

Detectando misticismo quântico em livros publicados no Brasil com Ciência de Dados⁺*

*Mairus Disconzi de Moura*¹

Secretaria Municipal de Educação

Porto Alegre – RS

*Renato P. dos Santos*²

Universidade Luterana do Brasil

Canoas – RS

Resumo

As livrarias têm sido inundadas por títulos incluindo a palavra “quântica”, mas de caráter do chamado “misticismo quântico”, o qual propõe a existência de ligações entre a mecânica quântica e o misticismo oriental, levando a sérios equívocos. Este trabalho tem como objetivo identificar termos que possam ajudar os leitores, especialmente professores do ensino médio, a reconhecer a qual dessas categorias um livro pertence, antes mesmo de lê-lo, ajudando a quebrar o círculo vicioso de estudantes que aprendem pseudociência e a transmitem como verdade. Os catálogos on-line das quatro maiores livrarias do Brasil foram pesquisados por livros que contêm as palavras “quântica” ou “quântico” em seus títulos ou em suas sinopses. O banco de dados resultante foi examinado com a ajuda de técnicas de ciência de dados e recursos de mineração de dados textuais da linguagem de programação R. Foram identificados 22 termos que discriminam com uma acurácia de 94% entre as categorias “misticismo quântico” e popularização científica ou ciência.

Palavras-chave: *Ensino de Física; Física quântica; Misticismo quântico; Aprendizado de máquina; Livros.*

⁺ Detecting quantum mysticism in books published in Brazil with Data Science

^{*} *Recebido: dezembro de 2016.*

Aceito: julho de 2017.

¹ Email: mairusdm@gmail.com, ² Email: renatopsantos@ulbra.edu.br

Abstract

The bookstores have been full of titles including the word “quantum”, but more regarding to the “quantum mysticism”, which supports the existence of links between quantum mechanics and Eastern mysticism, leading to serious misunderstandings. This work aims to identify terms that can help readers, especially high-school teachers, to recognize to which of those categories a book pertains even before reading it through, contributing to breaking the vicious circle of students learning pseudoscience and passing it on as truth. The online catalogues from the four biggest bookstores in Brazil were used in this research. Books containing the words “quântica” or “quântico” – both corresponding to “quantum” – in their titles or synopses were searched. The resulting database was scrutinized with the help of Data Science techniques and text mining resources from the R programming language, and 22 terms that discriminate with an accuracy of 94% between the categories “quantum mysticism” and science or science popularization were identified.

Keywords: *Physics Teaching; Quantum Physics; Quantum Mysticism; Machine Learning; Books.*

I. Introdução

O presente analfabetismo científico fica evidente, quando se consideram, por exemplo, as atuais negações das mudanças climáticas, os movimentos contra vacinação, as recorrentes profecias de fim de mundo e as discussões, em conselhos escolares, sobre a inclusão do criacionismo nos currículos de ciências (ZIMMERMAN; CROKER, 2014).

Em um país como o nosso, com sérias deficiências de escolaridade, indicadas pelos resultados no PISA, SAEB, IDEB e INAF², e uma cultura de abertura significativa para espiritualidade, novas religiões, profetas, gurus e magos (PIERUCCI, 2012), em que os jovens que se autotransformam como “sem religião”, afirmam acreditar em “energia”, astrologia, orixás, duendes e gnomos (NOVAES, 2004), é bastante fácil que ideias desviantes sejam apresentadas em sala de aula como corretas, haja vista que muitos professores de Ciências mantêm concepções pseudocientíficas (EVE; DUNN, 1990; KALLERY, 2001; MARTIN, 1994; MCCLOUGHLIN; KALLERY; PSILLOS, 2015). Especialmente popular tem sido o ponto de vista chamado “misticismo quântico”, o qual propõe a existência de ligações entre a mecânica quântica e o misticismo oriental (CREASE; MANN, 1982, p. 306) e que, na última década, se disseminou tão fortemente na mídia e em boa parte da sociedade (PESSOA JR., 2010, p. 280).

² Conforme informações disponíveis em <portal.inep.gov.br>.

Segundo Pessoa Jr., os adeptos do misticismo quântico são, “em sua maioria, pessoas com predisposição a uma visão mística de mundo, que não entendem exatamente do que trata a Física quântica, mas confiam ou acreditam nos livros de divulgação que leram” (2010, p. 293). No entanto, vários cientistas, “desafiadores” à ciência estabelecida (2010, p. 295), contribuíram, intencionalmente ou não, para o “*corpus* teórico” do misticismo quântico, enquanto outros escritores dedicam-se a ganhar dinheiro com livros, palestras e *workshops*.

Essa “tendência mística do nosso tempo”, nas palavras de Einstein (1979, p. 40), tem sido objeto de vários estudos que representam suas origens históricas e filosóficas, seus elementos religiosos e influências culturais (p.ex., PESSOA JR., 2010). Segundo os críticos, o misticismo quântico não consiste apenas de ciência popular ou mesmo mal feita, mas algo mais grave: imitações corruptas da ciência propriamente dita (GRIM, 1982b, p. 87–88); ou, como classificaria Bunge (1982), uma “pseudociência”.

Este estudo, no entanto, partiu em uma direção diferente, focando a seguinte pergunta: tendo sido as livrarias inundadas por títulos que incluem a palavra “quântica” que, no entanto, têm mais a natureza de misticismo quântico do que de científico ou de popularização da ciência, como pode o leigo, ou mesmo um professor menos informado, distinguir entre eles? Assim, procurou-se, aqui, analisar os títulos e as sinopses dos livros disponíveis no Brasil com o objetivo de identificar “marcadores”, termos que possam ajudar o visitante de uma livraria a reconhecer a qual dessas categorias um livro pertence, antes mesmo de lê-lo completamente.

Para a análise desses textos, em vez da tradicional análise de conteúdo, a qual tem sido uma ferramenta essencial das ciências sociais por muito tempo, os dados foram processados com a ajuda de técnicas de Ciência de Dados (*Data Science*), mineração de dados textuais (*text mining*) e de aprendizagem de máquina (*machine learning*). Tal abordagem se justifica com o avanço da tecnologia de informação que, por um lado, produz hoje muito mais textos cuja análise é de interesse e, por outro lado, também proporciona novas técnicas automatizadas ou não, para essa análise. Estas, desde a virada do milênio, vêm causando uma revolução de Big Data nessa área, permitindo, por exemplo, estudos envolvendo milhões de documentos em mídia social e conteúdo de jornais, propiciando análise de viés de gênero, legibilidade, similaridade de conteúdo, preferências dos leitores e, até mesmo, de sentimento.

Um exemplo recente dessa metodologia é a pesquisa de bases de dados do Inep, realizada por Fonseca e Namen (2016), com o intuito de identificar fatores que relacionam o perfil de professores que lecionam Matemática com a proficiência obtida por seus alunos. Por sua vez, Rocha e Massarani (2016) combinaram a mineração quantitativa de textos com o método qualitativo da etnografia numa metodologia mista para analisar comentários feitos no site da revista *Ciência Hoje das Crianças* (CHC Online), a fim de compreender a percepção de internautas sobre os textos de divulgação científica lá publicados.

