

## Simulação, Mineração e Implementação de Subprocessos em Turismo Inteligente e Sustentável (Smart & Sustainable Tourism)



- dominar técnicas de **mineração** de processos (*process mining*) a partir de um log de eventos, tais como a sua descoberta (*process discovery*), análise (*process analysis*) e mineração dos parâmetros de execução;
- otimização/balanceamento dos recursos atribuídos a um processo, através de **simulação** de vários cenários, evitando gargalos (*bottlenecks*), aumentando a eficiência e reduzindo o custo dos recursos utilizados e melhorando o tempo de resposta;
- experiência de utilização de uma plataforma *low-code* para gerar uma **implementação** executável de um subprocesso modelado na 1ª parte deste trabalho.

- Instale o interpretador da [linguagem R](#) (download [aqui](#)) e a versão open-source do IDE [R Studio](#) (download [aqui](#));
- Crie uma conta na versão académica da ferramenta *Apromore* [aqui](#), usando obrigatoriamente o endereço de email do ISCTE-IUL. Esta será a ferramenta usada para mineração e comparação de cenários de simulação.
- Faça login na conta criada em <https://academic-eu.apromore.org/>
- Leia o [tutorial introdutório](#) que contém uma série de vídeos (total 90 minutos) e registe o link para o [manual de utilizador](#) que provavelmente irá necessitar durante a execução deste trabalho.
- Relativamente à preparação da componente de implementação, siga as instruções descritas em [AgileKip](#)

## 1. Descoberta, animação e análise de um processo através de mineração

A **descoberta** de um processo (*process discovery*) é a recuperação de uma representação do processo que está a ser realmente executado, descrevendo as possíveis sequências (variantes) de execução das atividades, com base num log de eventos em que cada linha corresponde à execução de uma instância de um elemento de fluxo (atividade ou

evento) do modelo. A **animação** de um processo permite visualizar as suas instâncias em execução desde a sua criação até à sua conclusão. A **análise** de um processo tem como objetivo compreender os processos, nomeadamente identificando as atividades desenvolvidas, os recursos gastos (tempo e dinheiro) e as características das variantes no sequenciamento das atividades, entre outras coisas.

Na aula teórico-prática importe no *Apromore* o log de eventos do processo **BankLoan** disponibilizado no e-learning em formato **mxml.gz** e preencha a **Ficha de Mineração** (em anexo). Compare os resultados com os dos colegas de grupo e, se tiver dúvidas, com os do professor.

Também no e-learning, no separador do trabalho (*assignment*), é disponibilizado um log de eventos adicional (diferente para cada grupo), com o nome **G###.csv**, em que **##** é o número do seu grupo. Usando esse log, que está em formato CSV, contido no zip **PurchaseProcess\_EventLogs.zip**, execute os seguintes passos no *Apromore*:

- crie uma pasta com o nome **MIP\_S&M\_G##** (onde **##** será o número do grupo) e partilhe-a com privilégio “owner” com os seus colegas de grupo e os utilizadores **Joao.Caldeira** e **Fernando.BritoAbreu**;
- lance a importação do ficheiro CSV e identifique os atributos mais relevantes (*Case ID, Activity, Start timestamp, End timestamp, Other timestamp, Resource, Role, Case attribute, ou Event attribute*);
- depois de importado o log na pasta partilhada, prossiga para a descoberta do processo, em formato BPMN, usando a perspetiva das atividades e selecionando todos os nós, arcos e paralelizações; observe o modelo descoberto em que aparecem eventos de início e fim, atividades e portas;
- execute a **animação** do modelo e constate que os fluxos de sequência mais a cheio correspondem àqueles onde circulam mais instâncias do processo (marcas a azul);
- pare a animação e preencha, em conjunto com os seus colegas, a **Ficha de Mineração** para o log do seu grupo;
- no modelo descoberto no *Apromore*, em formato BPMN, selecione a perspetiva das atividades e a abstração por frequência de casos (*Frequency: Case*); mantendo o número de nós (atividades) e o grau de paralelismo a 100%, reduza o número de arcos a 0, o que permite obter um modelo simplificado; exporte o modelo resultante com o nome **DiscoveredModel.bpmn** e importe-o no *Signavio*;
- ainda no *Apromore*, guarde o diagrama no ficheiro **CaseFrequencies.png** e, com base no número de instâncias (casos) que saem de cada *fork gateway* de “ou exclusivo”, determinado por mineração do ficheiro de eventos, calcule as percentagens relativas de cada fluxo (arco) de saída (nota: a soma deve dar 100%); registre essas percentagens nos ramos respectivos no modelo importado no *Signavio* (ex.: 213 cases / 58%);
- no *Signavio*, usando a versão inicial e a informação contida no ficheiro de eventos sobre que tipo de recurso executou cada uma das atividades (o que pode ser obtido no *Apromore*), melhore o modelo recuperado, acrescentando-lhe elementos de modelação adicionais que lhe parecerem relevantes, como *pools, lanes, gateways*, eventos intermédios, mensagens, etc [**sugestão**: comece com apenas um processo e distribua as atividades pelas *lanes* relativas a cada papel (*role*)];
- guarde o modelo melhorado no ficheiro **ImprovedModel.bpmn**

