

## **Aplicabilidade do Teorema de Bayes no Monitoramento de Redes Sociais**

### **Resumo**

O teorema de Bayes tem se mostrado como uma ferramenta importante no auxílio à tomada de decisão. O site Causa Brasil monitora as questões e reivindicações nas redes sociais, é um site atualizado a cada hora e calcula os percentuais de cada causa baseado nas últimas 24h, contém um histórico de 55 dias. O *clipping* da Radiobrás divulga as principais notícias das mídias com maior impacto no Brasil. O teorema de Bayes calcula incertezas, fazendo previsões, baseado na experiência do pesquisador ou no histórico da situação. Utilizando o site Causa Brasil e relacionando-o com as notícias divulgadas no *clipping* da Radiobrás diariamente, criou-se um histórico para verificar se existe alguma relação entre as notícias da mídia com as causas das redes sociais. O objetivo desse estudo foi criar bases técnicas para fazer previsões da repercussão de fatos, utilizando o teorema de Bayes, nas redes sociais com as notícias ou acontecimentos. O teorema de Bayes é explicado e exemplificado ao longo do texto. Os resultados não foram significativos e não se pode chegar a conclusões quanto as possíveis probabilidades para o uso do teorema de Bayes. No entanto, desenvolveu-se uma metodologia para, com mais dados, futuramente utilizar o teorema de Bayes.

### **Introdução**

As redes sociais estão cada vez mais presentes no cotidiano e tem sido uma importante ferramenta de comunicação. Elas tem provado ser um meio muito poderoso de difusão e mobilização social, visto os recentes acontecimentos na mobilização dos protestos. O site Causa Brasil nasceu da ideia de verificar as causas e reivindicações que estão acontecendo na internet. Dessa forma, o site tem uma plataforma que monitora o movimento nas redes sociais detalhando quais manifestações mais foram feitas na últimas 24 horas. Assim, o objetivo deste artigo foi compreender se existe alguma maneira de mensurar, de forma a quantificar ou prever o impacto de acontecimentos antes de manifestações nas redes sociais acontecerem. Ao tratar de previsão estatística, o modelo que mais se encaixa neste tipo de comportamento é o teorema de Bayes. O Teorema de Bayes é uma ferramenta que auxilia na análise de decisão e tem sido usado na tomada de decisão. (MICHAELI, 2007). Ele trabalha com incertezas, contudo este teorema ainda é pouco usado apesar de ter uma matemática simples, pois a inserção dos dados não acontece de maneira tão trivial. O Teorema de Bayes utiliza incertezas do ambiente e para fazer a sua correta aplicação é necessário ter experiência no assunto ou ter dados que ajudem a formular as previsões.

Assim, nesse artigo utilizaram-se três tipos de informação: o fato, a notícia e a repercussão. Dessa forma, os três tipos foram caracterizados de acordo com a seguinte regra: o fato que ocorre quando algo acontece, a notícia é quando o fato é divulgado pela imprensa e a repercussão, que é o impacto da notícia na sociedade. Entretanto, percebe-se que no Brasil, as informações da internet excluem de certa forma uma boa parte da população, assim como a utilização de outros meios de acesso à informação.

### **Fundamentação teórica**

O teorema de Bayes relaciona informações, com a probabilidade de ocorrência, para gerar uma nova probabilidade quando os fatos acontecem de maneira relacionada ou são dependentes. Também é importante ressaltar que essas probabilidades podem e devem ser revistas à medida que são observados novos fatos que podem alterar a probabilidade de cada um dos acontecimentos envolvidos (SILVER, 2013). O teorema de Bayes utiliza incertezas, entretanto é alimentado por meio de probabilidades que podem ser estimas ou sugeridas.

Assim, o teorema de Bayes trabalha primeiramente com informações prévias, dessa forma, ao obter novas informações aplica-se o teorema e alcançam-se as probabilidades posteriores.

