

MBA⁺

Artificial Intelligence &
Machine Learning





Métricas de Machine Learning

Resultados publicados no mercado Brasileiro



>> Finanças

Chatbot realiza 70% dos atendimentos do BB em rede social

A acurácia, ou seja, o percentual de respostas corretas às perguntas dos clientes, é de 77%. O mercado considera muito bom uma acurácia entre 70 e 80%. Quando é necessário algum tipo de interação humana, o atendimento é encaminhado para funcionários do BB que atuam no atendimento em redes sociais.

Vivi: bot da Vivo faz 600 mil atendimentos por mês com 94% de assertividade

Bradesco passa a usar ferramenta de inteligência artificial para esclarecer dúvidas de clientes

Segundo Marcelo Câmara, gerente de inovação, até agora o BIA “aprendeu” a formular 4 mil respostas a 600 mil perguntas sobre 60 produtos, com 96% de assertividade. Em um projeto sobre plataforma Watson (o primeiro grande uso de nuvem pública do banco), o foco até agora são nos algoritmos de atendimento assistido por IA, sem grande carga de dados ou serviços transacionais.

Resultados publicados no mercado Brasileiro



>> Finanças

Chatbot realiza 70% dos atendimentos do BB em rede social

A **acurácia**, ou seja, o percentual de respostas corretas às perguntas dos clientes, é de 77%. O mercado considera muito bom uma acurácia entre 70 e 80%. Quando é necessário algum tipo de interação humana, o atendimento é encaminhado para funcionários do BB que atuam no atendimento em redes sociais.

Vivi: bot da Vivo faz 600 mil atendimentos por mês com **94% de assertividade**

Bradesco passa a usar ferramenta de inteligência artificial para esclarecer dúvidas de clientes

Segundo **Marcelo Câmara**, gerente de inovação, até agora o BIA **“aprendeu” a formular 4 mil respostas** a 600 mil perguntas sobre 60 produtos, com **96% de assertividade**. Em um projeto sobre plataforma Watson (o primeiro grande uso de nuvem pública do banco), o foco até agora são nos algoritmos de atendimento assistido por IA, sem grande carga de dados ou serviços transacionais.



O que isso significa?_

Esses resultados são confiáveis?_



Cálculo de métricas

Acurácia ou precisão geral do modelo

$$\frac{Acertos(A)}{Acertos(A) + Erros(E)}$$

Média das acurácias encontradas para cada intenção

$$\frac{1}{n} \sum_{i=1}^n AGCl(i) = \frac{1}{n} \sum_{i=1}^n \frac{Acertos(A_i)}{Acertos(A_i) + Erros(E_i)}$$

Calculo de métricas



Imagine que tenhamos um classificador binário treinado para detectar spam e obtivemos os seguintes resultados:

	Classificado como Spam	Classificado como normal
Spam	100	150
Normal	50	1000

Acurácia = $(100+1000)/1300 = 84.6\%$. O que representa um valor muito bom! **Mas e se trocássemos esse classificador por um classificador burro?** Que independente da entrada sempre retornaria como e-mail normal (sem spam):

-

	Classificado como Spam	Classificado como normal
Spam	0	0
Normal	150	1150

Nossa acurácia que antes era de quase 85% subiria para 88.5%! Ou seja, mesmo que o nosso classificador não esteja realizando nenhuma análise/previsão ele ainda é melhor que o anterior.

Calculo de métricas – Matriz de Confusão

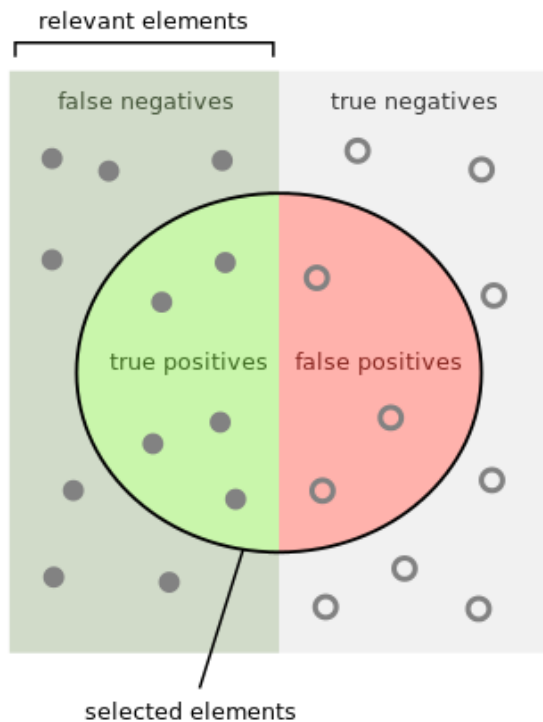


A matriz de confusão nos permite realizar uma análise mais detalhada da situação do nosso classificador uma vez que ela distingue nossos resultados em quatro classes.

