with concurrent transitions from the same source activity (below an example).
- It has been added new activities, so as to differentiate creation activities from reviewing activities, these activities are:
  - Prototyping refinements: this is connected with the **Technical Feasibility Analysis** and **Lean Market Validation** activities, because in case there are loops in this points what was done in the **Product Discovery Prototyping** will not repeated, but refined;
  - Business planning refinements: this is connected in a loop with the **Investors Roadshow** activity, for the same reasons of the activity above;
  - No Actions: this activity was inserted in order to detect the traces that didn't get investments, is connected with the **exclusive gateway** with the annotation "Is the co-founder the right one?";

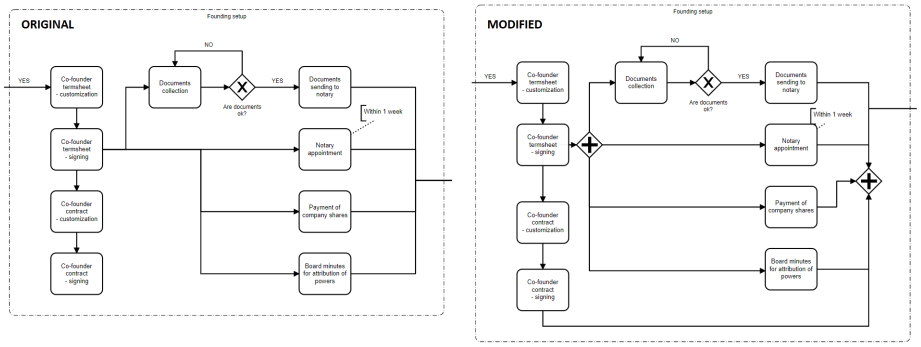

Figure 1: Example of updates done in the original BPMN, parallel gateways were added in order to get a parallel execution in the simulation of the resulting Petri Net)

The algorithm generates in particular so called traces, where each trace correspond to the Mix Phase of a given startup. Each trace contains all the activities executed. The algorithm provided by the PM4PY library, the **BASIC PLAYOUT** function, has been **modified** in order to introduce to each activity a start and an end timestamp. Such timestamps are based on the relation of a given activity with its **direct antecedents** and on the **time constraints imposed** (plus some randomness' factor).

The **unknown time constraints** of the activities has been deducted by the time constraints provided by the company. Each time constraint represent **the number of days** ideally needed to accomplish a specific activity. The time constraints for each activity can be shown in the **github repository**.

Some traces has been filtered out from the logs, because of their **alethic behaviour**. Such filtering is based on the number of repetitions of specific activities. It has been decided to get as valid traces the ones that respects these rules:

Activity	**MIN**	**MAX**
Co-founder engagement	1	4
Documents collection	1	5
Naming	1	5
Prototyping refinements	1	3
Investors roadshow	1	2
Lean market validation	1	2

Using these rules the number of traces have been reduced from **200 to 49**.

5 SOLUTION

5.1 SOLUTION FOR THE GOAL 1

For the first goal **a function has been developed** in order **to get** different **metrics** related to the **overall time execution** of a **trace** and **metrics** related to the **time execution** of **single events** of a trace. In particular what we obtain from such function is:

- For the trace:
  - **Time duration** of the trace: Tracked from the starting event to the ending event (expressed in days);
  - **Timely activities**: number of activities that respect the time constraint associated;
  - **Not timely activities**: number of activities that didn't respect the time constraint associated;
  - **Time accuracy**: the fraction of timely activities over the total number of events;
  - **Time divergence**: the time divergence between the time duration of the trace and the target duration, that is in this scenario set to 6 months (expressed in days);
  - **Process failure**: is evaluated through the fraction of the time divergence of the trace over the tolerance value. The **tolerance parameter** is expressed as the maximum number of days that a trace execution can waste, w.r.t the target duration, over which the risk of failure is certain. **It has been set this parameter to 60 days**;
- For each event of the trace:
  - **The name of the activity**;
  - **The start timestamp** of the activity;
  - **The end timestamp** of the activity;
  - **The time divergence** between **the time execution of the activity w.r.t the time constraint** associated (expressed in days);
  - If the event **fit or not the time constraint**;
  - **Current Process failure**: the measure of the risk of failure at that point in the timeline. Is evaluated by taking into account the time execution from the starting event to the current event and the target duration of the trace in the same point. Than the time divergence between these two elements will be used to evaluate the process failure w.r.t the tolerance parameter;