O teorema de Bayes utiliza dois tipos de inferências estatísticas: as intuitivas e as experimentais. A primeira é adquirida através do conhecimento prévio em relação a situações passadas, desta maneira pode-se fazer uma conjectura sobre a situação e formular uma possível probabilidade. Já a outra é obtida à medida através da experimentação, a partir dos dados são feitos procedimentos estatísticos com os quais se calcula a probabilidade, à medida que novos dados coletados são refeitos os cálculos estatísticos e calculadas as novas probabilidades. Assim, as probabilidades associadas a um evento são definidas previamente e atualizadas na medida em que se recebem novas informações. Quando os dados são mais informativos que as suposições iniciais, ou seja, quando a amostra aumenta, as probabilidades convergem independentes das distribuições iniciais. (SILVER, 2013)

Nesse sentido, Wheaton (2009), explica que o teorema de Bayes é muito útil quando se treina novos profissionais, pois ele auxilia a minimizar a diferença da experiência entre os profissionais antigos e os novos. Assim, pode-se quantificar quão provável é algo de acontecer baseado em informações prévias. No entanto o Teorema de Bayes não é usualmente utilizado por parecer muito complexo ou porque os analistas não o conhecem e não sabem como usá-lo.

O teorema de Bayes é utilizado por profissionais de inteligência. Atualmente é utilizado principalmente em problemas psicológicos, pois trabalha com incertezas condicionais computadas explicitamente, não estimadas intuitivamente, porque ao estimar-se intuitivamente, as crenças podem levar ao erro. (MICHAELI, 2007).

Wheaton (2009) explica que o teorema de Bayes é o fundamento da inteligência competitiva. Pode ser aplicado na análise de decisão, assim quando um produto é lançado, é necessária a formatação correta e adequada do teorema de Bayes para que este mostre probabilidades de modo condizente. Em situações de decisão é altamente recomendável o uso do teorema de Bayes, pois ele contribui para geração do cenário em conjunto com as probabilidades.

Para Silver (2013), ao se tratar de problemas buscam-se alternativas conhecidas para solucioná-las, porém também existe o caso em que as alternativas desconhecidas não são imaginadas, por isso muitas vezes não são questionadas outras hipóteses. Porém, dessa maneira podem-se estimar probabilidades de maneira errada, pois se descartam hipóteses por simplesmente desconhecer sua existência.

Silver (2013) complementa que um problema grave é que nem sempre são mostradas as incertezas de um cálculo, levando as pessoas a pensarem que o número bruto é o valor real e ele acontecerá desta maneira. Para Wheaton (2009) o teorema de Bayes é algo que exige treinamento, pois se necessita estipular as probabilidades iniciais e quanto mais familiarizados com o assunto mais precisará esta avaliação. Assim, para melhor compreensão do Teorema de Bayes é preciso compreender como o mesmo é calculado.

### **Definição – teorema de Bayes**

Apesar das probabilidades  $P(A|B)$  e  $P(B|A)$  serem parecidas, elas significam algo diferente cada uma. Sendo  $P(A|B)$  a probabilidade de uma loja especializada em uma marca de pneus prestarem um bom serviço dentro da garantia. Então,  $P(B|A)$  é a probabilidade de uma loja de pneus que prestar um bons serviços dentro da garantia serem pneus especializada em uma marca de pneus. No primeiro momento quando lemos muitas vezes não percebemos a diferença entre as duas afirmações, porém probabilisticamente são diferentes, pois na primeira selecionamos uma loja especializada em uma marca de pneus e depois dentre estas as que prestam bom serviço dentro da garantia; na segunda afirmação, selecionamos as lojas que prestam um bom serviço e depois dentre essas, as que são especializadas em uma marca de

pneu. A inversão das afirmações faz a probabilidade estatística ser completamente diferente, pois isso devemos cuidar muito bem no momento em que fazemos a formulação estatística.

