



Universidade Presbiteriana Mackenzie



Universidade Presbiteriana Mackenzie

Pró-reitora de Pesquisa e Pós-graduação

Faculdade de Computação e Informática

Pós-Graduação *Lato Sensu*

Análise da viabilidade de prever o comportamento decisório futuro do leitor assíduo, fundamentado na Inteligência Artificial.

Samuel Delmonte

São Paulo

2022



Universidade Presbiteriana Mackenzie



Samuel Delmonte

Análise da viabilidade de prever o comportamento decisório futuro do leitor assíduo, fundamentado na Inteligência Artificial.

Trabalho de Aplicação de Conhecimento apresentado ao Curso de Pós-Graduação em Espec. em Inteligência Artificial da Universidade Presbiteriana Mackenzie para a obtenção do título de Especialista em Espec. em Inteligência Artificial.

Orientador: Prof. Me. Vladimir Aparecido Esteves

São Paulo

2022



Resumo

O presente trabalho tem como objetivo, investigar o mercado editorial de livros, relacionando-o aos baixos e persistentes índices médios de leitura do brasileiro, assim como a recente e retumbante derrocada das grandes empresas do setor que, aparentemente, seguem sem rumo nenhum. E mais, muito mais do que um simples diagnóstico, há também, a aposta dos autores, em encontrar o indício correto, que emerja a solução para o grave quadro de desprezo ao livro em si, observado claramente em território brasileiro. A hipótese, aqui defendida, diz respeito à possibilidade de prever a compra futura de certo cliente, de acordo com seus hábitos de leitura, resultando a posteriori, na fidelização e aumento qualitativo do consumo de livros da editora capaz de viabilizar um modelo preditivo. Para tal, foi necessário instaurar uma metodologia descritiva, com base em estudo de casos, robusta o bastante, para filtrar e abstrair informações que alimentem uma rede neural construída no paradigma de aprendizado de máquina, mais precisamente no modelo estatístico de regressão logística, codificado em python, visando sugerir as compras futuras dos clientes analisados. Diante dos resultados, concluiu-se haver indícios significantes que corroboram o poder preditivo da hipótese, portanto, os autores sugerem a continuidade das pesquisas e investigações baseadas nos paradigmas de previsões temporais, com base nas máquinas computacionais capazes de aprender.

Palavras-chave: inteligência artificial, aprendizado de máquina, regressão logística, python



SUMÁRIO

INTRODUÇÃO	8
1 RELATÓRIO DA SITUAÇÃO	
1.1 A empresa.....	10
1.2 Desafio.....	10
1.3 Os sintomas	11
1.4 Objetivo	12
2 DIAGNÓSTICO	
2.1 As informações	13
2.2 Análise e diagnóstico	22
3 SOLUÇÃO	
3.1 Propostas de Solução	27
3.2 Conexão Proposta com os Resultados Esperados	33
4 PLANEJAMENTO	
4.1 Planos de ações	38
5 CONSIDERAÇÕES FINAIS.....	41
REFERÊNCIAS	43
APÊNDICES	45



LISTA DE QUADROS

Quadro 1 – Proposta de solução A: Modelo de regressão preditiva de vendas futuras..... 40



LISTA DE FIGURAS

Figura 1 - Capa da publicação de 2019 – Retratos da Leitura no Brasil.....	17
Figura 2 - Percentual e estimativa Populacional.....	18
Figura 3 - Média de livros lidos.....	19
Figura 4 - Normalização em Python.....	20
Figura 5 - As três médias dos livros lidos, em python.....	21
Figura 6 - Percentual de compradores.....	22
Figura 7 - Livros lidos e comprados.....	23
Figura 8 - Comparação Cenário 01 vs Cenário 02.....	26
Figura 9 - Consumo de livros do cenário 1.....	27
Figura 10 - Comparação Cenário 01 vs Cenário 02 II.....	28
Figura 11 - Arquivos pessoais do autor.....	33
Figura 12 - a - Simuladores estatísticos do site; 12 -b - Livro do autor publicado.....	35
Figura 13 - Protótipo regressão logística 01.....	37
Figura 14 - Protótipo regressão logística 02.....	38



LISTA DE TABELAS

Tabela 1 - Critério de leitura.....	18
Tabela 2 - Especificações técnicas.....	19
Tabela 3 - Compras de julho, agosto e setembro de 2019.....	24
Tabela 4 - Os valores da dispersão de dados dos compradores.....	25
Tabela 5 - Redução progressiva da amostra.....	32
Tabela 6 - Comparação final de ambos os cenários.....	33



