

Análise do Consumo de Combustível - Aerometrics

Modelagem Preditiva e Processamento de Dados

Levi Gurgel de Lima, Yves Gabriel Q. de Sousa, Marco Aurélio
C. Risardi

Introdução

A **Aerometrics** desenvolveu um sistema preditivo para estimar o **consumo de combustível** a partir de dados históricos de operações aéreas.

Nesta apresentação serão mostrados:

- ▶ Preparação e engenharia de atributos
- ▶ Análise exploratória
- ▶ Modelos preditivos implementados
- ▶ Combinação de modelos (*ensemble*)
- ▶ Previsões finais no conjunto de teste

Setup

```
library(tidyverse)
library(lubridate)
library(e1071)
library(neuralnet)
library(caret)
library(tree)

set.seed(123)

load("data_project_train.Rda")
load("data_project_test.Rda")
```

Estrutura de Dados

```
summary(data_train)
```

```
##      flight_id      orig      dest
##  Min.      :785929656  SBSP      :1239  SBGR      :1407
##  1st Qu.:797507106    SBGR      :1104  SBSP      :1377
##  Median :822383588    SBRJ      : 728  SBRJ      : 743
##  Mean    :821981147    SBBR      : 626  SBBR      : 697
##  3rd Qu.:848453636    SBSV      : 313  SBSV      : 311
##  Max.    :861779696    SBPA      : 302  SBCF      : 282
##                (Other):2967  (Other):2462
##      dep_time      arr_time
##  Min.      :2023-01-14 00:01:24.00  Min.      :2023-01-14 00
##  1st Qu.:2023-03-15 21:06:47.50    1st Qu.:2023-03-15 22
##  Median :2023-07-06 12:20:44.00    Median :2023-07-06 13
##  Mean    :2023-07-01 09:03:27.11    Mean    :2023-07-01 10
##  3rd Qu.:2023-10-25 00:28:12.00    3rd Qu.:2023-10-25 01
##  Max.    :2023-12-24 23:57:44.00    Max.    :2023-12-25 02
##
```

Feature Engineering

Nesta etapa realizamos a criação de novas variáveis derivadas das informações brutas do dataset.

A engenharia de atributos é fundamental para melhorar a representação dos dados e aumentar o poder preditivo dos modelos.

As principais transformações incluem:

- ▶ **Velocidade média (avg_speed):** estimada pela razão entre distância voada e tempo de voo.
- ▶ **Indicadores de eficiência:** métricas de consumo por milha náutica e por minuto.
- ▶ **Classificação do tipo de rota:** segmentação dos voos em *curto*, *médio* ou *longo alcance*.
- ▶ **Variáveis temporais:** extração da hora de partida e do período do dia (*manhã*, *tarde*, *noite*).
- ▶ **Índice total de ineficiência operacional:** combinação dos KPIs de decolagem e pouso.

Esses atributos aprimoram a capacidade dos modelos em capturar padrões relevantes associados ao consumo de combustível.

Feature Engineering - Código

```
df_train <- data_train %>%  
  mutate(  
    avg_speed = flown_distance_enr / (flight_duration / 60),  
    fuel_per_nm = fuel_burn / flown_distance_enr,  
    fuel_per_min = fuel_burn / flight_duration,  
    total_ineff = kpi_inefficiency_dep + kpi_inefficiency_a  
    flight_range = case_when(  
      flown_distance_enr < 400 ~ "Curto",  
      flown_distance_enr < 800 ~ "Médio",  
      TRUE ~ "Longo"  
    ),  
    flight_range = factor(flight_range, levels = c('Curto',  
    dep_hour = hour(dep_time),  
    dep_period = case_when(  
      dep_hour >= 5 & dep_hour < 12 ~ "Manhã",  
      dep_hour >= 12 & dep_hour < 18 ~ "Tarde",  
      TRUE ~ "Noite"  
    ),  
  ),
```