

## **Modelo de classificação de dados não estruturados para análise da competitividade de mercado**

### **Unstructured data classification model for competitiveness analysis competitiveness**

DOI:10.34117/bjdv7n9-377

Recebimento dos originais: 22/08/2021

Aceitação para publicação: 22/09/2021

#### **Delano Cordeiro Lima**

Mestrando em Administração de Empresas, professor de Graduação em Administração

Instituição: UNIFOR – Universidade de Fortaleza

Endereço: Av. Washington Soares, 1321 - Edson Queiroz – CEP: 60811-905

Fortaleza – CE-Brasil

E-mail: delano@unifor.br

#### **Carlos Eduardo Bittencourt Paiva**

Mestrando em Marketing, professor de Graduação em Administração

Instituição: UNIFOR – Universidade de Fortaleza

Endereço: Av. Washington Soares, 1321 - Edson Queiroz – CEP: 60811-905

Fortaleza - CE-Brasil

E-mail: carlosbittencourt@unifor.br

#### **Andrey Chaves**

Doutor em Física, professor de Pós-Graduação em Física

Instituição: UFC – Universidade Federal do Ceará

Endereço: Campus do Pici - UFC, Bloco 922, Fortaleza - Ceará, Brasil-Pici

Fortaleza - CE, CEP: 60440-900

E-mail: andrey@fisica.ufc.br

#### **Keysa Manuela Cunha de Mascena**

Doutora em Administração, professora de Pós-Graduação em Administração

Instituição: UNIFOR – Universidade de Fortaleza

Endereço: Av. Washington Soares, 1321 - Edson Queiroz – CEP: 60811-905

Fortaleza – CE-Brasil

E-mail: keysamascena@unifor.br

#### **José Wagner Borges Gondim**

Mestrando em Marketing, professor de Graduação em Jornalismo

Instituição: UNIFOR – Universidade de Fortaleza

Endereço: Av. Washington Soares, 1321 - Edson Queiroz – CEP: 60811-905

Fortaleza – CE-Brasil

E-mail: wagnerborges@unifor.br

**Josimar Souza Costa**

Doutor(a) em Administração, professor(a) de Graduação em Administração  
Instituição: UNIFOR – Universidade de Fortaleza  
Endereço: Av. Washington Soares, 1321 - Edson Queiroz – CEP: 60811-905  
Fortaleza - CE Brasil  
E-mail: josimarscosta@unifor.br

**RESUMO**

O artigo tem por objetivo apresentar um modelo de análise de posicionamento de mercado aplicado às instituições de ensino superior privadas. Foram coletados dados gerados em mídia social, totalizando 792.674 interações analisadas geradas por três instituições de ensino superior. Foi desenvolvido um modelo de classificação de dados não estruturados utilizando a ferramenta RapidMiner, considerando um grupo de cinco fatores prioritários, pelos quais é possível inferir as vantagens competitivas e o posicionamento de mercado das instituições de ensino superior privadas. O modelo utilizou-se de uma ferramenta matemática de comparação de curvas para estabelecer o grau de proximidade entre o posicionamento das marcas, desenvolvendo uma abordagem de comparação e análise de vantagens competitivas. Os resultados obtidos permitiram a identificação de curvas de posicionamento de mercado e a comparação da similaridade entre as curvas, contribuindo para a definição de um método de comparação de posicionamento entre concorrentes de mercado.

**Palavras-chave:** Posicionamento de Mercado, Dados Não Estruturados, Mídias Sociais, Big Data. Ensino Superior.

**ABSTRACT**

The article aims to present a model of market positioning analysis applied to private higher education institutions. Data generated in social media were collected, totaling 792,674 analyzed interactions generated by three higher education institutions. An unstructured data classification model was developed using the RapidMiner tool, considering a group of five priority factors, by which it is possible to infer the competitive advantages and market positioning of private higher education institutions. The model used a mathematical tool of curve comparison to establish the degree of proximity between the positioning of the brands, developing an approach to compare and analyze competitive advantages. The results obtained allowed the identification of market positioning curves and the comparison of the similarity between curves, contributing to the definition of a method for comparing positioning among market competitors.