INTRODUÇÃO

Os livros que se acumulam, continuamente, no setor daqueles que já foram lidos, da estante de um leitor entusiasta, são grosso modo, seu repertório adquirido, aquilo que ele sabe de fato; das suas leituras, extrai-se o tempo que gastou em busca dos seus interesses, o livro retrata, de quem o lê, aquilo que ele se tornou como indivíduo, manifesta também, aquelas atribuições que ele gostaria de ter, mas, somente o legente enxerga a contradição existente entre essas duas expressões (aquilo que o leitor é, e aquilo que gostaria de ser), gerada dentro de si próprio. O livro é, sobretudo, a única possibilidade do leitor, de usufruir da companhia dos autores que já não vivem, é poder conviver com seus mestres, e quiçá, superar algum deles.

De acordo com a FIPE (Fundação Instituto de Pesquisas Econômicas) (FIPE, 2018), o mercado editorial de livros, sofre desde 2014, vertiginosa redução de sua produção e quantidade de vendas, sobretudo, dos livros no formato físico. Possivelmente, a aparente expansão econômica, desfrutada pelas editoras e varejistas do setor, durante a primeira década do século XXI, as impediu de prevenir erros nas suas estratégias de inserção no mercado, aqui discutido, de acordo com o perfil específico dos consumidores de livros. Eles, os erros, foram implacáveis, e suscitaram a falência de editoras e lojas varejistas líderes, até pouco, do mercado.

Parece haver, pouco conhecimento, no conteúdo disponível na literatura do setor editorial; ademais, pode-se verificar empiricamente, que o livro não segue regras do livre mercado, dito neoliberal, capaz de autorregular seus parâmetros de oferta e demanda. Além disso, nota-se generalização, em demasia, das variáveis analisadas estatisticamente, a respeito dos hábitos de leitura no Brasil; não há, portanto, dados coletados, devidamente abstraídos ao que é importante, capazes de semear base de dados, composta por perfis robustos.

Diante deste cenário, a jovem editora Rebelião Intelectual, promoveu uma investigação alternativa, cujos novos indícios encontrados, viabilizam o aumento de consumação dos seus clientes, gerando à organização excedente no percentual de lucro.

O presente trabalho é resultado desta investigação, edificada em premissas cujo paradigma difere do *status quo* dos trabalhos anteriores, de tal modo que são desprezadas quaisquer intervenções do mercado, seus imperativos categóricos que prometem, desde o **Iluminismo (SOUZA)**, no século XVIII, prosperidade econômica na sociedade futura, tampouco serão levados em conta; portanto, a pesquisa aqui realizada, concentrou-se na dor em si do problema, identificando-a como a necessidade de aumentar a consumação dos clientes já existentes, através da predição de qual será o próximo título de livro a ser adquirido por certo



cliente. A hipótese última aqui defendida, resulta na mensuração da acurácia de predizer o quão significativas são, as citações e referências do livro que o leitor lê no tempo presente, na escolha do próximo volume. Fato que terminou examinado pela pesquisa, como verossímil.

E, para isso, desenvolveu-se uma base de dados, com informações robustas dos hábitos e comportamentos de leitura dos leitores que já são clientes, cujos testes e treinamentos, mensuraram uma acurácia superior a 70%, em relação ao seu poder de predição. O êxito foi atingido através de um algoritmo calibrado para aprender e predizer o desejo futuro de leitura do ser humano.

A fundamentação analítica deste trabalho, apoia-se na **metodologia** da pesquisa descritiva, cujo **procedimento técnico** a aplicar à coleta de dados, é um estudo de caso.

Por final, este trabalho foi estruturado em 5 partes: A 1ª se refere ao “Relatório da Situação” que abordará o cenário atual da empresa editora Rebelião Intelectual com relação ao uso dos dados. A 2ª parte se refere ao “Diagnóstico”, que explicará como se deu a coleta das pesquisas e entrevistas. A 3ª parte apresentará a “Solução”, propondo uma abordagem de governança para uso de informação dentro da empresa. A 4ª parte apresentará os “Planos de Ações” referentes a como implementar essa governança. Por final, a 5ª parte serão apresentadas as considerações finais deste estudo.