Inicialmente, faremos uma breve revisão da história da Mecânica Quântica, com o objetivo de contextualizar o surgimento do “misticismo quântico”. Em seguida, a metodologia utilizada será discutida em mais detalhe. Na sequência, será feita a análise dos dados, incluindo

uma exploração inicial e a mineração de texto, propriamente dita. Finalmente, teceremos algumas conclusões sobre o trabalho.

Destaca-se que, em se tratando de “pesquisa que utiliza informações de acesso público”, esta pesquisa foi dispensada de registro e avaliação pelo sistema CEP/CONEP³.

II. O Misticismo quântico

Não temos a pretensão de discorrer aprofundadamente sobre o tema num espaço tão reduzido. O leitor interessado é indicado aos capítulos de Crease e Mann(1982) e de Pessoa Jr. (2010) e ao livro de Freire Junior (2014).

Em 1900, Max Planck introduziu a ideia de que a energia era quantizada, com o objetivo de derivar uma fórmula, obtida anteriormente, para a dependência da frequência observada com a energia emitida por um corpo negro (PLANCK, 1900). Em 1905, Einstein explicou o efeito fotoelétrico por um postulado de que a luz, ou mais especificamente toda a radiação eletromagnética, pode ser dividida num número finito de “quanta de energia”, mais tarde conhecidos como fótons (EINSTEIN, 1905).

Em 1924, o físico francês Louis-Victor de Broglie apresentou a sua teoria de ondas de matéria, dizendo que as partículas podem exibir características de onda e vice-versa (DE BROGLIE, 1924). Em 1926, a função de onda foi interpretada por Max Born como uma medida da probabilidade de se encontrar a partícula em determinada posição e em determinado tempo (BORN, 1926). Em 1927, a partir da relação de comutação entre operadores que representam observáveis, Werner Heisenberg enunciou uma relação entre as inacurácias (em alemão, *Unge nauigkeit*) com que se podem efetuar medições simultâneas de pares conjugados de observáveis em nível subatômico (HEISENBERG, 1927). Posteriormente, ela se popularizou com o nome de “Princípio da Incerteza”, a partir do termo *Unsicherheitsrelation* que aparece numa nota no fim de seu artigo (HEISENBERG, 1927).

Como se sabe, a mecânica quântica é matematicamente bem compreendida, no sentido de que se sabe como usar seu aparato para calcular, para fazer previsões que são espetacularmente bem sucedidas em termos de potência e precisão. No entanto, é notável que, fisicamente (outros diriam, desdenhosamente, “filosoficamente”), no sentido de descrever como é que o mundo “lá em baixo” é, ela seja pelo menos controversa, com uma divergência acentuada entre físicos e filósofos, que se materializa em várias *interpretações* distintas da interpretação padrão “de Copenhague” (PYKACZ, 2015; SCHLOSSHAUER; KOFLER; ZEILINGER, 2013), cada uma delas, no entanto, “internamente consistente e, de modo geral, consistente com experimentos quânticos” (MONTENEGRO; PESSOA JR., 2002).

³ Resolução nº 510, de 07 de abril de 2016, do Plenário do Conselho Nacional de Saúde, Art. 1º, Parágrafo único, Item V.

Sendo talvez o mais eminente adversário da interpretação de Copenhague, Einstein, durante suas discussões com Bohr na conferência Solvay de 1930, concebeu um *Gedankenexperiment* (experimento mental) que, ele considerou, revelava a incompletude dessa interpretação. Mais tarde, este evoluiu para o bem conhecido Experimento EPR, mesmo que na versão altamente abstrata e estereotipada de Podolsky, em vez da do original de Einstein (EINSTEIN; PODOLSKY; ROSEN, 1935).

No decurso da sua correspondência com Schrödinger, Einstein descreveu um “exemplo macroscópico grosseiro”, uma “explosão de barril de pólvora”, cuja função de onda Ψ , “sem artes de interpretação”, “pode ser transformada em uma descrição adequada de um estado real das coisas; na realidade, não há intermediário entre explodido e não-explodido” (FINE, 2014). Schrödinger respondeu com “um exemplo muito semelhante”, o agora famoso experimento do “paradoxo do gato” (SCHRÖDINGER, 1935).

Para Schrödinger, o experimento do gato na caixa mostrava que havia algo de errado com a mecânica quântica. Ao contrário, para Bohr e Heisenberg, a mecânica quântica estava certa e o problema residia na falta de capacidade de visualizar esse novo mundo totalmente diferente “lá em baixo” (CREASE; MANN, 1982, p. 305). Além disso, variante da de Copenhague, a interpretação de London e Bauer (1939), de uma forma não tão filosoficamente rigorosa, afirmava que o “observador”, ao abrir da caixa de Schrödinger, colapsava a função de onda e tornava a situação quântica definida (FREIRE JUNIOR, 2014, p. 146). Disto, à conclusão de que a existência do mundo depende de consciência e que a realidade é nossa criação, foi apenas um pequeno passo (CREASE; MANN, 1982, p. 306). Pauli acreditava mesmo que o advento da mecânica quântica seria capaz de unificar a consciência como uma extensão da teoria quântica (MARIN, 2009).

Passada a segunda guerra mundial, Wigner, com seu artigo “*Remarks on the mind-body question*” (Observações sobre a questão mente-corpo) (1961), reacende a discussão sobre o papel que a consciência pode desempenhar nos processos quânticos. Nos anos seguintes, essa ideia “atraiu alguns físicos, alguns pesquisadores conscienciosos e um grande número de místicos” (SEIFE, 2000) e poucos de seus fundadores teriam imaginado que uma teoria com previsões tão extraordinariamente precisas “poderia abrir tal Caixa de Pandora de especulações nebulosas” (CREASE; MANN, 1982, p. 306):

Ao longo dos anos, as implicações ontológicas da interpretação de Copenhague têm inspirado comentários por físicos filosoficamente inclinados e filósofos fisicamente inclinados. Algumas de suas conclusões têm sido estranhas o suficiente para inspirar um crescimento secundário de escritos por excêntricos filosoficamente e fisicamente inclinados, o que tirou a mecânica quântica para fora do domínio subatômico e dotou-a de significado amplo e geral para a ação humana e do destino humano. [...] Especialmente popular, recentemente, tem sido o misticismo quântico, a visão de que existam relações entre a mecânica quântica e misticismo oriental (CREASE; MANN, 1982, p. 306).

A Física quântica está bem consolidada como a teoria mais bem aceita pela comunidade científica para a descrição das constituições internas de átomos e moléculas e das suas interações entre si e com diferentes formas de radiação. No entanto, o domínio de aplicação dessa teoria restringe-se à escala subnanométrica (abaixo de um nanômetro (10^{-9} m)), típica de uma molécula pequena e, quando se tenta estendê-lo ao nosso domínio macroscópico cotidiano, o grande número de átomos tende a “borrar” os efeitos quânticos, resultando num comportamento caracteristicamente de acordo com a física clássica pré-quântica (PESSOA JR., 2010, p. 279). Por exemplo, seria discutível determinar a função de onda de um caminhão, conforme apontaram Ridgen e James (1979) em editorial ao *American Journal of Physics*, seguido de proveitosa discussão (FANNER, 1979; FRIGG, 1979; HASLETT, 1980).