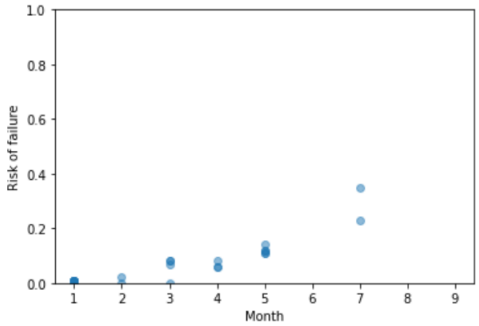

Figure 2: Process failure of a given trace during the months (old BPMN)

Taking into account **the 49 traces generated**, in the average **the results obtained** through this function are described in this table:

Metrics	Average results
Time duration of a trace	203 days
Process failure measure	0.43 with a tolerance of 60 days
Time divergence	23 days
Number of activities	46

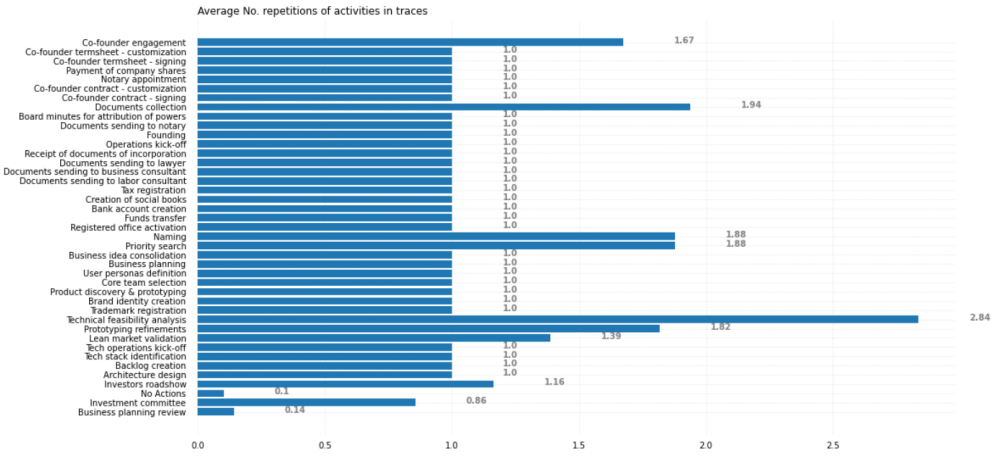

Figure 3: Average No. repetitions of activities in traces

Most of the traces **contains repetitions** of some activities, **the most time consuming repetitions** are the ones that involves the **Lean Market Validation** activity and the **Investors Roadshow** activity.

5.2 SOLUTION FOR THE GOAL 2

For the second goal my intuition is that there could be a **selected sub-team** of the company that put **the basis** for the activities of the Mix Phase for the **next potential new startups**.

In particular instead of choose a single best idea, **they could collect** time to time other potential good ideas, each of this ideas should have passed the so called "**problem solution fit**" in the **Pick Phase**.

In this way many **pre-requisites** for the most **time-consuming activities** in the Mix Phase can be prepared and so there could be a **reduction of the time needed** to create the next startups. Moreover, as stated in the previous solution **most of the traces has repetitions** [**Figure 3**], with the use of this sub-team it could be possible to **remove** some of these **loops**.