1 RELATÓRIO DA SITUAÇÃO

O objetivo deste capítulo é abordar os seguintes tópicos: (1) breve descrição da empresa usada como objeto do modelo de ambiente a ser estudado; (2) apresentar qual é o desafio, aqui proposto, cuja implicação é predizer qual será a decisão de compra futura do cliente; (3) promover simbiose da empresa, que arrecada lucro sazonal, à medida que são satisfeitos os desejos de compra do cliente; (4) discorrer sobre os sintomas gerados em uma sociedade com baixíssimos índices de utilização do livro; (5) contemplar sobre as reações sintomáticas, despendida a ambos (cliente e provedor de serviços), pela ineficiência e ignorância de estratégias de mercado das empresas deste setor; e (6) os objetivos que motivaram a confecção o deste trabalho.

1.1 A empresa

A Rebelião Intelectual, é uma editora, de porte médio, de livros impressos, com serviços de edição e impressão de obras literárias, serviço de aquisição parcial de direitos autorais; serviços de tradução de obras inéditas no português e serviço de livros virtuais específicos. Processo de vendas no varejo, diretas ao consumidor; seja por lojas físicas da própria empresa, seja por plataforma digital de e-commerce. O modelo de produção da Rebelião Intelectual será sob demanda (parcialmente inspirada no Toyotismo); e, em alguns casos, *Taylormade* (Sob medida).

1.2 Desafio

Há cinco anos, o mercado de livros brasileiro, evoluiu de um quadro já agônico, para um coma terminal. As razões para tal, evidenciam-se nas falências, durante este período, das três maiores editoras do país: Abril, Saraiva e Escala (onde trabalhei como programador front-end das plataformas de e-commerce, e será referência), além de gigantes do varejo, como a Livraria Cultura.

A conjectura atual como um todo, pressupõe a piora, do já agônico cenário, afinal somos assombrados pelas paranoias ideológicas de protestantes e evangélicos, notórios inimigos da ciência, do intelecto e do pensamento mais puro, o autônomo.



A priori, não é difícil perceber que a hipótese aqui defendida, é incapaz de aumentar o número de leitores, que na verdade, deveria ser competência da Educação.

O que se defende é a sofisticação dos serviços aos leitores já existentes, grosso modo, maior variedade de títulos de livros, de acordo com a métrica de leitura de cada um, com edição sob demanda.

Para tal, faz-se necessário introduzir redes neurais que aprendam a classificar, atributos de cada leitor, podendo prever para cada um, quais títulos faltam no repertório da Editora e teriam grandes chances de serem adquiridos por este cliente. A calibração do algoritmo, também classificará possíveis aquisições de obras sem edição no Brasil, que seriam comercializados em português e no formato e-book.

A inteligência Artificial criará *leads* publicitários, capazes de ofertar, individualmente, livros que foram classificados pela nossa rede neural, como portadores de alta acurácia em relação à predição de compras que serão feitas pelo cliente de dado perfil.

1.3 Os sintomas

De acordo com a última edição do senso sobre hábitos de leitura no Brasil, coordenado pelo Instituto Pró-livro (**RETRATOS..., 2020**), é considerado um leitor, aquele que, nos últimos três meses, leu um livro, inteiro ou em partes. Com base nesse parâmetro, estimou-se que, numa população de cem milhões de habitantes maior de cinco anos, 52 por cento são leitores, e os outros 48 não.

Dos livros lidos, incluem-se a bíblia e os didáticos (que nem sempre são lidos por vontade própria), tornando os resultados, um tanto, imprecisos e preocupantes.

Tais dados são um sintoma direto do insucesso educacional. A falência do ato de adquirir instrução através da leitura, é mais um sintoma de que não houve, não há, e provavelmente não haverá prosperidade nos índices de desenvolvimento humano, em terras brasileiras.

Ademais, a ineficiência comercial das editoras e varejistas, impossibilitou a fidelidade de seus clientes, afinal eles migraram para outras empresas em busca dos livros de interesse, sendo um infeliz sintoma, da ruína econômica, dos que vendem, e a



dependência das grandes varejistas online, daqueles que compram.

1.4 Objetivo

O objetivo último desta dissertação é prever a decisão de compra futura do leitor, com base nas referências e citações do seu último livro lido.