Apesar disso, as várias interpretações da Mecânica Quântica foram utilizadas para “embasar” diversas teses do misticismo quântico, tais como “o observador cria a realidade”, “o livre arbítrio é garantido pelo princípio de incerteza”, “a não-localidade quântica entre mentes permite transmissão instantânea de pensamentos”, “a mente pode transformar o universo com pensamento positivo”, “a alma pode viver em universos paralelos e estas contrapartidas podem se encontrar”, “religiões orientais já chegaram, no passado, às principais ideias da física moderna”, “a física quântica fundamenta terapias alternativas”, “a física quântica é o paradigma do novo milênio”, dentre outras (PESSOA JR., 2010, p. 285-292).

Segundo Pessoa Jr., o misticismo quântico, em linhas gerais, consiste em interpretações da Teoria Quântica que se inserem na tradição do naturalismo animista ou adotam um idealismo subjetivista, ou, ainda, que partem de elementos religiosos, atribuindo uma conexão íntima entre a consciência humana (ou a espiritualidade) e os fenômenos quânticos (2010, p. 284). Segundo esse autor, “ele faz parte de um movimento sociocultural que ressurgiu nas décadas de 1980-1990, herdeiro da “contracultura” das décadas de 1960-1970, mas com alguns traços diferentes” (2010, p. 296), tais como individualismo, caráter urbano em vez do ruralismo comunitário *hippie*, preocupações ecológicas e com a natureza e certo feminismo “mágico” (2010, p. 296-297).

Assim, em 1975, Fritjof Capra, físico teórico, surgiu e explodiu no cenário literário com o livro “O Tao da Física” (1975), onde traçou um paralelo entre a física moderna (relatividade, física quântica, física das partículas) e as filosofias e pensamentos orientais tradicionais, tais como o Taoísmo, o Budismo (incluindo o Zen) e o Hinduísmo, abrindo caminho para uma legião de futuros pseudocientistas. Depois disso, muitos de nós, físicos, já vivemos a experiência constrangedora de sermos convidados por alguém em um círculo de amigos para testemunhar que a mecânica quântica explica e valida a telepatia, a “Lei da Atração”, a cura por cristais ou qualquer outra crença pseudocientífica.

O uso do termo “pseudociência” pelos críticos sugere que parapsicologia, ufologia, astrologia, e o que mais tarde irá chamar de “misticismo quântico”, não são mera ciência popular, ou mesmo má ciência, mas que são algo totalmente diferente: pobres fac-símiles ou imita-

ções corruptas da ciência propriamente dita. Isto imediatamente levanta uma questão fundamental na filosofia da ciência. O que é a ciência, o que é pseudociência, e qual é a diferença essencial entre eles? Este, seguindo Karl Popper, é o que é chamado de “problema da demarcação” (GRIM, 1982a, p. 87-88).

Transcende o escopo deste trabalho uma discussão mais aprofundada sobre as principais teses do misticismo quântico. Da mesma forma, não discutiremos a importante questão ética de como a ciência estabelecida, de cunho eminentemente não-místico, poderia dialogar com o misticismo quântico, tema já devidamente tratado por Pessoa Jr. (2010).

III. Metodologia

Os catálogos on-line das quatro maiores livrarias do Brasil, Cultura, Saraiva, Amazon e FNAC, foram pesquisados em dois momentos: a primeira, entre março e abril de 2016, e a segunda em outubro de 2016. Para evitar subjetividades na seleção dos livros, critérios objetivos foram definidos: na recolha inicial dos livros, foram selecionados unicamente livros que continham as palavras “quântica” ou “quântico” em seus títulos; na segunda, foram selecionados não somente livros que, por algum motivo, escaparam da primeira recolha, mas, também, livros que, mesmo sem conter aquelas palavras no título, as contivessem na sinopse.

Para cada livro, foi construído um registro contendo dados adicionais, tais como nome do autor, sinopse e data de publicação. No entanto, apenas os títulos e as sinopses dos livros constituíram o *corpus* da análise de texto.

Esse banco de dados foi analisado com técnicas de ciência de dados, especificamente, utilizando os recursos de mineração de dados textuais (*text mining*), análise de texto e aprendizagem de máquina (*machine learning*) da linguagem R (R CORE TEAM, 2016) versão 3.3.1 (2016-06-21). Apesar de R ser muito versátil e contar com muitos recursos, todo o trabalho foi realizado em tempo razoável num desktop convencional, com arquitetura i386-w64-mingw32/i386, e sobre Windows7 (build 7600, 32-bit).

A mineração de textos, ou mineração de dados textuais, pode ser entendida, basicamente, como a aplicação da chamada mineração de dados (*data mining*) ao específico contexto em que os “dados” são textuais, em vez de puramente numéricos. Em ambos os casos, trata-se de explorar um conjunto de dados, estruturados ou não, de forma a identificar regularidades, padrões ou tendências que proporcionem um novo conhecimento. Quando os dados resultam de “pegadas digitais espontaneamente deixadas no universo digital em que vivemos” devido a “pesquisas na web, postagens, partilhamentos, *likes* em redes sociais, etc.”, temos o que Dos Santos (2015) denominou *crowdledge*, isto é, um conhecimento inesperado que emerge fracamente por meio de análise desses dados.

A mineração de dados era até pouco tempo considerada uma das cinco etapas do processo de Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases – KDD*) proposto por Fayyad, Piatetsky-Shapiro e Smyth (1996), que visava a identificação

de padrões novos, válidos, potencialmente úteis e compreensíveis, que estivessem embutidos nos dados:

1. Seleção de dados e de bases adequadas, a partir do domínio do conhecimento envolvido;
2. Limpeza dos dados, já que frequentemente apresentam inúmeras inconsistências que podem comprometer o resultado das análises;
3. Transformação dos dados em tamanhos e formatos mais adequados para as próximas etapas;
4. Escolher técnicas e algoritmos para a mineração de padrões;
5. Avaliação e interpretação dos padrões extraídos, em forma do novo conhecimento, o qual deve ser devidamente divulgado às partes interessadas, com todo o processo devidamente documentado.

Hoje, as três primeiras etapas são agrupadas numa fase denominada análise exploratória de dados (PENG, 2016).

Usando o *package tm* de mineração de texto na linguagem R (FEINERER; HORNIK; MEYER, 2008), os títulos e as sinopses desses livros foram, então, transformados em *corpora* linguísticos não-annotados, para tratamento posterior. Em seguida, as palavras dos *corpora* foram transformadas em letras minúsculas, retiradas pontuações e outros caracteres especiais, espaços em excesso e números. Palavras vazias (*stop words*)⁴ também foram removidas, utilizando-se as listas de palavras vazias do *package tm*, a partir do projeto *Snowball*⁵, e do site *ranks.nl*⁶. Além disso, como a sintaxe da linguagem R não permite caracteres acentuados ou especiais, como hífen, no nome das variáveis, as palavras dos títulos e das sinopses tiveram de ser transliteradas para seus correspondentes sem acento e para um ponto, em lugar do hífen.