**Keywords:** Market Positioning, Unstructured Data, Social Media, Big Data Data. Higher Education.

**1 INTRODUÇÃO**

O posicionamento de mercado relaciona-se com a capacidade de desenvolver e sustentar vantagens competitivas. Segundo Porter (1989, p. 4) “a vantagem competitiva surge fundamentalmente do valor que uma empresa consegue criar para seus compradores e que ultrapassa o custo de fabricação pela empresa.” Consolidado por Al Ries e Jack

Trout, ainda nos anos 1960, o conceito de Posicionamento aponta para a construção de uma percepção de marca na mente dos consumidores. Desta maneira, o Posicionamento faz com que os consumidores percebam características exclusivas e atribuam à marca em questão, diferenciais que a tornam memorável. Ries e Trout (2009) indicam, ainda, que a comunicação é, em última instância, o fator que constrói estas percepções. Nesse sentido, a percepção do posicionamento de mercado e a construção e disseminação desse posicionamento pode ser inferido pela informação produzida e disseminada por usuários nas mídias sociais.

Para entender o posicionamento de mercado a partir das mídias sociais é preciso primeiramente refletir sobre o fluxo de produção e a consequente disponibilidade de dados que ocorre atualmente, especialmente os dados não estruturados. Sobre este fato Stephens- Davidowitz (2018, p. 15) ressalta: “Pense em toda a informação que é disseminada pela internet em um determinado dia – temos um número, na verdade, para o quanto de informação existe. Em média, em um dia no início do século XXI, os seres humanos geraram 2,5 milhões de terabytes de dados. E esses bytes são pistas”. Segundo Amaral (2016, p. 33), dados não estruturados “não possuem qualquer tipo de estrutura definida, muitas vezes não possuem sequer qualquer tipo de metadado e ao mesmo tempo possuem e esses metadados não são de muita ajuda para os processos de análise. também são compostos por imagens, vídeos arquivos de áudio, plantas de engenharia”.

O advento das mídias sociais, dos dispositivos móveis e da constante expansão das tecnologias de acesso a internet favorecem uma realidade de consumidores conectados e ativos no processo de produção de dados. Tal contexto é fortemente relacionado com o que é dito por Kotler, Kartajaya e Setiawan (2017, p. 12): “Em um mundo altamente tecnológico, as pessoas anseiam por um envolvimento profundo. Quanto mais sociais somos, mais queremos coisas feitas sob medida para nós. Respaldados pela análise de big data (coleta, processamento e análise de megadados), os produtos tornam-se mais personalizados e os serviços, mais pessoais.”. É preciso destacar o papel transformador da sociedade em amplo sentido com o advento das redes sociais, considerando que em média, 38% dos usuários destas plataformas passam mais de quatro horas diárias conectados nestes ambientes (DINIZ, Natalha Herminio *et al*, 2020).

Como contribuição aos processos de inteligência de mercado e comunicação empresarial, este trabalho apresenta a modelagem de um processo de classificação preditiva de dados e a concepção de um modelo matemático de comparação entre concorrentes, tendo como base cálculos de similaridade entre curvas. Dessa forma é

possível estabelecer um ranque entre um grupo de marcas com relação a similaridade da percepção do público, no âmbito das vantagens competitivas.

Para a realização deste trabalho selecionou-se o mercado de instituições de Ensino superior privado brasileiro. Trata-se de universidades, centros universitários ou faculdades. Deste mercado foram elencadas as vantagens competitivas formadoras da percepção de qualidade do mercado considerando os seguintes aspectos: ensino; estrutura física; atividades de extensão; atividades relacionadas à inovação; preço das mensalidades. Cada um destes fatores se transmuta em iniciativas de comunicação para o público externo dessas instituições.

Este estudo tem como questão de pesquisa: quais instituições geram percepções similares de posicionamento de mercado, diante de vantagens competitivas fundamentais para um dado setor? O objetivo de pesquisa é analisar e comparar o posicionamento de mercado, considerando as vantagens competitivas relacionadas ao setor de instituições de ensino superior privado. Este estudo contribui com um método que pode ser replicado em diferentes setores.

