



Predição de Atraso de Voos Estado-Unidenses Utilizando Técnicas de Aprendizado de Máquina

Introdução

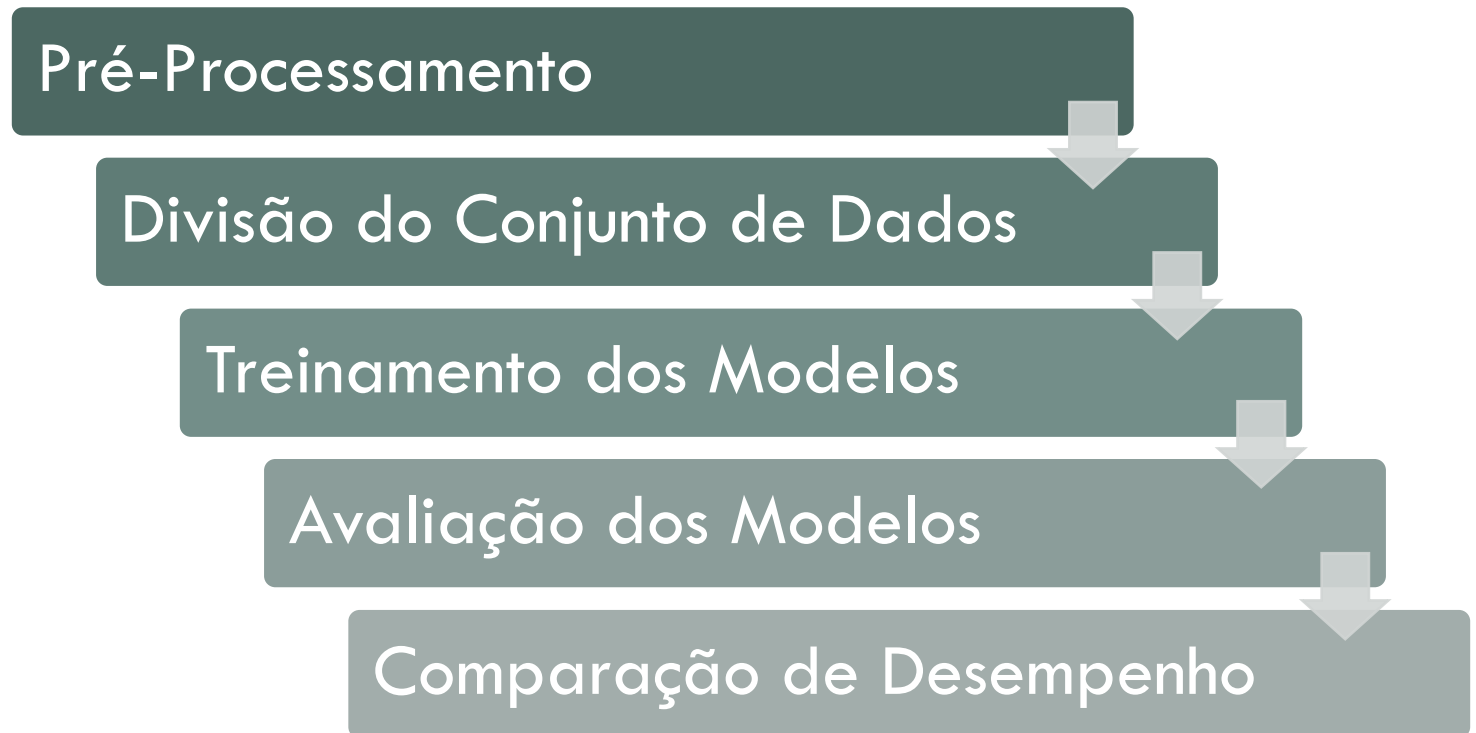
Este estudo busca prever a ocorrência de atrasos em voos comerciais nos EUA usando técnicas de aprendizado de máquina.

Com a crescente demanda por viagens aéreas, a pontualidade dos voos é crucial, mas os atrasos causam transtornos e custos substanciais.

Utilizando uma base de dados abrangente, o objetivo é desenvolver modelos capazes de identificar padrões e correlações para prever a probabilidade de atrasos, contribuindo para aprimorar a gestão do tráfego aéreo e beneficiar tanto as companhias aéreas quanto os passageiros.