Definição teórica do teorema de Bayes - O teorema de Bayes permite que o evento B tenha um número qualquer de categorias ( $B_1, B_2, \dots, B_n$ ) sendo este evento mutuamente excludentes.

$$P(B_i | A) = \frac{P(A | B_i)P(B_i)}{P(A | B_1)P(B_1) + P(A | B_2)P(B_2) + \dots + P(A | B_n)P(B_n)}$$

### Exemplo simples do teorema de Bayes

Suponha que 60% das mulheres que compram kits de gravidez instantâneos estão, de fato, grávidas. Para um kit de uma marca específica, se a mulher estiver grávida, o teste fornecerá resultado positivo 96% das vezes e negativo 4% das vezes (um “falso negativo”). Se ela não estiver grávida, o teste resultará positivo em 1% das vezes (um “falso positivo”) e negativo 99% das vezes (figura 1). Suponha que um teste resulte positivo. Qual a probabilidade de que a mulher esteja realmente grávida? (ANDERSON, 2003)

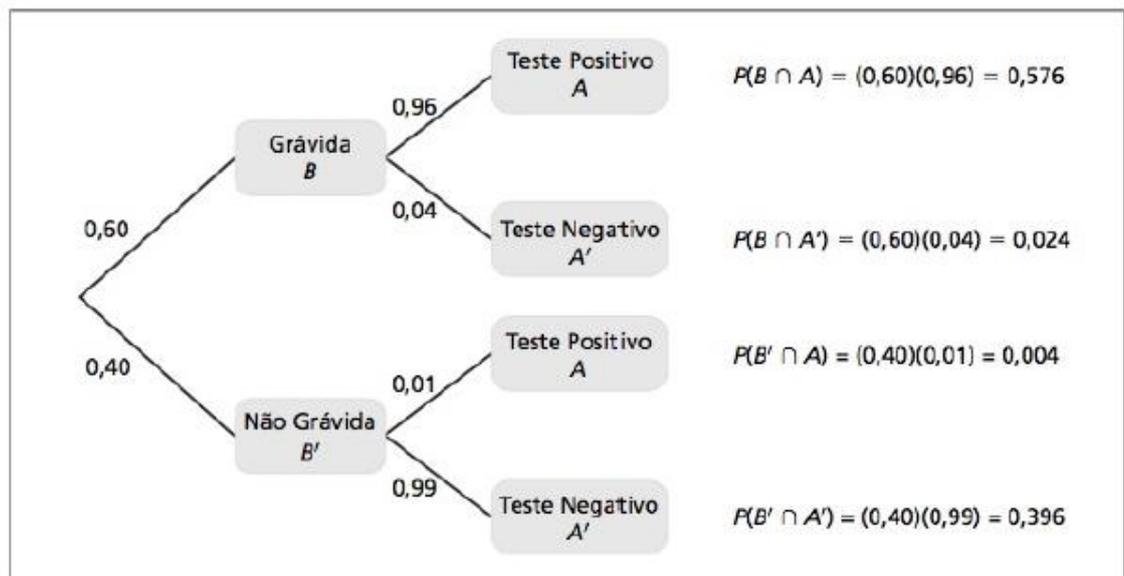


Figura 1. Diagrama de probabilidades do teste de gravidez.

Fonte: (ANDERSON, 2003)

$P(B|A)$  – probabilidade de grávida e o teste ter dado positivo

$P(A|B)$  – probabilidade de o teste ter dado positivo e estar grávida

$P(B)$  – probabilidade de estar grávida

$$P(B | A) = \frac{P(B)P(A | B)}{P(B)P(A | B) + P(B')P(A | B')}$$

$$P(A | I) = \frac{0,96 * 0,60}{0,96 * 0,60 + 0,01 * 0,40}$$

## **Método**

Foram coletados os dados de repercussão na rede social do site Causa Brasil ([www.causabrasil.com.br](http://www.causabrasil.com.br)) durante o período de tempo de um mês, entre 16 de junho e 16 de julho. A coleta era feita diariamente com intervalo de 24 horas, o horário de referência era à meia noite. Foram escolhidos alguns tópicos para serem monitorados. No momento da coleta os valores dos percentuais cada tópico e valores absolutos do dia eram passados para uma planilha e disponibilizados para análise.