Para este fim, são cumpridos os seguintes objetivos específicos:

1. Conceituar e abstrair certa coleta de dados, capaz de se tornarem informação robusta, através de perfis de leitores
2. Conceituar e definir um trecho de código em Python, com base no aprendizado de máquina, com características preditivas.
3. Definir os primeiros passos da calibração da rede neural, através das prioridades de rastreamentos, que ensinarão ao algoritmo, à medida que o tempo passa, a respeito dos hábitos de leitura de cada cliente inserido na base de dados.
4. Verificar a viabilidade desta hipótese, mensurando sua acurácia.



2 DIAGNÓSTICO

A finalidade deste capítulo é discorrer a respeito da metodologia escolhida para testar a hipótese, aqui defendida, esclarecer sobre como foi feita a coleta de dados, para então, analisar e diagnosticar o *status quo* (situação/ordem atual) do desempenho econômico da livraria Rebelião Intelectual, quando comparado, com os hábitos e comportamentos do leitor brasileiro.

A fundamentação analítica deste trabalho, apoia-se na **metodologia** da pesquisa descritiva, cujo **procedimento técnico** a aplicar à coleta de dados, é um estudo de caso.

Assim como o destaque de Jim Bergeson (**Frases Estatísticas, 2021**), mencionando que os dados falarão a quem estiver disposto a ouvi-los, e a definição de Silva & Menezes (**Silva & Menezes, 2000**), “a pesquisa descritiva visa descrever as características de determinada população ou fenômeno ou o estabelecimento de relações entre variáveis.”, apontam aos fins da pesquisa descritiva, aludem também, ao escopo do presente trabalho, cujo objetivo é investigar, detectar e aprimorar alguma circunstância fenomenológica, existente entre as variáveis abstraídas, capaz de instaurar o insucesso do livro (leitura, venda, oferta de exemplares internacionais traduzidos etc.), em mais de um cenário.

Portanto, decidiu-se adotar, como procedimento técnico, um estudo de caso, explanado por Oliveira (**Oliveira, 2002**) como um método, competente o suficiente para, identificar e analisar as múltiplas ocorrências de um mesmo fenômeno, em vários casos. São das palavras de Oliveira, que se extraiu o critério utilizado para adotar o estudo de caso, como técnica de coleta e organização dos respectivos dados, em duas perspectivas, aparentemente diferentes.

2.1 As informações

As informações de natureza conceitual e teórica, cuja composição conduziu o desenvolvimento deste trabalho, sobre os alicerces de um estudo de caso descritivo, foram



extraídos das literaturas disponíveis, na **WEB**, do repositório digital da Universidade Presbiteriana Mackenzie, cuja sugestão se encontra nas referências bibliográficas desta disciplina, além do repertório pessoal de estudos do autor.

A aquisição destas informações, acima relatadas, foram valiosas o bastante, para servirem como um meio, do qual fosse possível detectar, quais são as razões de existência da raiz do problema, aqui discutido, e grosso modo, definido como o insucesso da livraria Rebelião Intelectual, em aumentar seu volume de vendas, e conseqüentemente, sua margem de lucro.

Para tanto, procedeu-se com a coleta de duas amostras, ambas mensuradas em cenários diferentes; não obstante, são elas contemporâneas na dimensão temporal, e suscetíveis, segundo a tese aqui defendida, à influência de fenômenos circunstanciais, capazes de tornar o desempenho de cada cenário pertencente ao conjunto dos livros, insatisfatório.

Partindo, a priori, do pressuposto que ambas as amostras, resultam em evidências claras de desempenho negativo, coletou-se, no **primeiro cenário**, dados secundários, contidos na quinta e última edição do censo **Retratos da leitura no Brasil, 2019 (Figura 1)**.

A finalidade é garimpar as informações relevantes, do conjunto de dados dos indivíduos leitores, abstraindo-os à uma verossímil e adequada comparação com o cenário dois. Dados estes que, aqui no cenário um, serão extraídos, quase na sua totalidade, aplicando-se uma coleta indireta, pois, trata-se de dados secundários, previamente tratados pelo autor.

Figura 1 - Capa da publicação de 2019



Fonte: Retratos da leitura no Brasil (2019, capa)

2.1.1 Primeira Verificação

Refere-se ao desempenho de leitura do brasileiro, em geral, como descrito na **Tabela 1**. Nota-se que os autores do censo consideraram um leitor, aquele que leu, ao menos, um livro, inteiro ou em partes, nos últimos três meses.