Usando o exemplo de um classificador binário (classificar se um e-mail é ou não spam):

- **Verdadeiros Positivos (VP):** casos em que retornamos a classe SPAM e o e-mail realmente era de spam.
- **Falsos positivos (FP):** casos em que retornamos a classe SPAM e na verdade eram e-mails comuns.
- **Falsos Verdadeiros (FV):** casos que retornamos que era e-mail comum (ausência de spam) e realmente eram.
- **Falsos Negativos (FN):** retornamos que não eram spam e na verdade eram spam.

Calculo de métricas – Matriz de Confusão



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

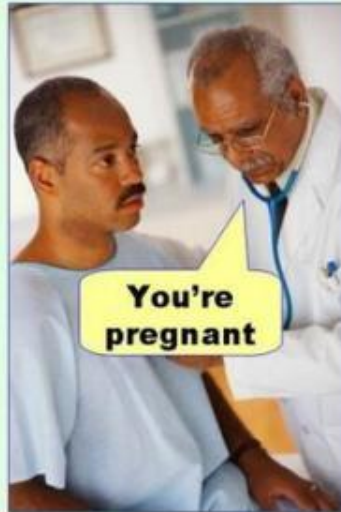
How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

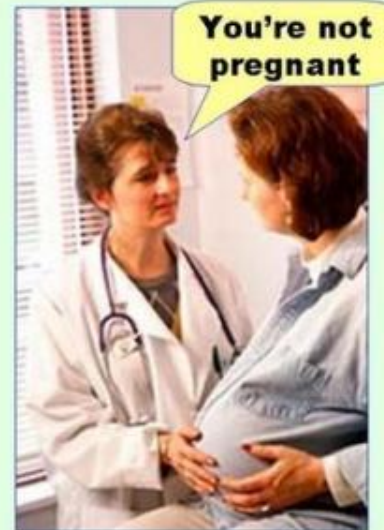
Calculo de métricas – Matriz de Confusão



Type I error
(false positive)



Type II error
(false negative)



Calculo de métricas – Matriz de Confusão



Vamos assumir que nosso ChatBot de uma pizzeria possui três classes (intenções): #realizarPedido, #obterStatus e #horarioAtendimento e nelas possuímos, respectivamente, dez, nove e doze exemplos de interações.

No caso do ChatBot em que temos múltiplas classes podemos calcular essas métricas da seguinte forma:

- Verdadeiros Positivos e Verdadeiros Negativos fazem compõem um mesmo conjunto (pintados de roxo na tabela abaixo)
- Falsos positivos ficarão na última linha da tabela enquanto os falsos negativos irão para a última coluna a direita, podemos ver que a soma destes conjuntos é a mesma.

		Previsto			FN
		realizarPedido	obterStatus	horarioAtendimento	
Classe Real	realizarPedido	6	4	0	4
	obterStatus	3	4	2	5
	horarioAtendimento	0	1	11	1
FP		3	5	2	10

Cálculo de métricas



Precisão

$$Precisão = \frac{Verdadeiros\ Positivos\ (TP)}{Verdadeiros\ Positivos\ (TP) + Falsos\ Positivos\ (FP)}$$

Recall

$$Recall = \frac{Verdadeiros\ Positivos\ (TP)}{Verdadeiros\ Positivos\ (TP) + Falsos\ Negativos\ (FN)}$$

F1 Score

$$F1 = \frac{2 * precisão * recall}{precisão + recall}$$

Resultados do Estado da Arte



AAAI Conference on Artificial Intelligence Copyright c 2017

Multi-Task Deep Learning for User Intention Understanding in Speech Interaction Systems

Models	Intention Prominence		
	Precision	Recall	F1-measure
SVM	0.627	0.618	0.621
BN	0.797	0.789	0.791
CRF	0.769	0.754	0.761
LSTM	0.792	0.803	0.797
LSTM+BN	0.868	0.865	0.866

Outras Métricas



G-Measure

Curva ROC

Grid Search

KS (Kolmogorov-Smirnov)

