

Gabriel Augusto

Curso: Engenheiro de Machine Learning - Udacity

Histórico do assunto

A necessidade de aumentar a receita em voos e evitar prejuízos é algo comum para qualquer empresa aérea. Entre causas que acabam causando prejuízo existe o mau funcionamento de partes do avião e atrasos e cancelamentos devido ao "mau tempo".

Para ser permitido a partida ou pouso de um voo é necessário condições meteorológicas aceitáveis. Caso as condições meteorológicas não sejam aceitáveis isso poderá acarretar mudanças de rotas, indenizações para os passageiros. Por isso é necessário um planejamento por parte das companhias aéreas para como se portar no meio dessas condições adversas.

Descrição do problema

A partir de dados meteorológicos de aeroportos dados (METAR¹) e de status de voos antigos Ex: realizado, atrasado por condições meteorológicas adversas, cancelado por condições meteorológicas adversas, ...

Prever horas antes se um outro voo terá atraso ou o risco de cancelamento do mesmo.

1. METAR - (METeorological Aerodrome Report - Informe meteorológico regular de aeródromo), é um informe codificado, associado às observações meteorológicas à superfície, e utilizado para fornecer informações sobre condições do tempo em um aeródromo específico.

Inputs

1º Conjunto de dados:

Entrada de dados METAR do Brasil retirada do site da Universidade de Iowa nos EUA:

https://mesonet.agron.iastate.edu/request/download.phtml?network=BR_ASOS

(Os arquivos dos aeroportos escolhidos estão no link enviado)

station: (Aeroporto)

three or four character site identifier (Identificador de 3 ou 4 caracteres)

valid: (válido) -> timestamp of the observation (Timestamp de observação)

tmpf: (Temperatura em Fahrenheit) -> Air Temperature in Fahrenheit, typically @ 2 meters

dwpf: (Ponto de orvalho em Fahrenheit) ->Dew Point Temperature in Fahrenheit, typically @ 2 meters

relh: (Umidade relativa) ->Relative Humidity in %

drct: (Direção do Vento em graus do norte) ->Wind Direction in degrees from north

sknt: (Velocidade do Vento em nós) -> Wind Speed in knots

p01i: Uma hora por um período de tempo de observação da última reposição de precipitação por hora. Valores em polegadas. -> One hour precipitation for the period from the observation time to the time of the previous hourly precipitation reset. This varies slightly by site. Values are in inches. This value may or may not contain frozen precipitation melted by some device on the sensor or estimated by some other means. Unfortunately, we do not know of an authoritative database denoting which station has which sensor.

alti: (Pressão do altímetro em polegadas) -> Pressure altimeter in inches

mslp: (Pressão do nível do mar em milibar) -> Sea Level Pressure in millibar

vsby: (Visibilidade em milhas) ->Visibility in miles

gust: (Velocidade do vento em nós) -> Wind Gust in knots

skyc1: (Nível 1 de cobertura do céu) -> Sky Level 1 Coverage

skyc2: Sky Level 2 Coverage

skyc3: Sky Level 3 Coverage

skyc4: Sky Level 4 Coverage

skyl1: (Altitude do nível de cobertura em pés) -> Sky Level 1 Altitude in feet

skyl2: Sky Level 2 Altitude in feet

skyl3: Sky Level 3 Altitude in feet

skyl4: Sky Level 4 Altitude in feet

wxcodes: (Códigos do Tempo presente separados por espaço) -> Present Weather Codes (space separated)

metar: (dados não processados no formato METAR) -> unprocessed reported observation in METAR format.

Todas as colunas: station,valid,lon,lat,tmpf, dwpf, relh, drct, sknt, p01i, alti, mslp, vsby, gust, skyc1, skyc2, skyc3, skyc4, skyl1, skyl2, skyl3, skyl4, wxcodes, ice_accretion_1hr, ice_accretion_3hr, ice_accretion_6hr, metar

2º Conjunto de dados

Dados retirados da Agência Nacional de Aviação Civil (ANAC) no Brasil de Janeiro de 2015 até Agosto de 2017 a partir de um projeto do Kaggle

Os dados estão no site do Kaggle e como ia ficar muito grande para enviar deixei o link para baixar:

<https://www.kaggle.com/ramirobentes/flights-in-brazil/home>

Voos,Companhia.Aerea,Codigo.Tipo.Linha,Partida.Prevista,Partida.Real,Chegada.Prevista,Chegada.Real,Situacao.Voo,Codigo.Justificativa,Aeroporto.Origem,Cidade.Origem,Estado.Origem,País.Origem,Aeroporto.Destino,Cidade.Destino,Estado.Destino,País.Destino,LongDest,LatDest,LongOrig,LatOrig

O primeiro conjunto de dados que contém as informações dos aeroportos será entrelaçado com o segundo que contém informações do status de cada voo para predição de atraso ou cancelamento de voos

Descrição da solução

Os dados de parâmetros do primeiro conjunto de dados (dados dos aeroportos a cada hora como temperatura) e os com status de cada voo desses mesmos aeroportos no mesmo horário serão utilizados sendo escolhidos inicialmente 8 aeroportos (Pela quantidade de voo e por ter estações do ano relativamente definidas: Guarulhos (SBGR), Congonhas (SBSP), Viracopos (SBKP) do Galeão (SBGL); Clima subtropical: o de Curitiba (SBCT), Porto Alegre (SBPA); por ter o clima tropical litorâneo: Recife(SBRF); Clima equatorial: de Manaus (SBMN):

Exemplo de identificador: SBKP, SBCT...

Para treinamento e classificação de 3 tipos de voos:

1. Os que poderiam decolar sem atrasos ocasionados pelo mau tempo
2. Os que podem atrasar
3. Os que tendem a ser cancelados

E fornecer uma estatística com os dados de teste de quantos acertos de atraso ou cancelamentos foram adquiridos a partir dessa base de dados treinada.

Os modelos de benchmark

Utilizar algoritmos como:

Árvores de decisão, random forest, Adaboost e KNN verificando o melhor modelo para minimizar os erros de treino e de teste.

<https://ieeexplore.ieee.org/document/7777956>

Métricas de avaliação

Taxa de erro (%) Precisão (%) Recall (%) seguindo as métricas de <http://cs229.stanford.edu/proj2012/CastilloLawson-PredictingFlightDelays.pdf> e adicionando o FBScore

Design do projeto:

