

Script 2 — 2_Config.R (Configuração do projeto)

Objetivo do script

- define **qual dataset** será usado e onde está localizado
- lê o dataset de forma **robusta** ao separador (, vs ;)
- cria a estrutura de pastas de outputs
- fixa a **seed** para reprodutibilidade

1) Leitura do dataset

```
csv_file_name <- "Dubai_data.csv"
csv_file <- file.path(project_dir, csv_file_name)
df_raw <- read_dataset_auto(csv_file)
```

Output

- cria `df_raw`: dataset "bruto".

2) Diretório principal dos outputs

```
result_dir <- file.path(project_dir, "result_dir")
```

3) Seed do projeto (reprodutibilidade)

```
seed <- 20260119
```

garante que

- split train/val/test é replicável
- folds de CV são replicáveis
- random forest / GBM (partes estocásticas) ficam o mais estáveis possível

4) Pastas de outputs por capítulo

```
out_cap3_tables <- file.path(result_dir, "outputs_tabelas_cap3")
out_cap3_figs   <- file.path(result_dir, "outputs_figuras_cap3")
out_cap4        <- file.path(result_dir, "outputs_cap4")
out_cap5        <- file.path(result_dir, "outputs_cap5")
out_cap6        <- file.path(result_dir, "outputs_cap6")
```

5) Criação das pastas

```
dir.create(..., recursive = TRUE, showWarnings = FALSE)
```

6) Controlos de execução (flags)

```
RUN_CAP3 <- TRUE
RUN_CAP4 <- TRUE
RUN_CAP5 <- TRUE
RUN_CAP6 <- TRUE
```

O que faz

Define "switches" para permitir executar a pipeline de forma parcial e orquestrada tal como está no main.R.