Em vez de construir matrizes de indexação (*document-term matrices*), a partir do *corpus*, foram produzidas *dataframes* em que as palavras consistiam nas colunas e as suas frequências relativas ao total de palavras em cada em cada livro, nas linhas.

Devido à pequena extensão da base, e por causa das sinopses e, mais ainda, os títulos dos livros consistirem de textos curtos, as palavras restantes apresentavam-se esparsas, isto é, com pequena frequência nos textos, dificultando sobremaneira a análise por métodos usuais, tais como o modelo saco-de-palavras (*bag of words*) (HARRIS, 1954).

Embora palavras com frequências quase constantes em todos os textos do *corpus*, também chamadas de preditores de variância próxima de zero, sejam geralmente removidas, sob o pressuposto de que elas são pouco informativas, este procedimento frequentemente resulta em remover os preditores mais fortes do modelo (GELMAN *et al.*, 2008) e, portanto, esta prática

⁴ Palavras vazias são palavras consideradas pouco relevantes, tais como verbos auxiliares, conjunções e artigos, que são removidas antes do processamento de um texto em linguagem natural (IKONOMAKIS; KOTSIANTIS; TAMPAKAS, 2005).

⁵ <<http://snowball.tartarus.org/algorithms/portuguese/stop.txt>>.

⁶ <<http://www.ranks.nl/stopwords/brazilian>>.

não foi seguida aqui. No entanto, as palavras que têm frequências constantes em todos os textos do *corpus* têm verdadeiramente variância zero e foram removidas.

Outro problema decorre de variáveis altamente correlacionadas entre si; a abordagem usual PCA (Análise de Componentes Principais) (KUHN, 2008) poderia tornar mais difícil de interpretar os preditores e, portanto, optou, em vez disso, por simplesmente remover variáveis com uma correlação entre elas, medida pelo coeficiente de correlação de Pearson, maior que 80%.

Gelman *et al.* (2008) sugere que as variáveis contínuas sejam centradas em um valor médio de 0 e reescaladas para um desvio padrão de 0,5 para assemelhá-las a variáveis binárias, as quais assumem apenas os valores 0 e 1. Neste caso, no entanto, como se está apenas interessado na presença ou não das palavras, em vez de sua frequência, as variáveis foram reescaladas como binárias, isto é, valores de frequência maiores que zero foram transformados em 1.

A prática usual de *machine learning* recomenda subdividir aleatoriamente o conjunto de dados e destinar 70-60% dos dados para o treinamento do modelo e os restantes 30-40% para sua validação. A linguagem R tem recursos para realizar essa subdivisão de forma automática, levando em conta um equilíbrio entre as categorias dos dados nos dois subconjuntos. A fim de garantir a reprodutibilidade, foi estabelecida uma semente global (42) para o gerador de números pseudoaleatórios.

Quando o conjunto de dados é pequeno em relação ao número de variáveis, a fim de evitar sobreajuste (*overfitting*) e reduzir os erros de previsão quando aplicados fora da amostra (*out-of-sample errors*), realiza-se uma validação cruzada *k-fold* (*k-fold cross-validation*) (SENI; ELDER, 2010, p. 26–28) sobre o conjunto de dados de treinamento. Nesta, divide-se o conjunto de dados em *k* subconjuntos mutuamente exclusivos do mesmo tamanho, *k-1* subconjuntos são utilizados para o treinamento do modelo e o restante para sua validação e cálculo da acurácia do modelo; este processo é, então, realizado *k* vezes, alternando o subconjunto de validação e, ao final, calcula-se a acurácia geral esperada, resultando numa medida mais confiável da expectativa da capacidade do modelo em representar os dados.

IV. Resultados e discussão

A recolha inicial dos livros, pelos critérios descritos acima, resultou em 131 itens distintos, enquanto a segunda resultou em 50 títulos adicionais.

Os livros eram originalmente classificados pelas livrarias em gêneros diversos e, por vezes, arbitrários, tais como “Autoajuda”, “Ciências Exatas – Física”, “Medicina Alternativa”, ou mesmo “Infantil”, como se vê na legenda da Fig. 1. Essa classificação talvez merecesse um estudo por si só, mas transcenderia o escopo deste trabalho.

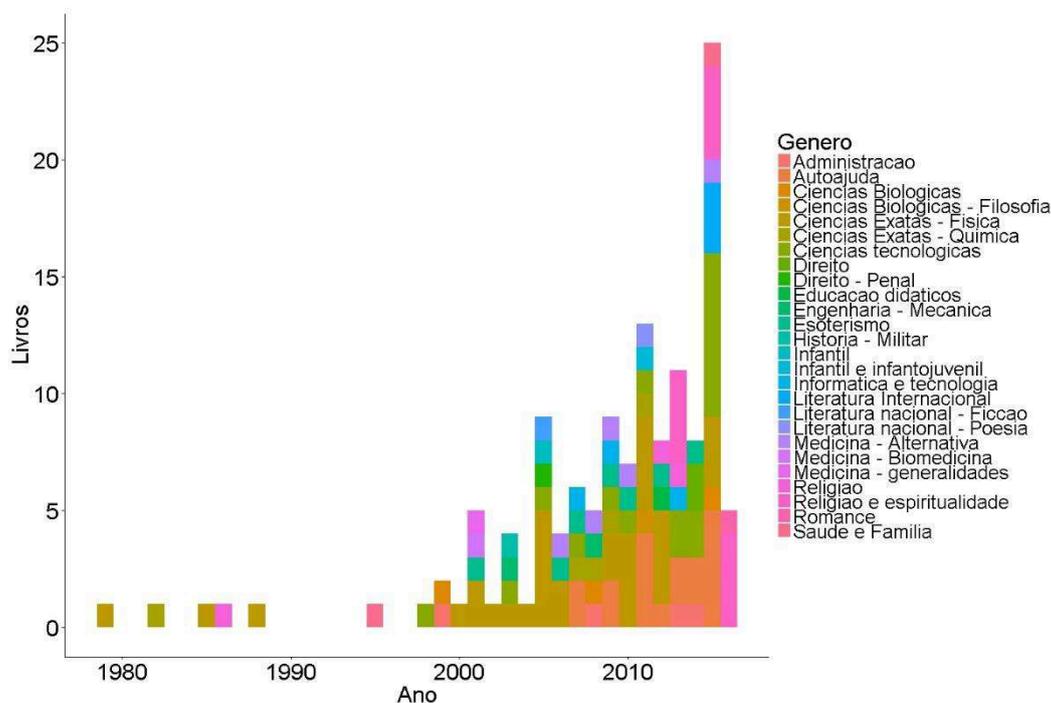


Fig. 1 – Classificação original dos livros por gênero. Fonte – Pesquisa.

Para os fins deste trabalho, os livros foram reclassificados em apenas três categorias, de acordo com os “critérios de demarcação” de Bunge (1982): “ciência” (incluindo tecnologia), “pseudociência” e “não ciência” (incluindo áreas do conhecimento que não se baseiam em ciência sem, no entanto, entrarem em conflito com ela, tais como literatura). Como exemplo destas reclassificações, temos:

- “Bohr e a Interpretação Quântica da Natureza”: de “Infantil” para “ciência”;
- “Criatividade Quântica”: de “Ciências biológicas” para “pseudociência”;
- “Entre Vácuos”: de “Ciências tecnológicas” para “não ciência”.

