



FORMAÇÃO INTELIGÊNCIA ARTIFICIAL E MACHINE LEARNING

MACHINE LEARNING – TÓPICOS AVANÇADOS
AVALIANDO A VARIABILIDADE EM UM MODELO

Prof. Fernando Amaral – Todos os Direitos Reservados

Cenário

- Seu cliente vende um produto on-line a preço fixo: R\$ 50,00
- São feitas 30 mil transações por mês
- 15% das transações são fraudulentas (4500)
- Faturamento mensal: R\$ 1.275.000,00
- Perdas com Fraude: R\$ 225.000,00

- Margem de lucro (5%): 63.750,00

- Posso reduzir minha perda com fraudes?



Cenário

Você coleta dados, faz testes e consegue criar um modelo com acerto em 90%

- Proposta:
 - Reduzir fraude de 15% para 10%
 - Faturamento mensal: de R\$ 1.275.000,00 para 1.350.000,00
 - Perdas com Fraude: de R\$ 225.000,00 para 150.000,00
- Custo do projeto: R\$ 300.000,00
- ROI em apenas 4 meses!



Questão:

É seguro afirmar um acerto de 90%?

Não, porque estamos trabalhando com amostras, e estas estão sujeitas a variabilidade

Agravante: Modelos de negócio mudam rapidamente, e isso pode influenciar o seu resultado, as vezes de forma mais rápida do que sua atualização do modelo



Como fazer

Prever uma variabilidade no resultado

De que forma: intervalos de confiança

- Intervalo de Confiança para a média (rodando várias vezes e usando a média), ou
- Intervalo de Confiança para a proporção (rodando o modelo uma única vez)



Intervalo de Confiança para a Média

100 modelos criados – amostra (n)

Intervalo de Confiança: 95%

Valor de Z = 1,96

Desvio padrão: 12,61

Média: 88,2

O acerto médio do modelo de Machine Learning deve variar entre 85,7 e 90,67, com um nível de confiança de 95%

$$\bar{X} \pm z * \left(\frac{\sigma}{\sqrt{n}} \right)$$

$$\bar{X} \pm 1,96 * \left(\frac{12,61}{\sqrt{100}} \right)$$

$$\bar{X} \pm 2,47$$

Os dados de treino dos modelos devem ser escolhidos por funções que deem aos mesmos as mesmas chances de serem selecionados!



Intervalo de Confiança para a Proporção

1000 registros (n) (dados de treino)

Intervalo de confiança: 95%

Valor de Z = 1,96

P(a) = 0,9

P(e) = 0,1

$$\hat{p} \pm z * \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$\hat{p} \pm 1,96 * \sqrt{\frac{0,9(1 - 0,9)}{1000}}$$

$$\hat{p} \pm 0,02$$

A proporção de acertos esperado da aplicação de Machine Learning na detecção de fraudes, é entre 88 e 92%, com um nível de confiança de 95%