Os dados referentes ao acontecimento foram retirados do site da Radiobrás que tem um *clipping* ([clipping.radiobras.gov.br](http://clipping.radiobras.gov.br)) dedicado as principais notícias dos jornais de maior impacto no país, este site é de uma empresa estatal do governo federal. Neste site foram analisadas as notícias diariamente, estas eram separadas de acordo com o termo de interesse e posteriormente foi dado um peso para cada tipo de notícia de acordo com o termo pesquisado por relevância atribuída pelo pesquisador. Também foi ponderado com um fator de impacto da mídia na população, este fator foi determinado pelo pesquisador. Foram atribuídos pesos de 0 a 10 de acordo com a significância da notícia e a significância da mídia.

As notícias ponderadas pela sua relevância foram combinadas e graficadas diariamente, resultando claramente em um impacto desse tópico ao longo do período estudado.

A análise dos dados foi feita comparando diretamente os dados de repercussão absolutos ou relativos e os dados das notícias ponderadas.

## **Aplicação o teorema de Bayes**

Quando se tem um problema com duas alternativas possíveis, primeiro pode-se ponderar que cada uma tem 50% de chance de acontecer, como se fosse um jogo de cara e coroa. Porém nem sempre é assim que as coisas funcionam, isso porque mesmo duas soluções com 50% de chance de acontecer, podem ter pesos diferentes de acordo com o problema. No caso de um governo de um país, restringindo as hipóteses das pessoas que aprovam e das pessoas que não aprovam, podemos ter, por exemplo, 30% e 70%, mas estes dados podem alterar de acordo com as notícias e acontecimentos no país. Por isso, é importante ressaltar que as probabilidades devem ser revistas à medida que as informações são alteradas ou chegam novas informações. Outro ponto importante é que como as informações circulam de maneira muito rápida podem ocorrer ruídos que podem afetar o correto discernimento sobre o problema. Neste estudo, o ruído considera as informações que podem parecer verdadeiras, porém nada tem haver com o sinal (SILVER, 2013).

O site Causa Brasil acompanha o movimento das informações nas redes sociais da internet, relacionando termos a alguns tópicos principais. As causas são identificadas a partir do conteúdo de postagens que contenham ao menos uma das principais hashtags ligadas às manifestações, segundo o site [causabrasil](http://causabrasil.com.br). Esses movimentos são atualizados de hora em hora a porcentagem de cada causa é referente ao número acumulado de menções no período de 24 horas segundo o site [causabrasil](http://causabrasil.com.br). São atribuídos percentuais de acordo com a presença dos termos nas redes sociais. Assim, temo-se uma noção do que está acontecendo nas redes sociais, mas com retardo de um dia, pois precisam de 24 horas para atualizar a ocorrência dos termos nas redes sociais. É importante ressaltar que o site mede o fluxo de informações sobre determinados termos nas redes sociais, não se preocupando com as notícias, somente com a repercussão, porém a repercussão acontece por causa das notícias. As causas são categorizadas por macrotemas: político, economia, direitos básicos do cidadão, liberdades individuais e copa do Brasil ([www.causabrasil.com.br](http://www.causabrasil.com.br)). Cada causa tem diversos termos e *hashtags* que são combinados a fim de construir o percentual de manifestações do assunto.. A