Não vimos necessidade de incluir a quarta categoria proposta por Bunge (1982), a “protociência”, que inclui campos de conhecimento em processo de se tornarem científicos. O total de livros em cada categoria é apresentado na Tabela 1.

Tabela 1 – Distribuição dos livros segundo as categorias.

Categoria	Freq.	%
Ciência	91	50
Pseudociência	81	45
Não ciência	9	5
Total	181	100

Fonte – Pesquisa

Como primeiro resultado dessa análise exploratória, verificou-se que a atual disponibilidade de livros científicos remonta aos anos 70, enquanto uma tendência crescente de obras pseudocientíficas surgiu no Brasil desde 2000 (

Fig. 2).

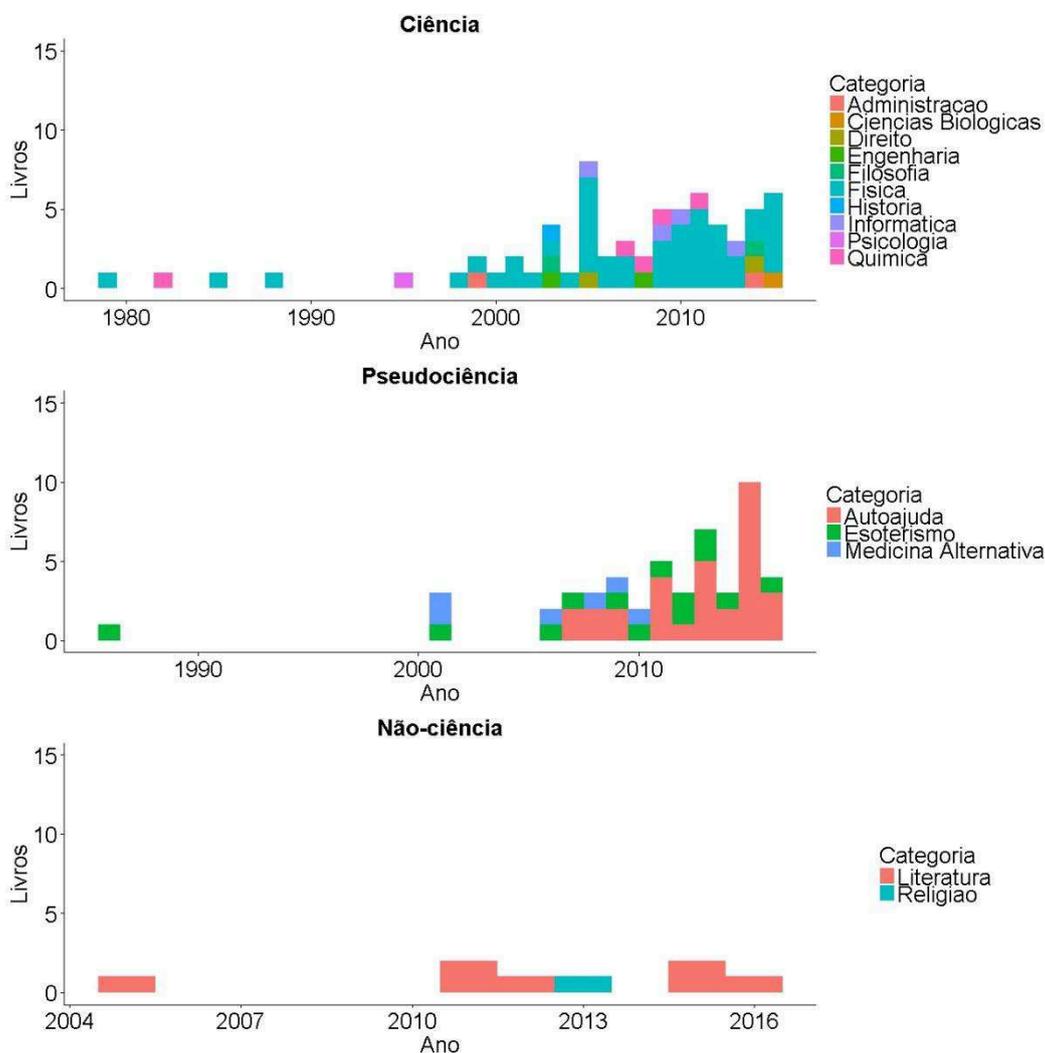


Fig. 2 – Distribuição dos livros por categoria e ano de publicação. Fonte – Pesquisa

Continuando a análise exploratória, dos 181 livros contendo as palavras “Quântica” ou “Quântico” em seus títulos, selecionamos apenas os 176 escritos em português para o conjunto de dados de treinamento.

Várias explorações do texto foram, então, realizadas, incluindo *K-Means Clustering* e dendrogramas. A Fig. 3 mostra alguma possibilidade nesse sentido, mas ainda longe do resultado que se desejava.

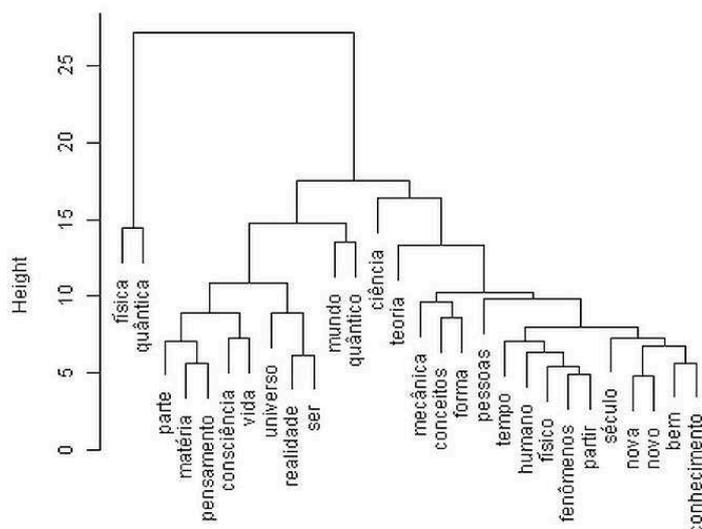


Fig. 3 – Dendrograma para as palavras nas sinopses dos livros. Fonte – Pesquisa

A categoria “não ciência” foi abandonada por ter pouca presença no conjunto de livros listados e, principalmente, por não causar prejuízos ao aprendizado em ciência, já que incluía apenas obras de literatura, tais como as de ficção científica “Utopia Quântica” e “O mistério da esfera negra e a caixa Quântica”. Esta escolha reduziu o conjunto de dados a 167 livros.

Para garantir a melhor discriminação entre as categorias “ciência” e “pseudociência”, palavras associadas simultaneamente a ambas as categorias foram descartadas. Após todos os procedimentos de limpeza descritos acima, o conjunto de dados foi reduzido de 6366 para 86 palavras (colunas) dos títulos e 272 palavras das sinopses.