## 2 REFERENCIAL TEÓRICO

A importância da conectividade é notória para o marketing do novo milênio. Empresas que não estiverem conectadas, perderão sua relevância junto aos consumidores. Assim, as transformações a partir do final dos anos 90 demandaram uma nova transformação dos processos. Transformação essa que o Marketing, a comunicação e as empresas passaram a vivenciar a partir da consolidação dos meios digitais. O que antes se impunha através da intervenção humana, agora passa, necessariamente, pelo processamento dos dados gerados em rede. Ainda apontado por Gabriel (2010, p. 74): “Conforme as tecnologias digitais passam permear cada vez mais as atividades humanas, mais influência digital passa a ter no marketing”.

O conceito de posicionamento aponta para a construção de uma percepção de marca na mente dos consumidores (Ries e Trout, 2009). Desta maneira, o posicionamento faz com que os consumidores percebam características exclusivas e atribuam à marca em questão, diferenciais que a tornam memorável. Ries e Trout (2009) indicam, ainda, que a comunicação é, em última instância, o fator que constrói estas percepções. No entanto, embora o conceito permaneça absolutamente válido, as formas de construção transformaram-se ao longo dos anos 2000. Redes sociais digitais surgiram e, com elas, um novo tipo de abordagem na comunicação. Expressões como inbound marketing, que

englobam ações como Search Engine Optimization - SEO (Otimização dos Mecanismos de Busca), Marketing de Conteúdo e as estratégias de abordagem nas redes sociais, passaram a ser determinantes para a efetividade no processo de construção e percepção das marcas.

Percebido como um campo novo onde ciências exatas e ciências humanas encontram pontos de convergência, a mineração de dados em redes sociais, quando utilizada cumprindo os protocolos de privacidade acordados na própria rede social, é uma poderosa ferramenta de pesquisa e de grande impacto nas ações de Marketing. Thibes aponta (2018) que os algoritmos "são uma série de instruções passo-a-passo que descrevem explicitamente várias operações." e, por isso mesmo revelam-se de grande utilidade no campo do Branding, do Marketing e, mais especificamente, no universo da Comunicação.

Na era do *big data* é primordial para os processos de estratégia em comunicação empresarial apropriar-se de dados não estruturados, públicos e produzidos sobre estas empresas. Mais do que isso: relacionar estes dados com as vantagens competitivas de mercados específicos, com as métricas de audiência digital, geradas no âmbito da web social, considerando problemáticas de comunicação. Trata-se de um momento em que à comunicação empresarial e diversas outras áreas passam a considerar os dados como parte primordial dos processos estratégicos (MAYER-SCHÖBERGER e CUKIER, 2013).

Neste artigo, busca-se apontar que, com o suporte da mineração de dados (Data Mining). em grandes volumes de dados (Big Data) é possível encontrar um ponto de convergência entre o que é produzido pelos usuários em redes sociais e o processo de criação de identificação de Vantagens Competitivas e Comparativas para impactá-los. Ou seja, a própria informação gerada pelos usuários, contribui para a construção do Posicionamento das marcas de qualquer setor.

### **3 MÉTODO DA PESQUISA**

O método de classificação proposto relaciona postagens na mídia social Instagram com cada uma das vantagens competitivas citadas, geradas pelos perfis oficiais das instituições de ensino analisadas. Foram selecionadas três instituições de ensino superior privado atuantes no Ceará, entre as sete instituições com maior volume de alunos no estado. Convencionou-se chamar as três instituições de Instituição A, Instituição B e Instituição C. Os dados foram coletados entre os meses de novembro de 2017 e março de 2019, totalizando 792.674 (setecentos e noventa e duas mil seiscientos e setenta e quatro)

interações analisadas. Todos os dados foram coletados através de requisições à API pública da mídia social Instagram, obedecendo todas as premissas de privacidade e aplicação de processos de anonimização de dados nos processos de análise.