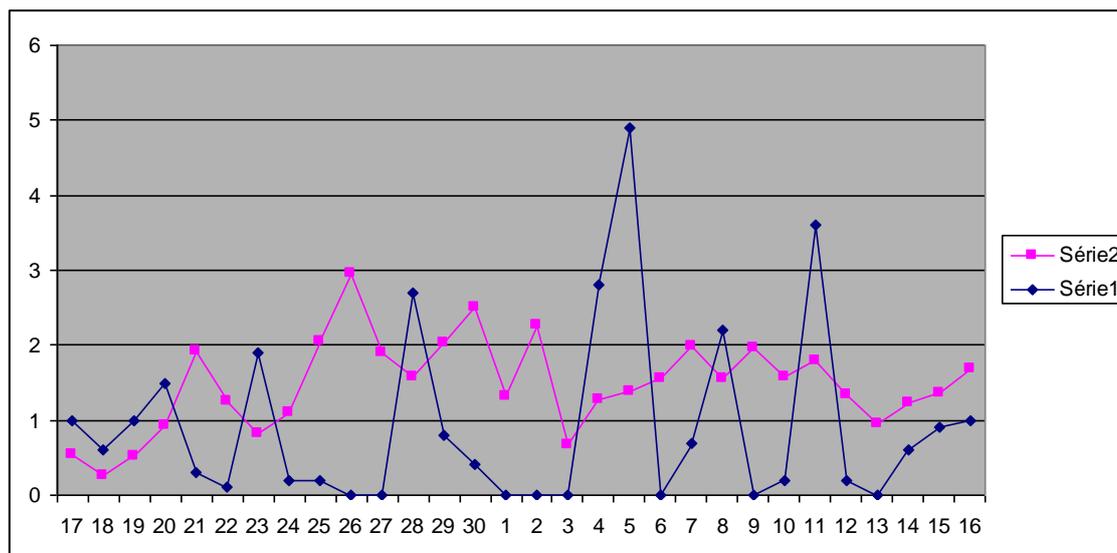


Figura 4. Comparação entre a causa “Combate à inflação” no site CausaBrasil (série 2) e o termo “Inflação” no clipping da Radiobrás (série 1).

Fonte: Elaboração própria

### Conclusão

Comparando os dados do site da Radiobrás com os dados do site da Causa Brasil não foi possível à obtenção de nenhum resultado definitivo fundamentar a aplicação do teorema de Bayes, pois não se conseguiu uma validação entre os dois. Porém com a verificação durante um tempo maior, possivelmente, seja mais clara para esta comparação. Também é importante verificar que alguns tópicos podem ter maior repercussão social que outros devido à politização da população.

É importante ressaltar que o teorema de Bayes é mais eficaz quando é utilizada uma série de dados histórica para embasar as previsões. Por isso é importante continuar fazendo o acompanhamento e corrigir os possíveis erros de estruturação da metodologia aplicada, pois pequenos erros podem se propagar de maneira se tornarem grandes erros quando aplicamos o teorema de Bayes.

A comparação direta com o outra forma de obtenção de tendências permite a validação e compreensão do comportamento do teorema de Bayes, a maneira como foi conduzido este estudo pode ter influenciado no resultado, pois a ponderação e a generalidade do tema influenciaram no resultado.

### Referências

- ALBRECHT, K. Um Modelo de Inteligência Organizacional. HSM Management 44 maio-junho 2004.
- ANDERSON, D.R.;SWEENEY,D.J., WILLIMAS,T.A. Estatística Aplicada à Administração e à Economia. São Paulo: Pioneira Thonson Learning, 2003. 642 p.
- ARIELY, D. O Fim da Economia Racional. Harward Business Review, Jul. 2009.
- BARTON, D., COURT, D. Faça a Analítica Avançada Trbalhar para sua Empresa. Harward Business Review, Out. 2012.
- BLANK, M. , JANISSEK-MUNIZ, R. Inteligência Estratégica Antecipativa Coletiva e Crowdfunding: aplicação do método L.E.SCanning em empresa social de economia P2P. XXXVI Encontro da ANPAD. Rio de Janeiro,2012.
- BRYNJOLFSSON,E.; HITT,L.M.; KIM,H.H. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? (April 22, 2011). Available at SSRN: <http://ssrn.com/abstract=1819486> or <http://dx.doi.org/10.2139/ssrn.1819486>

BUCHANAN, L e O'CONNELL, Uma breve história da tomada de decisão. IN. Harvard Business Review. P.: 20-29, Jan. 2006.