Given a potential good idea, what they can do is:

- Periodically **check** if there are potential **Co-founders interested** on the given business ideas (in order to avoid the loop of the co-founder engagement activity);
- **Pick** different possible names of the startup, and make **periodically the priority search** (in order to delete the loop of the naming and the priority search activity);
- Identify the **user personas** before the Mix Phase (In order to reduce the time needed to execute the business idea consolidation);
- Identify the **key-features** of the **software prototype** and **check** before the Mix phase if such features are **technically feasible** (in order to reduce the time needed for the business idea consolidation, the product discovery prototyping activity);

Based on **these assumptions** a **new BPMN** is proposed and the **time constraints** of some activities have been modified:

- The time constraint of the **Product discovery prototyping** activity **has been reduced** of **7 days**;
- The time constraint of the **Business idea consolidation** activity **has been reduced** of **2 days**;
- The **loop** between the **naming and the priority search** has been **removed**;
- **The loop** of the **Co-Founder engagement** has been **removed**;

The **new BPMN** can be shown in the **Github repository**. Using the updated BPMN **56 valid traces** as been generated and evaluated with the function developed in the first solution.

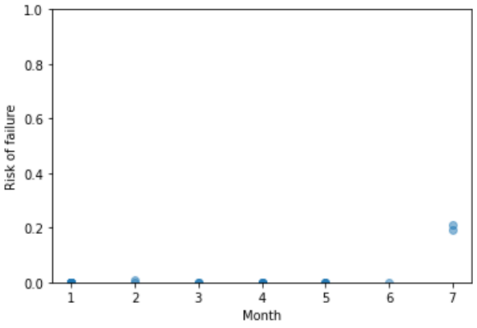

Figure 4: Process failure of a given trace during the months (updated BPMN)

In average the results obtained are:

Metrics	Old BPMN results	New BPMN results
Time duration of a trace	203 days	187 days
Process failure measure	0.43	0.28
Time divergence	23 days	7 days
Number of activities	46	43

6 KEY CONTRIBUTIONS

In this project **two main key contributions** have been introduced. First, the **algorithm provided** by the **PM4PY library** for the **simulation of event logs** over a Petri Net **has been extended** in order to have a functionality that generates more realistic logs, based on **time constraints associated** to each activity. **The second key** contribution is the **realisation** of a function able to make a **conformance check** of the logs on the **time domain** (explained in the **SOLUTION section**).

6.1 THE BASIC PLAYOUT ALGORITHM

Starting from a model, the **BASIC PLAYOUT** algorithm is able to **produce** an output that **follows** the model and the different rules that have been provided by the user.
Such function accepts a Petri Net along with an initial marking, and returns a specified number of process executions.
The problem is that **the timestamp** associated to an event of a give process execution **is not meaningful**, after an event is generated the timestamp is **incremented** by one second each time. Moreover, because of the **lack of an end timestamp** of the event it becomes difficult to understand the **time duration** of an **activity**.
For these reasons **the algorithm has been modified**, what was done is:

- The introduction of **four** new parameters:
  - **END_TIMESTAMP_KEY**: is the **key** useful **to retrieve** the **timestamp** related to **the end** of an event;
  - **TIME_CONSTRAINTS**: a **dictionary** where **each key is an activity** of the given Petri Net and **the value** indicates the **time normally needed** to **accomplish** that **activity**;
  - **ANTECEDENTS**: a **dictionary** where **the key** is **an activity** of the process model and **the value** is **a list** containing **the nearest activities antecedents to the target one**. Such parameter has been introduced **in order to put the right starting timestamp** to each event generated, that **could be dependent** to one or more different **concurrent activities**;
  - **RANDOMNESS_OF_TIMESTAMP**: Such parameter is **a fraction** used **to generate random time duration** of the activities based their time constraints;

The parameters introduced and **the extension of the code doesn't compromise** the previous **functionality** of such **algorithm**.

7 CONCLUSIONS

In conclusion, **the solutions found** for the achievement of the two goals proposed **seems to be reasonable**.

The first solution can be a good way **to monitor the behaviour of the process** and step by step **understand** what are **the main activities that bring most of the problems**, in terms of time. For **the solution of the second goal**, **many assumptions** has been taken, because of the lack of some more information.

For future works, in order to prove in a more better way the solution about the second goal, **could be useful** the introduction of **real logs** related to the Mix Phase **of different created startups** in the Bakery Studio. Moreover, **can be also helpful** the definition of a **BPMN of the other phases** (the Pick, Bake and Serve phase) in order to get a wider view of the overall business process, in this way the solutions provided will be more consistent, for the reduction of the Mix phase's risk of failure.