

SPRINT 4 - Melhorias dos modelos

In & Out Devs (Grupo 2)

Hiperparâmetros

HIPERPARÂMETROS SÃO PARÂMETROS SELECIONADOS EM QUE O VALOR É USADO PARA MELHORAR O APRENDIZADO DE MÁQUINA. O OBJETIVO DE SUA UTILIZAÇÃO É APERFEIÇOAR A PRECISÃO DOS MODELOS MANIPULADOS.



Hiperparâmetros

HIPERPARÂMETROS SÃO PARÂMETROS SELECIONADOS EM QUE O VALOR É USADO PARA MELHORAR O APRENDIZADO DE MÁQUINA. O OBJETIVO DE SUA UTILIZAÇÃO É APERFEIÇOAR A PRECISÃO DOS MODELOS MANIPULADOS.

Grid search

BUSCA SIMPLES EM
GRADE DENTRO DO
ESPAÇO DELIMITADO
PELOS PARÂMETROS

Random search

BUSCA ALEATORIA
DENTRO DO ESPAÇO
DELIMITADO PELOS
PARÂMETROS

Árvore de Decisão

RESULTADO SEM HIPERPARÂMETROS

Acurácia do treino: 1.0
Acurácia do teste: 0.97

GRID SEARCH

Acurácia do treino: 0.98
Acurácia do teste: 0.95

RANDOM SEARCH

Acurácia do treino: 0.98
Acurácia do teste: 0.96

KNN

RESULTADO SEM HIPERPARÂMETROS

Acurácia do treino: 0.80
Acurácia do teste: 0.74

GRID SEARCH

Acurácia do treino: 1.0
Acurácia do teste: 0.81

RANDOM SEARCH

Acurácia do treino: 0.86
Acurácia do teste: 0.76

Naive Bayes

RESULTADO SEM HIPERPARÂMETROS

Acurácia do treino: 0.77
Acurácia do teste: 0.60

GRID SEARCH

Acurácia do treino: 0.93
Acurácia do teste: 0.93

RANDOM SEARCH

Acurácia do treino: 0.93
Acurácia do teste: 0.93

SVM

RESULTADO SEM HIPERPARÂMETROS

Acurácia do treino: 0.63
Acurácia do teste: 0.64

GRID SEARCH

Acurácia do treino: 0.99
Acurácia do teste: 0.96

RANDOM SEARCH

Acurácia do treino: 0.99
Acurácia do teste: 0.96

Regressão Logística

RESULTADO SEM HIPERPARÂMETROS

Acurácia do treino: 0.98
Acurácia do teste: 0.97

GRID SEARCH

Acurácia do treino: 0.98
Acurácia do teste: 0.97

RANDOM SEARCH

Acurácia do treino: 1.0
Acurácia do teste: 0.74

Qual modelo foi escolhido?

Modelo Escolhido

**Árvore de
Decisão**

**Regressão
Logística**

- SVM possui grande gasto computacional, isto é, demora para obter o resultado.
- Demais modelos possuem resultados menos satisfatórios





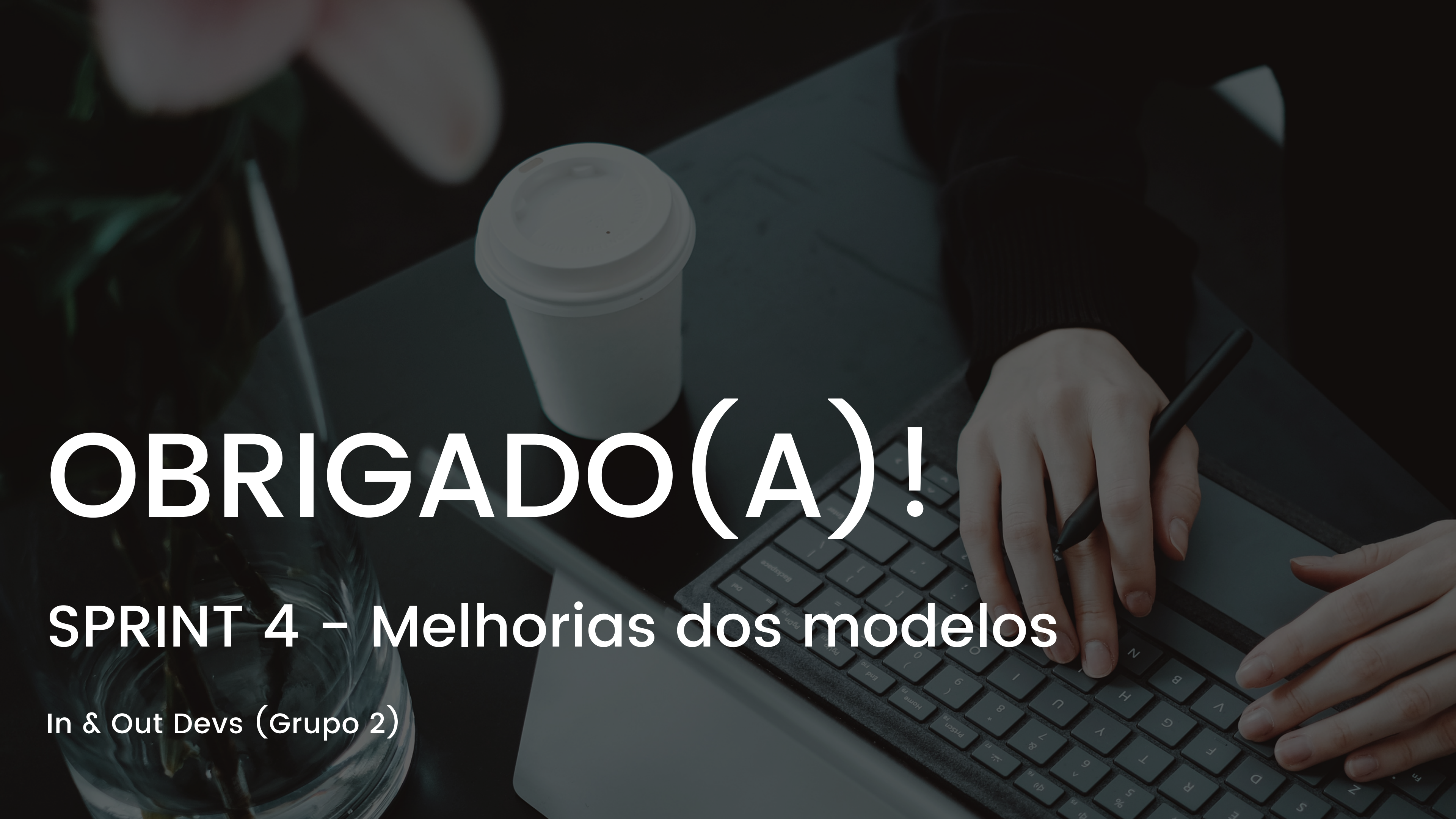
Random State

Árvore de decisão + Regressão Logística (42, 43, 44 e 45)

- Apresentam baixa variação;
- Média de variabilidade gira em torno de 1% a 2%;
- A partir das análises o que teve uma maior constância de valores foi o teste com o número 44;

Teste 44

- Maior balanceamento de dados;
- Altos valores de acurácia no treino e no teste;
- Seus parâmetros possuem uma média de 96% de acerto;



OBRIGADO(A)!

SPRINT 4 – Melhorias dos modelos

In & Out Devs (Grupo 2)