“Para que uma máquina preditiva mude sua estratégia, alguém tem que criar uma que seja útil para o seu caso específico. Isso depende de várias coisas que estão fora do controle da sua organização.” (Agrawal; Gans; Goldfarb, 2019, p. 158). É exatamente considerando a necessidade de criar processos preditivos customizados, para mercados específicos, que o modelo de classificação desenvolvido neste trabalho se insere, onde modelos preditivos de classificação de dados tornam-se um aspecto inerente à inteligência de mercado e às estratégias de comunicação.

Do processo de coleta foram constituídos grupos de palavras prioritárias para cada uma das vantagens elencadas. Trata-se de palavras extraídas das postagens no Instagram, das instituições de ensino e que fazem clara referência a cada uma das vantagens competitivas. O processo de definição destes grupos baseou-se na técnica de bag of words, definida por Provo e Fawcet (2016, p. 254) como “tratar cada documento como uma coleção de palavras individuais. Essa abordagem ignora a gramática, a ordem das palavras, a estrutura da frase e (geralmente) a pontuação. Ela trata cada palavra em um documento como uma palavra-chave potencialmente importante”. Estes grupos prioritários foram utilizados para treinar os modelos preditivos de classificação de postagens. Posteriormente foi feito uso de técnicas de mineração de texto, utilizando um processo de morfologia, atribuindo a classe gramatical para cada uma das palavras utilizadas nas postagens analisadas. Após esta etapa, utilizou-se um processo de filtragem, para gravar em um banco de dados apenas as palavras classificadas como verbos, adjetivos ou substantivos, reduzidas aos seus respectivos “lemas”.

Foi desenvolvido um modelo de classificação de dados não estruturados utilizado a ferramenta RapidMiner, tendo por base o algoritmo Suport Vector Machine. Segundo Amaral (2016, p. 44) “Máquinas de vetores de suporte, em inglês Suport Vector Machine, são um algoritmo de classificação que maximizam as margens entre as instâncias mais próximas, dessa forma, é criado um vetor otimizado que é então utilizado para classificar novas instâncias.”.

Os resultados obtidos indicam, para cada postagem analisada, a referência relacionada a cada uma das vantagens competitivas analisadas. Uma classificação “TRUE”, indica que uma postagem faz alusão a uma determinada vantagem competitiva.

Uma classificação do tipo “FALSE”, indica que não há referência na postagem analisada e a vantagem analisada.

Para analisar as postagens, foi aplicada técnica de reconstrução textual das postagens, usando apenas os verbos, substantivos e adjetivos de cada postagem e reduzindo cada palavra aos seus respectivos lemas, através de um processo de lematização. Nesta etapa do trabalho, utilizou-se a funcionalidade de morfologia da API de processamento de linguagem natural Rosette, acessada através do software RapidMiner.

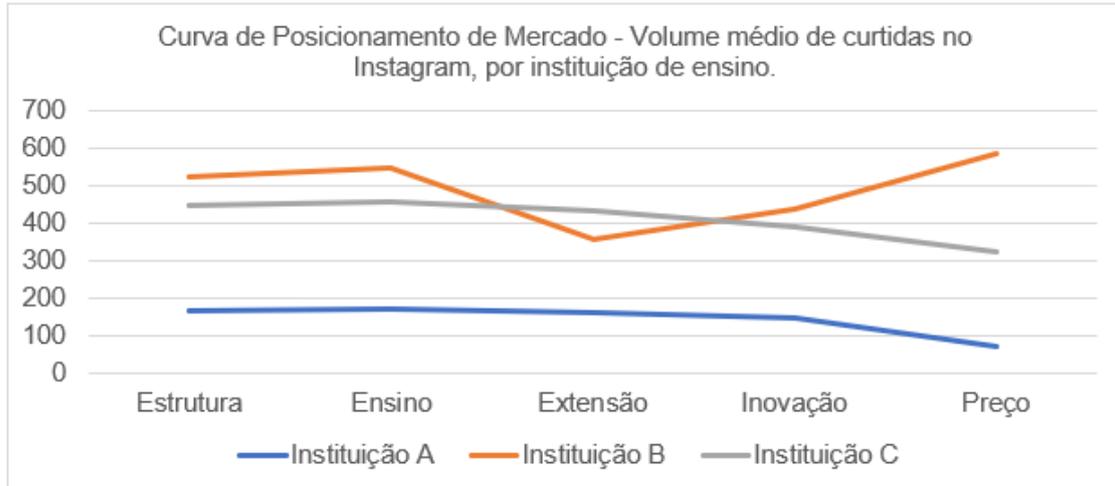
O processo de reconstrução das postagens descrito acima tem por objetivo aumentar a acurácia do processo de classificação. Considerando o comportamento dos cinco modelos de classificação, chegou-se a uma taxa de acerto média de classificação de postagens de 97,64%, considerando uma amostra previamente classificada.

