Decisões



Tarefa de ML

sua categoria.

Avaliação

Offline



Proposições de Valor



Fontes de Dados



Coleta de Dados



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

Evitar o trabalho de ler um documento por inteiro para apenas classificá-lo.

Entrada, saída a ser

prevista, tipo de problema.

Receber um documento oficial e classificá-lo em O que estamos tentando

fazer para o(s) usuários(s) final(is) do sistema preditivo? Quais objetivos estamos cumprindo?

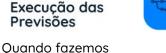
Automatizar a tarefa de classificação dos documentos da SEFAZ.

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

Os documentos disponíveis na SEFAZ. Como obtemos novos dados para aprendizado (entradas e saídas)?

Os dados são obtidos por web scrapping do site da SEFAZ.

Previsões



previsões sobre novas

entradas? Em batch ou

temos para trabalhar uma

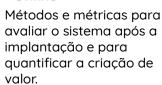
nova entrada e fazer uma

online? Quanto tempo

Métodos e métricas para avaliar o sistema antes da implantação.

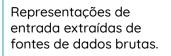
Acurácia, Precisão e Sensibilidade do Modelo.

Avaliação e Monitoramento Online



Coerência da classificação dos dados

Atributos (Features)



Os atributos são o Texto e o Título do documento.

Criação de Modelos

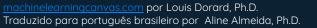
Categorias



Ouando criamos/atualizamos modelos com novos dados de treinamento? **Quando surgem novas**

Previsões onlines a medida que são adicionados documentos.

previsão?





Decisões

Tarefa de ML



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.

Entrada, saída a ser prevista, tipo de problema.

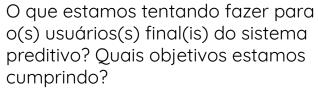
Entrada: um documento oficial de legislação da SEFAZ

Saída: tipo de documento.

Problema: Analisar o documento para conseguir uma classificação adequada.



Proposições de Valor



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 <u>site da Sefaz</u>.



Coleta de Dados



Execução das Previsões

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

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





<u></u> {ô

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.

