











<b>Decisões</b>  <p>Como as previsões são usadas para tomar decisões que fornecem o valor proposto ao usuário final?</p> <p>Evitar o trabalho de ler um documento por inteiro para apenas classificá-lo.</p>	<b>Tarefa de ML</b>  <p>Entrada, saída a ser prevista, tipo de problema.</p> <p><b>Receber um documento oficial e classificá-lo em sua categoria.</b></p>	<b>Proposições de Valor</b>  <p>O que estamos tentando fazer para o(s) usuário(s) final(is) do sistema preditivo? Quais objetivos estamos cumprindo?</p> <p><b>Automatizar a tarefa de classificação dos documentos da SEFAZ.</b></p>	<b>Fontes de Dados</b>  <p>Quais fontes de dados brutos podemos usar (interno e externo)?</p> <p><b>Os documentos disponíveis na <a href="#">SEFAZ</a>.</b></p>	<b>Coleta de Dados</b>  <p>Como obtemos novos dados para aprendizado (entradas e saídas)?</p> <p><b>Os dados são obtidos por web scrapping do <a href="#">site da SEFAZ</a>.</b></p>
<b>Execução das Previsões</b>  <p>Quando fazemos previsões sobre novas entradas? Em batch ou online? Quanto tempo temos para trabalhar uma nova entrada e fazer uma previsão?</p> <p><b>Previsões online a medida que são adicionados documentos.</b></p>	<b>Avaliação Offline</b>  <p>Métodos e métricas para avaliar o sistema antes da implantação.</p> <p><b>Acurácia, Precisão e Sensibilidade do Modelo.</b></p>	<b>Avaliação e Monitoramento Online</b>  <p>Métodos e métricas para avaliar o sistema após a implantação e para quantificar a criação de valor.</p> <p><b>Coerência da classificação dos dados</b></p>	<b>Atributos (Features)</b>  <p>Representações de entrada extraídas de fontes de dados brutas.</p> <p><b>Os atributos são o Texto e o Título do documento.</b></p>	<b>Criação de Modelos</b>  <p>Quando criamos/atualizamos modelos com novos dados de treinamento?</p> <p><b>Quando surgem novas Categorias</b></p>



## Decisões

Como as previsões são usadas para tomar decisões que fornecem o valor proposto ao usuário final?

**O algoritmo irá encontrar as palavras-chave de cada tipo de documento, permitindo ao usuário classificar documentos de acordo com os termos usados.**



## Tarefa de ML

Entrada, saída a ser prevista, tipo de problema.

**Entrada:** um documento oficial de legislação da SEFAZ

**Saída:** tipo de documento.

**Problema:** Analisar o documento para conseguir uma classificação adequada.



## Proposições de Valor

O que estamos tentando fazer para o(s) usuário(s) final(is) do sistema preditivo? Quais objetivos estamos cumprindo?

**Nossa proposta é obter um método para classificar automaticamente os documentos da SEFAZ, para que novos documentos entrem automaticamente no seu modelo de organização.**



## Fontes de Dados

Quais fontes de dados brutos podemos usar (interno e externo)?

**Os dados usados para nossa tarefa foram retirados do site da Sefaz.**



## Coleta de Dados

Como obtemos novos dados para aprendizado (entradas e saídas)?

**A Coleta de dados é feita pelo site da Sefaz, através de operações de web scraping.**



## Execução das Previsões

Quando fazemos previsões sobre novas entradas? Em batch ou online?

**As previsões devem ser feitas assim que um novo documento for cadastrado no sistema.**



## Avaliação Offline

Métodos e métricas para avaliar o sistema antes da implantação.

**Acurácia ,Precisão e Sensitividade do Modelo.**



## Avaliação e Monitoramento Online

Métodos e métricas para avaliar o sistema após a implantação e para quantificar a criação de valor.

**Após a implantação deve ser analisada a coerência das novas classificações realizadas pelo modelo com as já existentes observando se o modelo apresentou dúvidas em relação às probabilidades obtidas.**



## Atributos (Features)

Representações de entrada extraídas de fontes de dados brutas.

**Texto e Título(Label) do Documento .**



## Criação de Modelos

Quando criamos/atualizamos modelos com novos dados de treinamento?

**O modelo deve ser atualizado quando um novo tipo, ou subtipo, de documentos é inserida.**