O método de análise inclui a definição da curva de posicionamento de mercado das instituições e o cálculo da similaridade entre as curvas com o objetivo que ranquear as instituições por similaridade de posicionamento de mercado.

#### **4 RESULTADOS E DISCUSSÃO**

Inicialmente no processo de classificação desenvolvido, as postagens foram separadas considerando suas classificações para cada uma das vantagens competitivas. Utilizou-se então o cálculo de média da métrica nativa da mídia social Instagram, curtidas em uma postagem. A métrica referida relaciona-se com o nível de interesse positivo do público com relação a uma determinada postagem. A análise concentra-se na reação da audiência, com relação aos aspectos que fazem parte do discurso corrente destas marcas, levantando o interesse do público com relação aos temas propostos. O resultado é apresentado na Figura 1.

Figura 1 – Curva de Posicionamento de Mercado.



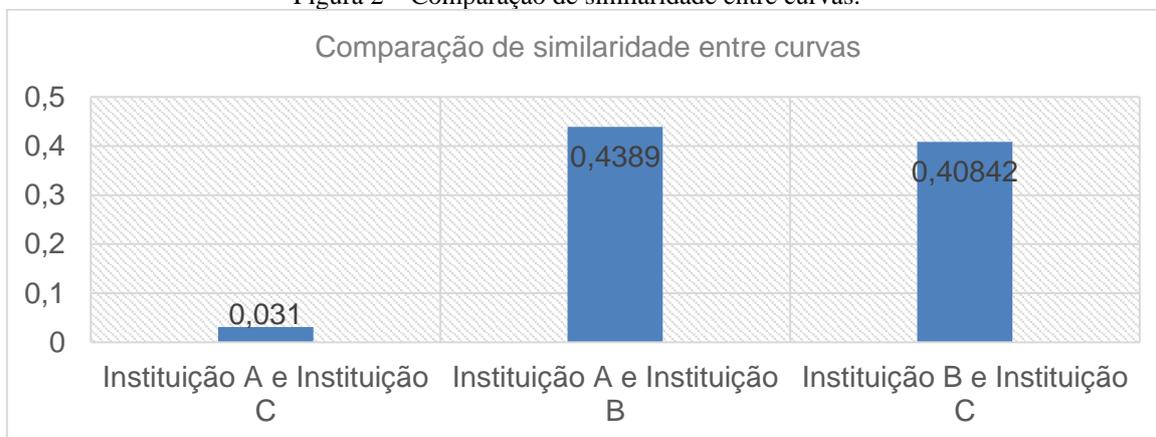
Este trabalho não tem como foco a comparação entre os volumes “curtidas” entre as instituições. A questão que se coloca relaciona-se com o contexto da percepção externa, com relação às vantagens competitivas. Trata-se de compreender, considerando a percepção do público externo, e responder de forma precisa à pergunta: quais instituições geram percepções similares de posicionamento de mercado, diante de vantagens competitivas fundamentais para um dado setor?

