

## DECISION TREES - PARTE 3: REGRESSÃO

VAMOS ANALISAR COMO O ALGORITMO DE ARVORES DE DECISÃO CALCULA PROBLEMAS DE REGRESSÃO.

TEMPERATURA	DOMINGO	VENDAS
QUENTE	SIM	286
FRIO	NÃO	147
AMENO	NÃO	169
FRIO	SIM	172
AMENO	NÃO	176
QUENTE	NÃO	253
QUENTE	NÃO	238
FRIO	NÃO	151
FRIO	SIM	168
QUENTE	NÃO	264
AMENO	SIM	207
QUENTE	SIM	309
QUENTE	NÃO	245

FORMULA :  $\sum P(c) \cdot S(c)$   
UTILIZADA : CEX

#DESVIO PADRÃO DO VARIÁVEL TARGET

$$S_{VENDAS} = 52,35$$

#DESVIO PADRÃO PARA CADA UM DAS VARIÁVEIS + FORMULA

TEMP	S	AMOSTRAS
QUENTE	24,65	6
FRIO	10,69	4
AMENO	16,51	3

$$S_{TEMP} = \frac{24,65 \cdot 6}{13} + \frac{10,69 \cdot 4}{13} + \frac{16,51 \cdot 3}{13}$$

$$= 18,48$$

$$\Delta S_{TEMP} = 52,35 - 18,48$$

$$= 33,87$$



\_\_/\_\_/\_\_

S T Q Q S S D

OBS. O DESVIO PADRÃO (SD) É O QUANTO QUE OS VALORES ESTÃO DISTANTES DA MÉDIA. SE O DP FOR ALTO, OS VALORES ESTÃO DISTANTES DA MÉDIA, ENQUANTO QUE SE FO BAIXO, OS VALORES ESTÃO PRÓXIMOS DA MÉDIA.

VAMOS CONSIDERAR QUE OS VALORES DA VARIÁVEL VENDAS VARIAM ABSTAMENTE, COM UM ALTO DP, SENDO OS VALORES DISTANTES DA MÉDIA. MAS PORQUE ISSO OCORRE? PORQUE DA TEMPERATURA? NESTE SENTIDO, AO CALCULAR O DP PARA A TEMP QUENTE, PODÉ SER QUE ESSE VALOR SEJA BAIXO, UMA VEZ QUE ESTAMOS CONSIDERANDO APENAS A TEMP QUENTE E QUE TODOS OS SEUS VALORES VÃO SER ALTOS, ENQUANTO QUE PARA A TEMPERATURA FRIA TODOS OS VALORES VÃO SER BAIXOS. CONSIDERANDO QUE O DP DOS VALORES QUENTE É FRIA SERÃO MUITO BAIXOS, ENTÃO O DP TOTAL DA VARIÁVEL VAI SER PEQUENA. ISSO SIGNIFICA QUE AO SE DIVIDIR AS CATEGORIAS DESSA VARIÁVEL ELA EXPLICA MUITO BEM OS DADOS, OU SEJA, EXPLICA O COMPORTAMENTO DE NOSSA VARIÁVEL TARGET.

SE O DP DE UMA VARIÁVEL FOR BEM MENOR QUE O DA VARIÁVEL TARGET SIGNIFICA QUE ESSA VARIÁVEL SEJA UMA BOA VARIÁVEL PARA CLASSIFICAR OS DADOS POR ISSO ESTAMOS INTERESSADOS NA MELHOR RELAÇÃO DO DESVIO PADRÃO. QUANTO MENOR O DP DE UMA VARIÁVEL EM RELAÇÃO A VARIÁVEL TARGET, MELHOR VAI SER ESSA VARIÁVEL PARA A CRIAÇÃO DE NOSSA FUNÇÃO DE DECISÃO.

DOMINGO	S	AMOSTRAS
SIM	58,48	5
NÃO	49,94	8

$$S_{DOM} = \frac{58,48 \cdot 5}{13} + \frac{49,94 \cdot 8}{13}$$
$$= 50,77$$

$$DS_{DOM} = 52,55 - 50,77$$
$$= 1,78$$



A maior variação de DP foi da varável temperatura, portanto utilizaremos essa para início e criação de nosso árvore de decisão

Vamos estabelecer um critério de parada: É necessário ter mais de 3 amostras para continuar os cálculos.

→ Min amostras = 4