Tabela 1 - Critério de leitura

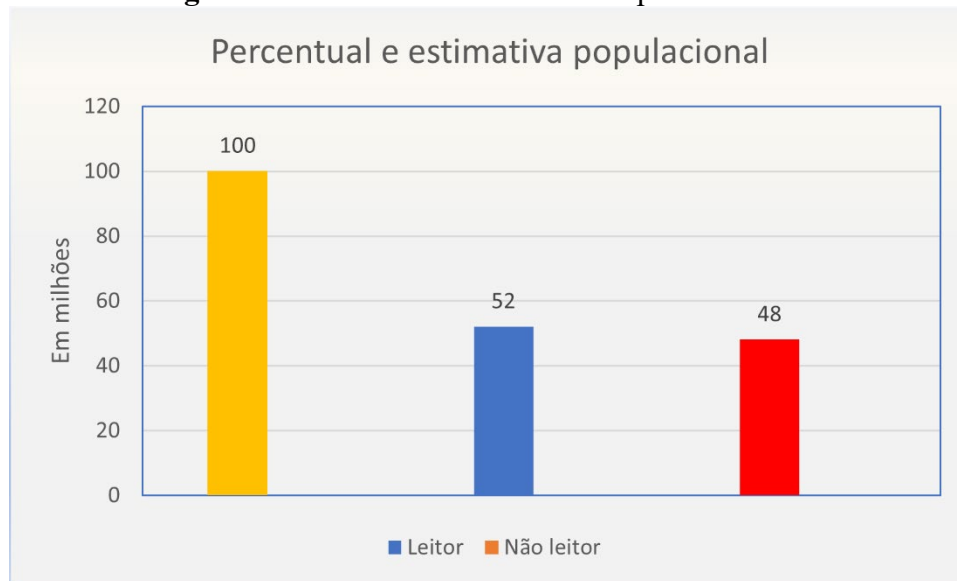
Critério de leitura	
Leitor	Não leitor
É aquele que leu, inteiro ou em partes, pelo menos 1 livro nos últimos 3 meses.	É aquele que declarou não haver lido nenhum livro nos últimos 3 meses.

Fonte: Retratos da leitura no Brasil (2019, p. 19)

Na figura 2, é possível verificar que a pesquisa estimou uma população total (barra amarela), de cem milhões de indivíduos, com um percentual de 52% para os que são leitores (barra azul), e 48% aos não leitores (barra vermelha).



Figura 2 - Percentual e estimativa Populacional



Fonte: Retratos da leitura no Brasil (2019, p. 20)

2.1.2 Segunda Verificação

A **Tabela 2** exibe a escolha do público, estratificada de acordo com o número de leitores, como parâmetro de comparação com o segundo cenário, além do vigor das suas especificações técnicas para tal.

Tabela 2- Especificações técnicas

Especif. técnicas	Total	Leitor
Amostra	8076	4270
intervalo de confiança	95%	95%
margem de erro	1.1%	1.5%

Fonte: Retratos da leitura no Brasil (2019, p. 13)

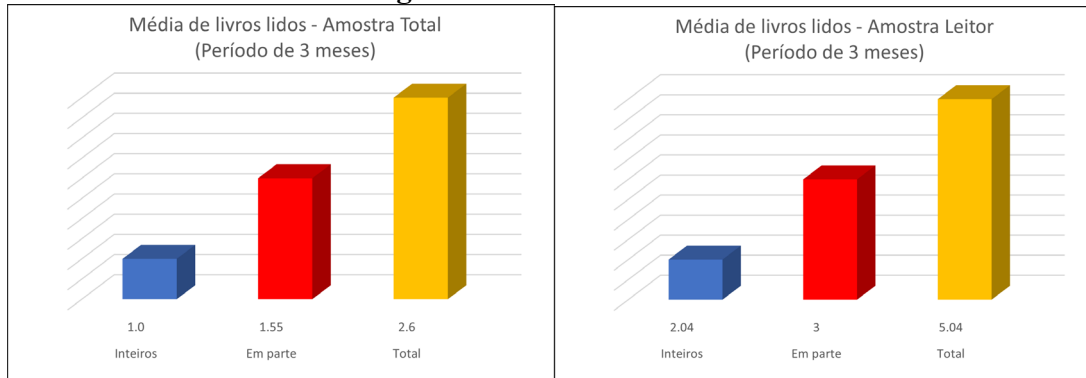
Do universo populacional, extraiu-se uma amostra, com tamanho total de 8076 indivíduos, preservando um intervalo de confiança da média populacional de 95%, além de um erro da média amostral de 1,1%.