Após a subdivisão aleatória do conjunto de dados em aproximadamente 70% deles para o treinamento do modelo e 30% para sua validação, resultou em conjuntos de 119 e 49 observações (livros), respectivamente, para as mesmas 86 e 272 palavras em ambos. Para evitar sobreajuste (*overfitting*), foi efetuada uma validação cruzada de 7 vezes.

Vários dos atualmente mais populares algoritmos de *machine learning* de classificação, disponíveis no *package* Caret (KUNH, 2008), foram testados, incluindo *Support Vector Machines with Radial Basis Function Kernel SVMRadial* (KARATZOGLOU *et al.*, 2004), *Extreme Gradient Boosting (XGB)* (CHEN; HE; BENESTY, 2016; FRIEDMAN, 2001), *Random Forests* (BREIMAN, 2001; LIAW; WIENER, 2002), e *Recursive Partitioning and Regression Trees (RPART)* (BREIMAN *et al.*, 1984; THERNEAU; ATKINSON; RIPLEY, 2015), dentre vários outros.

Após a aplicação de cada modelo aos conjuntos de dados de treinamento, o *package* Caret também calcula automaticamente os usuais parâmetros acurácia de Jordan e índice Kappa de Cohen para avaliar as expectativas da capacidade dos modelos em classificar os dados de teste. A acurácia será mais facilmente descrita a partir das tabelas 4 a 7, adiante. Aqui, basta dizer que, quanto mais próximo de um desses valores, melhores são as perspectivas dos modelos. Já, o índice Kappa de Cohen é calculado por:

$$\kappa = \frac{p_0 - p_e}{1 - p_e},$$

onde p_0 é a acurácia e p_e é a probabilidade hipotética dos acertos terem sido fruto de mero acaso. Os resultados desses parâmetros para títulos e sinopses são apresentados nas Tabelas 2 e 3.

Tabela 2 – Comparação entre os modelos para títulos.

Modelo	Acurácia	Kappa
SVMRadial	0.838	0.669
XGB	0.745	0.475
Random Forests	0.763	0.511
RPART	0.822	0.636

Fonte – Pesquisa.

Tabela 3 – Comparação entre os modelos para sinopses.

Modelo	Acurácia	Kappa
SVMRadial	0.873	0.741
XGB	0.788	0.565
Random Forests	0.839	0.670
RPART	0.831	0.653

Fonte – Pesquisa.

Dessas tabelas, observa que o modelo SVMRadial é o que parece mais promissor, com o RPART em segundo lugar, no conjunto títulos e sinopses.

Agora, os modelos devem ser validados, aplicando-os aos conjuntos de dados de teste, tanto de palavras de títulos quanto de sinopses, com 49 livros cada, simulando o caso real de um visitante escolhendo livros nas livrarias por seus títulos e sinopses. Os resultados podem ser vistos nas matrizes de confusão (*confusion matrices*) das Tabela 4, Tabela 5, Tabela 6 e Tabela 7.

Tabela 4 – Matriz de confusão do modelo SVMRadial para títulos.

Classe real	Classe predita		Erro de Classificação
	Pseudociência	Ciência	
Pseudociência	19	4	0.21
Ciência	0	26	0.00

Fonte: Pesquisa

Tabela 5 – Matriz de confusão do modelo SVMRadial para sinopses.

Classe real	Classe predita		Erro de Classificação
	Pseudociência	Ciência	
Pseudociência	16	7	0.44
Ciência	0	26	0.00

Fonte: Pesquisa

Tabela 6 – Matriz de confusão do modelo RPART para títulos.

Classe real	Classe predita		Erro de Classificação
	Pseudociência	Ciência	
Pseudociência	20	3	0.15
Ciência	0	26	0.00

Fonte: Pesquisa

Tabela 7 – Matriz de confusão do modelo RPART para sinopses.

Classe real	Classe predita		Erro de Classificação
	Pseudociência	Ciência	
Pseudociência	14	9	0.64
Ciência	0	26	0.00

Fonte: Pesquisa

Como se vê nessas tabelas, em ambos os modelos, todos os 26 livros científicos do conjunto de teste foram corretamente classificados na categoria “ciência”, mas nem todos os livros pseudocientíficos foram reconhecidos como tal pelos seus títulos. O modelo RPART foi o que teve a menor taxa de erro de classificação de livros pseudocientíficos, por palavras dos títulos ($3/20 = 15\%$). Os 3 livros que foram incorretamente classificados por ele são:

- “O Salto Quântico”
- “Terapias Quânticas. Cuidando do Ser Inteiro”
- “A Era do Raciocínio Quântico”

A acurácia é calculada pelo quociente entre a soma das classificações corretas e, no caso, a quantidade total de livros. Assim, o modelo RPART teve uma acurácia de $46/49 \cong 94\%$ na classificação dos livros por palavras dos títulos.

O resultado sensivelmente pior para as sinopses em relação aos títulos deve-se, provavelmente, ao elevado número de palavras nestas, chegando a um máximo de seiscentas e quarenta e seis palavras em um único livro, contra um máximo de apenas vinte e uma nos títulos, o que pode ter “confundido” o algoritmo de aprendizagem de máquina por uma extensa variedade de palavras em cada categoria.

O RPART é um modelo econômico, pois consiste numa árvore de regressão com apenas 22 “folhas” relacionadas às seguintes 22 palavras dos títulos: espiritualidade, mente, ajudar, amor, indiano, humanidade, prática, objetivos, sabedoria, cura, doenças, espírito, seitas, problemática, justiça, menezes⁷, singulares, renome, ética, funcionamento, concisa e crenças. De fato, é difícil imaginar palavras tais como “espiritualidade”, “amor”, “cura” e “espírito” como fazendo parte de um texto científico sobre mecânica quântica, sendo, no entanto, bastante compatíveis com as “teses” do misticismo quântico, mencionadas anteriormente.

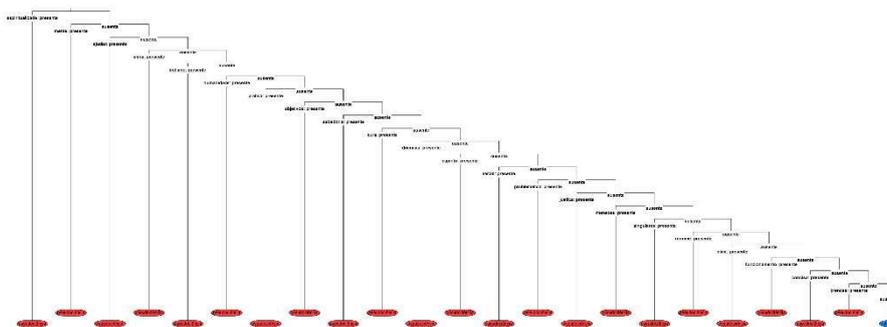


Fig. 4 – Árvore de regressão produzida. Fonte: Pesquisa.

A Fig. 4, produzida com o *package rpart.plot* (MILBORROW, 2017), apresenta a árvore de regressão. Embora essa figura não permita identificar essas palavras nas folhas, ela é importante porque mostra a estrutura dessa árvore, indicando que, como desejado, as 22 palavras são todas “positivas” (estão do lado esquerdo da ramificação), no sentido em que a presença de todas e qualquer uma daquelas palavras no título de um livro indica que ele pertence à categoria “misticismo quântico” e nenhuma delas se relaciona à categoria “ciência” (lado direito da ramificação). Para melhor visualização, a Fig. 5 apresenta as primeiras 5 “folhas” da árvore da Fig. 4.