Para tanto, faz-se necessário compreender, dentre um grupo de instituições, qual delas tem uma distribuição dos níveis médios de curtidas em uma postagem semelhantes, comparada com as outras instituições, com relação às vantagens analisadas. Dessa forma é possível estabelecer um ordenamento preciso entre concorrentes com relação ao posicionamento de marca percebido. Esse ordenamento considera a comparação similaridade entre curvas, tendo por base o gráfico apresentado na Figura 1. As curvas que serão comparadas são denominadas “Instituição A”, “Instituição B” e “Instituição C”. Cada curva é composta por cinco pontos, os quais representam cinco características analisadas. Para permitir o uso de operações matemáticas com estas curvas, associamos um número para cada uma dessas características e, como cada característica leva a apenas um número em cada curva, podemos associar essas curvas a funções. Matematicamente, estas curvas agora são representadas por  $y_{Instituição A}(x)$ ,  $y_{Instituição B}(x)$  e  $y_{Instituição C}(x)$ , respectivamente, enquanto as características são agora representadas por estrutura ( $x = 1$ ), ensino ( $x = 2$ ), extensão ( $x = 3$ ), inovação ( $x = 4$ ), e preço ( $x = 5$ ). Note que a numeração escolhida é arbitrária: variar esses números não altera qualitativamente os resultados.

Primeiramente, deve-se minimizar a possível influência do intervalo de valores das curvas sobre o resultado. Em outras palavras, deve-se apenas em comparar as formas das curvas, não os valores em si. Para isso, comparou-se as derivadas das curvas. A derivada representa o quanto varia uma função quando mudamos o seu argumento  $x$ . Comparar as derivadas das funções em questão, então, é basicamente comparar suas formas. Em um segundo passo, criou-se uma espécie de intervalo de valores universal; um intervalo como  $[0, 1]$ , tal que um valor próximo de 0 (1) represente sempre uma alta (baixa) similaridade entre as funções. Para isso, então, não se considera as derivadas das funções em si, mas o arco de tangente da derivada. O arco de tangente de um valor  $y$  é um ângulo cuja tangente é igual a  $y$ . Sendo um ângulo, ele é naturalmente restrito ao intervalo  $[-\pi/2, \pi/2]$ . Dividiu-se então esse resultado por  $\pi/2$  para fazer com que todas as curvas a serem comparadas estejam no intervalo  $[-1, 1]$ . As funções que resultam de todo esse processo passam a ser  $y_{Instituição A}(x)$ ,  $y_{Instituição B}(x)$  e  $y_{Instituição C}(x)$ . Finalmente, a comparação entre as curvas se dá pela operação  $\langle |Y_a(x) - Y_b(x)|/2 \rangle$ , ou seja, a média do módulo da subtração entre duas curvas  $a$  e  $b$  (onde  $a, b =$  Instituição A, Instituição B ou Instituição C) dividido por dois. O módulo serve para analisar apenas o valor absoluto da diferença (sem sinal positivo ou negativo). A divisão por 2 serve para corrigir o fato de que duas funções podem estar separadas por um valor máximo 2 (quando uma delas é  $-1$  e a outra  $+1$ ), de forma que agora a operação terá resultados apenas entre 0 e 1, qualquer que seja o  $x$ . A média então é calculada sobre todos os valores de  $x$  e, com isso, obtêm-se a comparação entre as curvas  $a$  e  $b$ .

Os resultados obtidos com esse procedimento mostram a comparação entre as instituições e são apresentados na Figura 2.

Figura 2 – Comparação de similaridade entre curvas.



Analisando a Figura 2, observa-se que a relação de similaridade entre as curvas de posicionamento é maior entre as instituições A e a instituição C, com o menor fator de comparação entre as curvas de similaridade. O maior fator de comparação de similaridade entre as curvas é verificado entre as instituições A e a instituição B. É importante destacar que o modelo proposto permite a comparação entre um grupo maior de instituições, marcas ou empresas, estabelecendo um ordenamento preciso entre um grupo de marcas analisadas, tendo por base o intervalo  $[0, 1]$ , definido anteriormente.