CARON-FASAN, Marie-Laurence; JANISSEK-MUNIZ, Raquel. Análise de informações de inteligência estratégica antecipativa coletiva: proposição de um método, caso aplicado e experiências. Revista de Administração, São Paulo, v. 39, n. 3, p. 205-219, jul./ago./set. 2004.

CHARAN R. Guerra contra a indecisão. IN. Harvard Business Review. Página: 76-83, Jan. 2006

DAVENPORT, T.H. , PATIL, D.J. Cientista de Dados: O Profissional mais Cobiçado do Século. Harvard Business Review, Out. 2012.

DOANE, D.P.;SEWARD,L.E. Estatística Aplicada à Administração e à Economia – Dados Eletrônicos.Porto Alegre: AMGH, 2011. 839 p.

DRUCKER, Peter F. Gerenciando a si mesmo. Harvard Business Review, São Paulo, v. 83, n. 1, p. 89-97, jan. 2005.

FACHINELLI, A.C., RECH, J., MATTIA, O.M., ROVEDA, V. Vigília e inteligência Estratégica: Ferramentas Metodológicas para definição de estratégias de Relações Públicas. Conexão – Comunicação e Cultura, UCS, Caxias do Sul, v.6, n.11, jan./jun. 2007.

FERNANDES, F.C. Inteligência e Gestão Estratégica: Uma Relação Sinérgica. Revista Brasileira de Inteligência. Brasília:Abin, n.7,jul.2012.

FREUND,J.E.;SIMON,G.A. Estatística Aplicada Economia, Administração e Contabilidade. Porto Alegre: Bookman, 2000. 404 p.

HAMMOND S. J.; KEENEY, R e RAIFFA H. Armadilhas ocultas na tomada de decisão. Harvard Business Review. P. 84-92, Jan. 2006.

KANTER, R.M. O que Diria Drucker. Harvard Business Review, Nov. 2009.

MARCHAND, D., PEPPARD, J. Por que a TI se Atrapalha com a Analítica. Harvard Business Review, Fev. 2013.

McAFEE, A. , BRYNJOLFSSON, E. Big Data: A Revolução da Gestão. Harvard Business Review, Out. 2012.

MICHAELI,R.;SIMON,L. (2007). An illustration of Bayes' theorem and its use as a decision-making aid for competitive intelligence and marketing analysts. European Journal of Marketing. v. 42, n.7/8, p. 804-813, 2008.

OHL, M. Como Tomar Decisões Melhores. Revista VoceSA. Mai. 2013.

PETRY, A. O Berço do Big Data. Revista Veja, 15 de maio, 2013.

STAUFFER, David. Bem-vindo ao mundo da inteligência competitiva. HSM Management Update, São Paulo, n. 10, p. 5-7, jan.-fev. 2004.

SILVER, NATE. O sinal e o ruído: porque tantas previsões falham e outras não. Rio de Janeiro: Intrinsic,2013. 544p.

CAUSA BRASIL. Disponível em < <http://www.causabrasil.com.br> >. Acesso em: 17/07/2013

CLIPPING RADIOBRAS. Disponível em: <[http:// clipping.radiobras.gov.br](http://clipping.radiobras.gov.br)>. Acesso em: 17/07/2013

TEIXEIRA. D.R. Inteligência de Mercado: Transformando as Olimpíadas em Negócios. Revista Inteligência Competitiva,São Paulo, v.2, n.1, p15-21, jul./set. 2012.

THOLT,C. Decida com Inteligência. Brasília, thesaurus, ABRAIC, 2006.

VALENTIM,M.L.P. et al. O processo de inteligência competitiva em organizações. *DataGramaZero - Revista de Ciência da Informação*. v. 3, n. 4, 2002.

WHEATON,K.J.;LEE,J.;DESHMUKH,H.Teaching Bayesian statistics to intelligence analysts: lessons learned. Journal of Strategic Security. v. 2, n.1/3, p. 39-58, 2009.