Metodologia

A metodologia adotada neste estudo consistirá em uma abordagem abrangente que compreende várias etapas distintas, desde o pré-processamento dos dados até a avaliação dos modelos de aprendizado de máquina.



Obtenção e Preparação dos Dados

Flight	Identificador do voo, representado por números inteiros.
Time	Horário de partida do voo, expresso no formato de minutos.
Length	Duração do voo, também expressa no formato de minutos.
Airline	Sigla que identifica a companhia aérea responsável pelo voo.
AirportFrom	Sigla que identifica o aeroporto de origem do voo.
AirportTo	Sigla que identifica o aeroporto de destino do voo.
DayOfWeek	Dia da semana em que o voo ocorreu, representado por números inteiros de 1 a 7, onde 1 corresponde a segunda-feira e 7 a domingo.
Class	Variável alvo que indica se o voo sofreu atraso (1) ou não (0).

Obtenção e Preparação dos Dados

Flight	Time	Length	Airline	AirportFrom	AirportTo	DayOfWeek	Class
23130	12960	1410	DL	ATL	HOU	1	0
69480	3600	1460	OO	COS	ORD	4	0
12470	11700	1430	B6	BOS	CLT	3	0
310	14100	3440	US	OGG	PHX	6	0
5630	6920	980	FL	BMI	ATL	4	0
36920	5800	600	WN	MSY	BHM	4	0
11350	6900	2390	CO	EWR	DFW	4	0
13000	12100	800	AA	DFW	MEM	3	0
5870	12950	1050	FL	BWI	GRR	7	0
7640	5300	1080	FL	ATL	PBI	3	0

Obtenção e Preparação dos Dados

- Pré-Processamento:
 - Remoção de valores ausentes
 - Normalização das variáveis numéricas
 - Conversão das variáveis categóricas
 - Remoção de atributos não relevantes

```
### PROCESSADOR ###  
caminhoArquivo = 'D:/dev/projetos/machineLearning/classificacao-delay-airlines'  
  
configs = {  
    'col_classe': 'Class',  
    'nome_arquivo': caminhoArquivo+'/Classificação/airlines_delay.csv',  
    'remover_colunas': ['Flight'],  
    'concatenacao': {  
        'col1': 'AirportTo',  
        'col2': 'AirportFrom',  
        'concat': '-',  
        'col_nova': 'Airports',  
        'drop_cols': True  
    },  
    'cols_categoria_nominal': ['Airports'],  
    'cols_dummy': ['Airline'],  
    'padronizacao': False  
}  
  
processador = PreProcessador(configs)
```

Obtenção e Preparação dos Dados

- Divisão do conjunto de dados:
 - Base de treinamento (75%)
 - Base de teste (25%)

```
### PROCESSADOR ###
caminhoArquivo = 'D:/dev/projetos/machineLearning/classificacao-delay-airlines'

configs = {
    'col_classe': 'Class',
    'nome_arquivo': caminhoArquivo+'/Classificação/airlines_delay.csv',
    'remover_colunas': ['Flight'],
    'concatenacao': {
        'col1': 'AirportTo',
        'col2': 'AirportFrom',
        'concat': '-',
        'col_nova': 'Airports',
        'drop_cols': True
    },
    'cols_categoria_nominal': ['Airports'],
    'cols_dummy': ['Airline'],
    'padronizacao': False
}

processador = PreProcessador(configs)
```

Obtenção e Preparação dos Dados

- Treinamento dos Modelos:
 - Classe majoritária
 - K-Nearest Neighbors (KNN)
 - Naive Bayes
 - Árvore de Decisão
 - Random Forest
 - Support Vector Machine (SVM)
 - Redes Neurais

```
### CLASSIFICAÇÃO ###  
classificador = Classificador(processador)  
classificador.RedesNeurais(verbose = True,  
                             max_iter= 2500,  
                             tol = 0.0000001,  
                             solver = 'sgd',  
                             hls=[10,10],  
                             activation='relu',  
                             random_state = 1)
```


Obrigado

Saulo Klein Nery

saulo.nery@aluno.cefet-rj.br

Thaís Munier Wermelinger

thais.wermelinger@aluno.cefet-rj.br