A amostragem total contém, o conjunto de dados dos Leitores, cuja análise isolada, permite abstrai-lo à condição de amostra, composta de 4270 indivíduos, dotada de um intervalo de confiança de 95%, e um erro da média de 1,5%

A **Figura 3** ilustra as medidas de tendência central de posição. As medias estão

representadas em três caracterizações diferentes do ato de ler, de tal modo, que o gráfico da esquerda elucida as medias da amostra Total, e o da direita, a amostra leitora.

Figura 3 - Média de livros lidos



Fonte: Retratos da leitura no Brasil (2019, p. 38)

A normalização dos dados da amostra dos leitores, por meio da linguagem de programação python, pode ser apreciada na **Figura 4**.

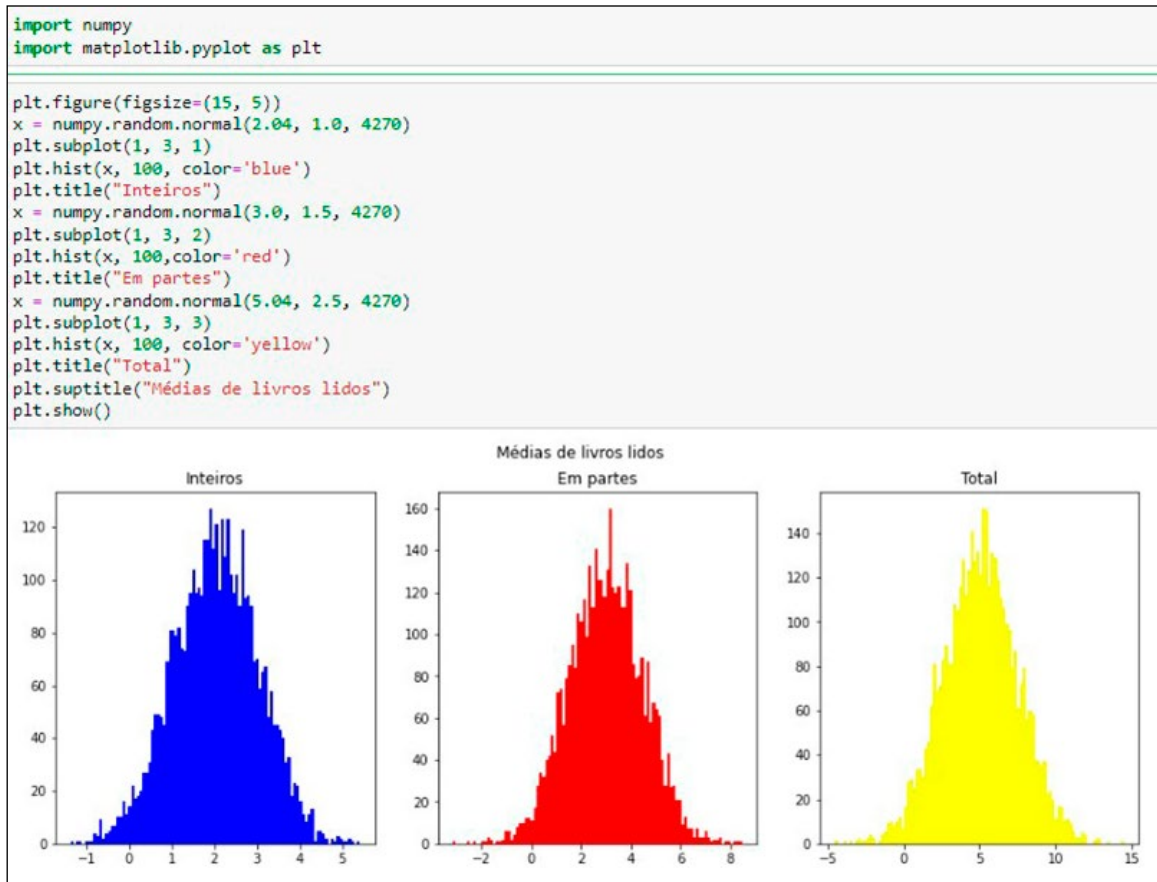
Para o desvio padrão:

$$\theta = \frac{\sqrt{n}}{Z} \cdot E$$

- ⊖ – desvio padrão da média;
- E – Erro padrão;
- Z – Escore do nível de confiança;
- N – Tamanho da amostra

Foram normalizadas, as três formas de dispersão de dados em relação às medias; sendo que a média da amostra dos livros lidos Inteiros (curva azul), foi distribuída com média = 2.04, desvio padrão=1 e tamanho da amostra = 4275. A amostra dos lidos em partes (curva amarela), foi normalizada com média=3, desvio padrão =1.5 e tamanho da amostra=4275 e finalmente a amostra total (curva vermelha), teve média =5.04, desvio padrão=2.5 e tamanho amostral = 4275.