## 5 CONCLUSÃO

Os desafios impostos pela era do *big data* já transcendem o campo da tecnologia da informação. Este trabalho relaciona três aspectos principais, a citar: modelos preditivos, baseado em tecnologias estabelecidas, aplicadas de forma customizada, no sentido de conseguir classificar dados de um mercado específico, no contexto de suas principais vantagens competitivas. Desta etapa, fica aplicabilidade para qualquer outro mercado, dado que o ponto central é a perspectiva de relacionar as vantagens competitivas com um modelo preditivo, tendo por base um grupo de palavras, igualmente específicas, aumentando a acurácia do modelo. Em segundo, o estabelecimento de uma relação entre as métricas em mídias sociais com o posicionamento de mercado das marcas. E em terceiro, uma modelagem matemática que permite de forma precisa estabelecer um ordenamento considerando a similaridade de posicionamento percebido de um grupo de marcas. O processo de comparação de curvas é fundamental, pois evita comparações imprecisas e permite que um grupo de marcas concorrentes possam ser relacionadas entre si, com relação ao posicionamento percebido pelo público externo.

A implementação do modelo proposto neste trabalho representa uma forma de compreender mercados e moldar estratégias de marketing e comunicação mais precisas e eficientes. Neste sentido, compreender a relação do público consumidor, com relação às principais vantagens competitivas ou comparativas de setores específicos, permite que as empresas compreendam como criar e combinar vantagens, gerando ativos, atividades ou recursos que gerem uma percepção única para uma dada marca, perante seu público.

## REFERÊNCIAS

AGRAWAL, Ajay; GANS, Joshua; GOLDFARB, Avi. Máquinas Preditivas: A Simples Economia da Inteligência Artificial. Rio de Janeiro: Alta Books, 2019. 257 p.

AMARAL, Fernando. Introdução à Ciência de Dados: Mineração de Dados e Big Data. Rio de Janeiro: Alta Books, 2016. 320 p.

DINIZ, Natalha Herminio et al. A utilização do marketing digital na promoção e divulgação dos serviços de fisioterapia. **Brazilian Journal of Development**, v. 6, n. 9, p. 72186-72195, 2020.

GABRIEL, M. Marketing na Era Digital: Conceitos, Plataformas e Estratégias - São Paulo, Novatec Editora, 2010.

GUPTA, Sunil. Implantando Estratégia Digital: Guia Para Reinventar Sua Empresa. São Paulo: M. Books, 2018. 296 p.

KOTLER, Philip; KARTAJAYA, Hermawan; SETIAWAN, Iwan. Marketing 4.0: Do Tradicional ao Digital. Rio de Janeiro: Sextante, 2017. 208 p.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. Big Data: Como Extrair Volume, Variedade, Velocidade e Valor da Avalanche de Informação Cotidiana. Rio de Janeiro: Elsevier, 2013. 176 p.

PORTER, Michael E. Vantagem Competitiva: Criando e Sustentando um Desempenho Superior. 15. ed. Rio de Janeiro: Campus, 1989. 512 p.

RIES, Al & TROUT, Jack - Posicionamento: a Batalha por sua Mente 2009 - São Paulo - M. Books do Brasil Ltda.

PROVOST, Foster; FAWCETT, Tom. Data Science Para Negócios: O que Você Precisa Saber Sobre Mineração de Dados e Pensamento Analítico de Dados. Rio de Janeiro: Alta Books, 2016. 383 p.

STEPHENS-DAVIDOWITZ, Seth. Todo Mundo Mentira: O Que A Internet E Os Dados Dizem Sobre Quem Realmente Somos. Rio de Janeiro: Alta Books, 2018. 335 p.

THIBES, V. (2018) - Afinal, o que é um algoritmo e o que isso tem a ver com computação? - Recuperado a partir de <https://canaltech.com.br/produtos/Afinal-o-que-e-um-algoritmo-e-o-que-isso-tem-a-ver-com-computacao/> - acessado em 29/05/2019.

ZENGER, Todd. Muito Além da Vantagem Competitiva: Como Resolver o Dilema de Manter um Crescimento Sustentável e Simultaneamente Criar Valor. São Paulo: M. Books, 2016. 192 p.