Como pode se ver na Fig. 5, a presença de qualquer das palavras “espiritualidade”, “mente”, “ajudar”, “amor”, ou “indiano”, por exemplo, no título de um livro que inclua também as palavras “quântico” ou “quântica” já o indica como sendo de “misticismo quântico”. Se nenhuma das 22 palavras acima aparecer no título, há mais de 90% de chance de se tratar de um livro científico sobre Quântica.

⁷ A palavra ‘menezes’ aparece em vários títulos, tais como “A Força da Calma no Xamanismo de Jorge Menezes”. Por isso, foi identificada pelo programa como uma das 22 principais palavras que, aparecendo no título de um livro que inclua também as palavras “quântico” ou “quântica”, o indica como sendo de “misticismo quântico”.

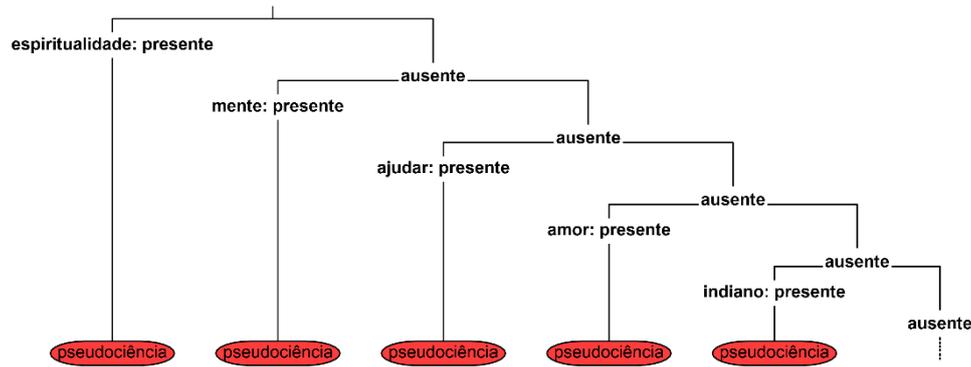


Fig. 5 – Parte superior da árvore de regressão da Fig. 4. Fonte: Pesquisa.

V. Conclusões

A lista de 22 palavras obtida discrimina com uma acurácia de 94% entre as categorias. Desta forma, um visitante de uma livraria, tendo em mãos um livro que supostamente se refira a “quântica”, antes mesmo de lê-lo, caso identifique qualquer uma dessas palavras em seu título, poderá assumir, com grande segurança, que ele pertence à categoria “misticismo quântico” e não à “ciência”.

Por outro lado, é preocupante o fato de que, conforme a Tabela 1, quase metade dos livros cujos títulos sugerem serem científicos, consistam, na verdade, de obras pseudocientíficas, livros esses que estão disponíveis nas principais livrarias do país e até mesmo em livrarias associáveis a institutos de ciências exatas de universidades respeitáveis.

Acreditamos, assim, que este trabalho reveste-se de importância para a área de Pesquisa em Ensino de Ciências e esperamos que ele auxilie os leitores, especialmente os professores do ensino médio, a estarem menos suscetíveis a serem enganados por esses livros e, com isso, ajudando a evitar que ideias desviantes sejam apresentadas em sala de aula como corretas.

Finalmente, o sucesso em ter obtido um conjunto relativamente pequeno de termos que discriminam com uma acurácia de 94% entre as categorias “ciência” e “pseudociência” demonstra a contribuição que a Ciência de Dados em geral, a mineração de dados e a aprendizagem de máquina em particular podem dar para pesquisas em Educação. Acreditamos, por isso, que, embora raros até o momento, estudos que as utilizem devam ser incentivados, inclusive nos Programas de Pós-Graduação.

Agradecimentos

Mairus Disconzi de Moura agradece à FAPERGS pela bolsa de iniciação científica.

Referências

BORN, M. Zur Quantenmechanik der Stoßvorgänge. **Zeitschrift für Physik**, v. 37, n. 12, p. 863-867, Dez. 1926.

BREIMAN, L. *et al.* **Classification and Regression Trees**. London: Chapman and Hall/CRC, 1984. (Wadsworth Statistics/Probability series).

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001.

BUNGE, M. Demarcating Science from Pseudoscience. **Fundamenta Scientiae**, v. 3, n. 3 e 4, p. 369-388, 1982.

CAPRA, F. **The Tao of Physics**: An exploration of the parallels between modern physics and eastern mysticism. Boulder, CO: Shambhala, 1975.

CHEN, T.; HE, T.; BENESTY, M. **xgboost**: Extreme gradient Boosting. R package version 0.4-4. Disponível em: <<https://cran.r-project.org/package=xgboost>>. Acesso em: 8 out. 2016.

CREASE, R.; MANN, C. C. The Yogi and the Quantum. In: GRIM, P. (Org.). **Philosophy of Science and the Occult**. SUNY Series in Philosophy. Albany, NY: SUNY Press, 1982. p. 302–314.

DE BROGLIE, L. **Recherches sur la Théorie des Quanta**. 1924. Thèse de doctorat. Sorbonne, Université de Paris, Paris.

DOS SANTOS, R. P. Big Data: Philosophy, Emergence, Crowledge, and Science Education. **Themes in Science and Technology Education**, v. 8, n. 2 Special Issue on Big Data in Education, p. 115-127, 2015.

EINSTEIN, A. **Albert Einstein, the human side**: Glimpses from His Archives. Princeton, NJ: Princeton University Press, 1979.

_____. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. **Annalen der Physik**, v. 322, n. 6, p. 132-148, Jun. 1905.

EINSTEIN, A.; PODOLSKY, B. Y.; ROSEN, N. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? **Physical Review**, v. 47, n. 10, p. 777-780, 15 May 1935.

EVE, R. A.; DUNN, D. Psychic Powers, Astrology & Creationism in the Classroom? Evidence of Pseudoscientific Beliefs among High School Biology & Life Science Teachers. **The American Biology Teacher**, v. 52, n. 1, p. 10-21, Jan. 1990.

FANNER, A. C. On the Broglie Wavelength of 10 ton Kenworth Truck (Letter). **American Journal of Physics**, v. 47, n. 2, p. 573, Feb. 1979.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 87, Sep. 1996.

FEINERER, I.; HORNIK, K.; MEYER, D. Text Mining Infrastructure in R. **Journal of Statistical Software**, v. 25, n. 5, p. 1-54, 31 Mar. 2008.

FINE, A. The Einstein-Podolsky-Rosen Argument in Quantum Theory. E. N. ZALTA (Org.) **The Stanford Encyclopedia of Philosophy**. Winter ed. Stanford, CA: Stanford University, 21 Dec. 2014.

FONSECA, S. O. D.; NAMEN, A. A. Mineração em bases de dados do Inep: uma análise exploratória para nortear melhorias no sistema educacional brasileiro. **Educação em Revista**, v. 32, n. 1, p. 133-157, mar. 2016.