## 2. Mineração dos parâmetros estatísticos relativos à execução de um processo

Seja para preparar um modelo de simulação, seja para compreender a realidade existente, é fundamental conhecer os **parâmetros estatísticos** que caracterizam a execução de um processo. Para otimizar um processo é preciso conhecer a sua *baseline*, isto é, o ponto de partida, que será o caracterizado pelo log de eventos carregado no *Apromore* no passo anterior. Essa *baseline* é caracterizada por:

- recursos atribuídos a cada *lane* do processo, que podem ser determinados no *Apromore* (nº de *resources* por cada *role*);
- horário de trabalho (*work schedule*) associado a cada *lane/role*; considere que um *Requester* “trabalha” 24/7, o que é habitual em compras online e que todos os outros *roles* operam num horário regulado pelo sindicato: 9 às 17h de 2ª a 6ª feira;
- probabilidades de cada ramo nas *gateways* de “ou exclusivo”, já anteriormente determinadas;
- distribuições probabilísticas do ritmo de criação de instâncias do processo e do esforço de execução de cada atividade, cujo tipo e correspondentes parâmetros podem ser obtidos através do script R disponibilizado no separador *Process Mining & Simulation* no e-learning.

Importe o modelo **ImprovedModel.bpmn** no [BIMP](#) e introduza-lhe toda a parametrização precedente. Altere o horário “default” para um horário de 24/7 e dispare uma simulação inicial, selecionando a opção de geração do log em formato MXML. Quando não tiver erros, guarde este primeiro cenário de simulação (opção “Save scenario”) com o nome **DiscoveredScenario.bpmn** e o ficheiro MXML (opção “download MXML”) com o nome **DiscoveredSimulationLog.mxmml.gz**

### 3. Otimização/balanceamento dos recursos atribuídos a um processo

Leia os *slides* e outro material fornecido no e-learning sobre o processo de otimização do processo através de simulação.

Começando com o modelo **DiscoveredScenario.bpmn** crie os seguintes cenários de simulação e guarde os resultados respetivos nos ficheiros indicados. Note que terá de calcular o custo de cada alternativa. A descoberta do cenário otimizado obrigará a experimentar várias simulações, ajustando os recursos disponíveis para cada *lane* e mesmo a mudança dos horários de trabalho (ex.: part time, horas extraordinárias, horários semanais mais flexíveis).

CENÁRIO	DESCRIÇÃO	FICHEIROS
Minimalista	Cada pool só com 1 recurso	<i>MinimalistScenario.bpmn</i> <i>MinimalistSimulationLog.mxmml.gz</i>
Esbanjador	Cada pool com tantos recursos como o número de casos (instâncias do processo)	<i>WastefulScenario.bpmn</i> <i>WastefulSimulationLog.mxmml.gz</i>
Otimizado	Cenário que otimiza o rácio $1/(\text{custo} \times \text{duração média do ciclo})$	<i>OptimizedScenario.bpmn</i> <i>OptimizedSimulationLog.mxmml.gz</i>

Importe os 4 ficheiros MXML produzidos anteriormente na ferramenta *Apromore*, usando a opção *Analyze / View performance dashboard*. Faça download dos diagramas produzidos para o seu relatório e comente-os. Identifique que SLOs (*Service Level Objectives*) poderá oferecer num SLA com base no cenário otimizado.

### 4. Implementação de um processo

Por **implementação**, ou automação de um processo entende-se a construção de um sistema de software (*Process Aware Information System*) que apoia e controla a execução das tarefas associadas a um modelo de processo. Pretende-se neste passo utilizar a plataforma *low-code* [AgileKip](#) que permite gerar uma implementação executável de um subprocesso BPMN modelado na 1ª parte deste trabalho.

Dado que este passo do trabalho se destina a perceber e obter experiência “hands-on” na implementação de processos e não a implementar um sistema real e completo, apenas iremos considerar o seguinte subconjunto das abstrações de modelação disponíveis em BPMN: *Start Event*, *End Event*, *Timer Event*, *User Event*, *User Task*, *Service Task*, *Message Task*, *Subprocess* e *Gateways*. O subprocesso escolhido deve ter **pelo menos 6 (seis) elementos de modelação por membro do grupo**. Por elementos de modelação entendem-se Eventos, Tarefas e Gateways (os fluxos não contam). Deve usar pelo menos (i.e. no mínimo), um (1) *Intermediate Event*, uma (1) *Service Task*, uma (1) *Message Task*, quatro (4) *User Tasks*, duas (2) *Domain Entities* e uma (1) *Gateway*.