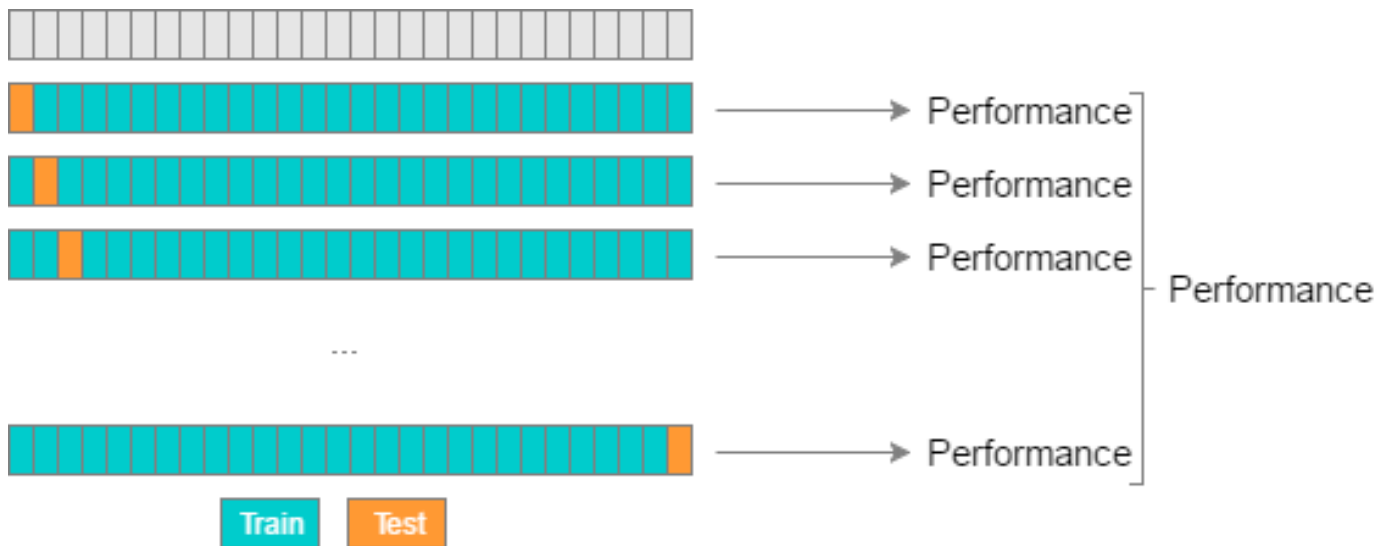
Area Under Curve (AUC)

Gini



Formas de Validação

Leave one out



K Fold



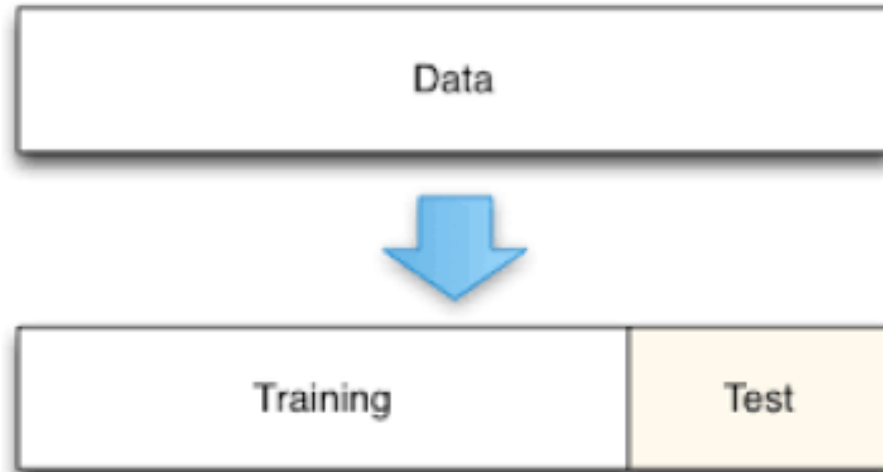
- The dataset is partitioned into K equal sized samples

5-fold CV

DATASET

Estimation 1	Test	Train	Train	Train	Train
Estimation 2	Train	Test	Train	Train	Train
Estimation 3	Train	Train	Test	Train	Train
Estimation 4	Train	Train	Train	Test	Train
Estimation 5	Train	Train	Train	Train	Test

Holdout



Questions and Feedback



[Thank you!](#)

Obrigado !



Vinicius Fernandes Caridá
vfcarida@gmail.com



@Vinicius Caridá



@Vinicius Caridá
@machine learning Brasil



@vfcarida

MBA⁺

Copyright © 2018 **Prof. Vinicius Fernandes Caridá**
 Todos direitos reservados. Reprodução ou divulgação
 total ou parcial deste documento é expressamente
 proibido sem o consentimento formal, por escrito, do
 Professor (autor).