FREIRE JUNIOR, O. **The Quantum Dissidents: Rebuilding the Foundations of Quantum Mechanics (1950-1990)**. Berlin, Heidelberg: Springer, 2014.

FRIEDMAN, J. H. Greedy Function Approximation: A Gradient Boosting Machine. **The Annals of Statistics**, v. 29, n. 5, p. 1189-1232, Oct. 2001.

FRIGG, G. L. de Broglie Wavelengths of Trucks and Other Objects (Letter). **American Journal of Physics**, v. 47, n. 2, p. 926, Feb. 1979.

GELMAN, A. *et al.* A weakly informative default prior distribution for logistic and other regression models. **The Annals of Applied Statistics**, v. 2, n. 4, p. 1360-1383, Dec. 2008.

GRIM, P. Quantum Mysticism. In: _____ (Org.). **Philosophy of Science and the Occult**. SUNY Series in Philosophy. Albany, NY: SUNY Press, 1982a. p. 267-271.

GRIM, P. Science or Pseudoscience? The Problem of Demarcation. In: _____ (Org.). **Philosophy of Science and the Occult**. SUNY Series in Philosophy. Albany, NY: SUNY Press, 1982b. p. 87-91.

HARRIS, Z. S. Distributional Structure. **WORD**, v. 10, n. 2-3, p. 146-162, 4 Aug. 1954.

HASLETT, J. de Broglie Phase Waves of a Truck? (Letter). **American Journal of Physics**, v. 48, n. 5, p. 335, May 1980.

HEISENBERG, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. **Zeitschrift für Physik**, v. 43, n. 3-4, p. 172-198, März 1927.

IKONOMAKIS, E. K.; KOTSIANTIS, S. B.; TAMPAKAS, V. Text Classification Using Machine Learning Techniques. **WSEAS Transactions on Computers**, v. 4, n. 8, p. 966-974, Aug. 2005.

KALLERY, M. Early-years Educators' Attitudes to Science and Pseudo-science: The case of

astronomy and astrology. **European Journal of Teacher Education**, v. 24, n. 3, p. 329-342, Oct. 2001.

KARATZOGLOU, A. *et al.* Kernlab - An S4 Package for Kernel Methods in R. **Journal of Statistical Software**, v. 11, n. 9, p. 1-20, 2 Nov. 2004.

KUHN, M. Building Predictive Models in R Using the caret Package. **Journal of Statistical Software**, v. 28, n. 5, Nov. 2008.

LIAW, A.; WIENER, M. Classification and Regression by randomForest. **R News**, v. 2, n. 3, p. 18-22, Dec. 2002.

LONDON, F. W.; BAUER, E. La théorie de l'observation en mécanique quantique. **Actualités scientifiques et industrielles : Exposés de physique générale**, n. 775, 1939.

MARIN, J. M. "Mysticism" in quantum mechanics: the forgotten controversy. **European Journal of Physics**, v. 30, n. 4, p. 807-822, 1 Jul. 2009.

MARTIN, M. Pseudoscience, the paranormal, and science education. **Science and Education**, v. 3, n. 4, p. 357-371, Oct. 1994.

MC CLOUGHLIN, T.; KALLERY, M.; PSILLOS, D. The Coconstruction of Scientific and Nonscientific Belief Systems in Educators. **Cybernetics and Systems**, v. 46, n. 6-7, p. 543-553, 3 Oct. 2015.

MILBORROW, Stephen. **rpart.plot**: Plot "rpart" Models: An Enhanced Version of "plot.rpart". R package version 2.1.2. 20 Apr. 2017. Disponível em: <<http://CRAN.R-project.org/package=rpart.plot>>. Acesso em: 10 jul. 2017.

MONTENEGRO, R. L.; PESSOA JR., O. F. Interpretações da Teoria Quântica e as Concepções dos Alunos do Curso de Física. **Investigações em Ensino de Ciências**, v. 7, n. 2, p. 107-126, maio 2002.

PENG, R. D. **Exploratory Data Analysis with R**. Victoria, CA-BC: Leanpub, 2016.

PESSOA JR., O. F. O fenômeno cultural do misticismo quântico. In: FREIRE JR., O.; PESSOA JR., O. F.; BROMBERG, J. L. (Org.). **Teoria Quântica - Estudos históricos e implicações**. São Paulo: Livraria da Física, 2010. p. 281-302.

PIERUCCI, A. F. O crescimento da liberdade religiosa e o declínio da religião tradicional: a propósito do censo 2010. **Anuac - Rivista dell'Associazione Nazionale Universitaria degli Antropologi Culturali**, v. 1, n. 2, Jul. 2012.

PLANCK, M. Entropie und Temperatur strahlender Wärme. **Annalen der Physik**, v. 306, n. 4, p. 719-737, Apr. 1900.

PYKACZ, J. A Brief Survey of Main Interpretations of Quantum Mechanics. **Quantum Physics, Fuzzy Sets and Logic**. SpringerBriefs in Physics. Berlin: Springer, 2015. p. 5-13.

R CORE TEAM. **R: A language and environment for statistical computing**. Disponível em: <<http://www.r-project.org/>>. Acesso em: 10 jun. 2016.

RIDGEN, J. S.; JAMES, P. B. Where are the Problems? (Editorial). **American Journal of Physics**, v. 47, n. 2, p. 125, Feb. 1979.

ROCHA, M.; MASSARANI, L. M. Divulgação científica na internet: um estudo de caso de comentários feitos por leitores em textos da Ciência Hoje das Crianças Online, **Alexandria: Revista de Educação em Ciência e Tecnologia**, v. 9, n. 1, p. 207-233, jan. 2016.

SCHLOSSHAUER, M.; KOFLER, J.; ZEILINGER, A. A snapshot of foundational attitudes toward quantum mechanics. **Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics**, v. 44, n. 3, p. 222-230, Aug. 2013.

SCHRÖDINGER, E. R. J. A. Die gegenwärtige Situation in der Quantenmechanik - II. **Die Naturwissenschaften**, v. 23, n. 49, p. 823-828, 6 Dez. 1935.

SEIFE, C. Cold Numbers Unmake the Quantum Mind. **Science**, v. 287, n. 5454, p. 791, 4 Feb. 2000.

SENI, G.; ELDER, J. F. **Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions**. San Rafael, CA: Morgan & Claypool, 2010.

THERNEAU, Terry; ATKINSON, Beth; RIPLEY, Brian. **rpart: Recursive Partitioning and Regression Trees**. R package version 4.1-10. [S.l: s.n.]. Disponível em: <<https://cran.r-project.org/package=rpart>>. Acesso em: 10 jun. 2016.

WIGNER, E. P. Remarks on the Mind-Body Question. In: GOOD, IRVING JOHN (Org.). **The scientist speculates: an anthology of partly-baked ideas**. London: Heinemann, 1961. p. 284-302.

ZIMMERMAN, Corinne; CROKER, Steve. A Prospective Cognition Analysis of Scientific Thinking and the Implications for Teaching and Learning Science. **Journal of Cognitive Education and Psychology**, v. 13, n. 2, p. 245-257, 1 Jun. 2014.