Uma vez que a plataforma [AgileKip](#) é baseada no [Camunda Workflow Engine](#) (*Community Edition*), cuja cobertura do metamodelo da linguagem BPMN é a [aqui](#) indicada (símbolos na cor laranja), o subprocesso BPMN escolhido, cumprindo com os requisitos mínimos anteriormente indicados, deve ser importado na ferramenta [Camunda Modeler](#), para confirmação da compatibilidade.

Para a geração do código será necessário produzir um conjunto de especificações em formato [JSON](#) descrevendo entidades auxiliares como telas (écrans) e entidades de domínio, tal como descrito na documentação encontrada em

[AgileKip](#). Será ainda necessário produzir alguns trechos de código em Java e JavaScript, para os quais são disponibilizados exemplos ilustrativos.

Para demonstrar a implementação do sistema, devem ser disponibilizados os seguintes artefatos:

- modelo de processo utilizado (***ImplementationModel.bpmn***);
- endereço no [GitHub](#) com todo o código fonte (JSON, Java, Javascript, ...) que for produzido;
- endereço no [YouTube](#) contendo um vídeo de 5-10 minutos explicando o modelo BPMN utilizado e demonstrando a execução (passo-a-passo) do sistema desenvolvido.

## Dúvidas

As aulas teórico-práticas não se destinam a tirar dúvidas sobre este trabalho, mas sim a sedimentar conceitos que apresentados na(s) aula(s) teórica(s) precedente(s). As dúvidas sobre este trabalho poderão ser tiradas nos horários que serão anunciados através da plataforma de e-learning, os quais, sempre que justificável, implicarão a escolha de um slot de atendimento através da ferramenta [Doodle](#).

## ENTREGA

### Data, local e formato

A entrega desta parte do trabalho deve ser efetuada num link identificado para o efeito, na página “**ASSIGNMENTS**” no *e-learning* até às **23:59 do dia 22 de Dezembro de 2021**. Qualquer membro do grupo pode efetuar uma entrega. Podem ser realizadas várias entregas, mas apenas a última será considerada para avaliação. O formato da entrega de cada grupo aparecerá descrito na página supracitada.

## Penalizações

Entregas após a data limite terão uma penalização de **2 valores por cada dia incompleto de atraso** (ex.: uma entrega no dia 23 de Dezembro às 07:05 terá 2 valores de penalização).

O envio de informação não solicitada irá causar perturbações no processo de correção e por isso será penalizado na classificação do trabalho com **1 valor**.

## AVALIAÇÃO DO TRABALHO

A avaliação desta parte terá os pesos constantes na seguinte tabela. A componente de implementação tem um peso superior às restantes porque envolverá um esforço relativo maior que cada uma das outras em separado.

TÓPICO	PESO
SIMULAÇÃO	30%
MINERAÇÃO	30%
IMPLEMENTAÇÃO	40%

### Aferição individual dentro do grupo

Caso existam discrepâncias entre o trabalho levado a cabo por cada um dos membros do grupo, cada um poderá dar no máximo 1 valor seu a outro membro do grupo para compensar, ainda que parcialmente, eventuais injustiças. A liberalidade de cada membro do grupo fazer esta oferta deve ser expressa no final do relatório, numa seção com o mesmo título que aqui: “*Aferição individual dentro do grupo*”. A pauta dos trabalhos refletirá estas aferições. Assim, se por exemplo a um trabalho de um grupo de 2 alunos for dada a nota 15, um dos alunos poderá ficar com 16 e o outro com 14. Num grupo de 3 alunos num trabalho com 14 valores, um dos alunos poderá ficar com 16 e os outros dois com 13, ou então um com 15, outro com 14 e outro com 13. O grupo pode obviamente decidir não usar este mecanismo de aferição.

## FICHA DE MINERAÇÃO DE PROCESSO

**Modelo:** \_\_\_\_\_

**Ficheiro de eventos:** \_\_\_\_\_ / \_\_\_\_\_ (KB)

Instruções: se numa dada pergunta existir mais do que um caso ou atividade que correspondam à resposta correta, basta indicar um@ del@s.

Parâmetro	Valor
Período de tempo total	
Nº de casos (instâncias de processo)	
Nº de variantes	
Nº de atividades distintas	
ID e duração do caso mais curto e do mais demorado	
Nome e nº de ocorrências das atividades mais e menos frequentes	
Nome e duração média das atividades não instantâneas com maior e menor duração média	
Nº de recursos distintos	
Nome dos recursos com mais e menos casos associados e correspondente nº de ocorrências	
Nome do recurso com mais e menos atividades executadas e correspondente nº de ocorrências	
Nome do recurso com maior e menor tempo médio gasto na execução de atividades e correspondente valor	
Intervalo de tempo em que ocorreu um maior número (pico) de casos ativos (indique também esse nº de casos)	
Variante mais frequente (indique o nº de atividades, nº casos e o intervalo de variação da duração no formato [min, max])	
Duração média da actividade com maior duração correspondente à variante mais frequente	

### Distribuição global dos recursos

Represente um gráfico com a alocação (frequências e percentagens) dos vários recursos à totalidade dos casos