Figura 4 - Normalização em Python

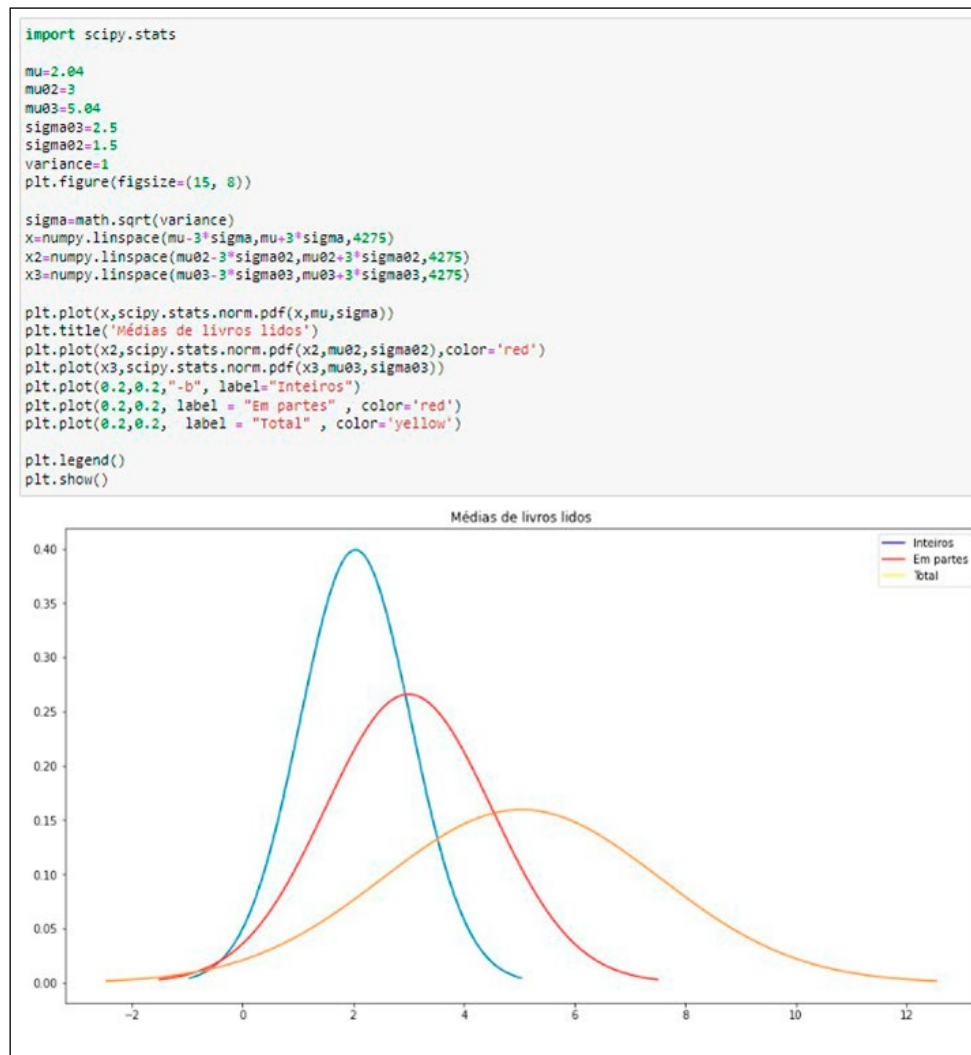


Fonte: O autor

2.1.3 Terceira Verificação

Refere-se à escolha do indicador de leitura (livros lidos inteiros, em partes, ou no total) mais adequado. A **Figura 5** demonstra, através da linguagem de programação python, as diferenças distributivas das três quantificações da amostra dos livros lidos; grosso modo, é fácil verificar visualmente que, quanto menor o desvio padrão, mais próximo da média estarão a maioria dos dados adjacentes, tanto à esquerda como à direita.

Figura 5 - As três médias dos livros lidos, em python



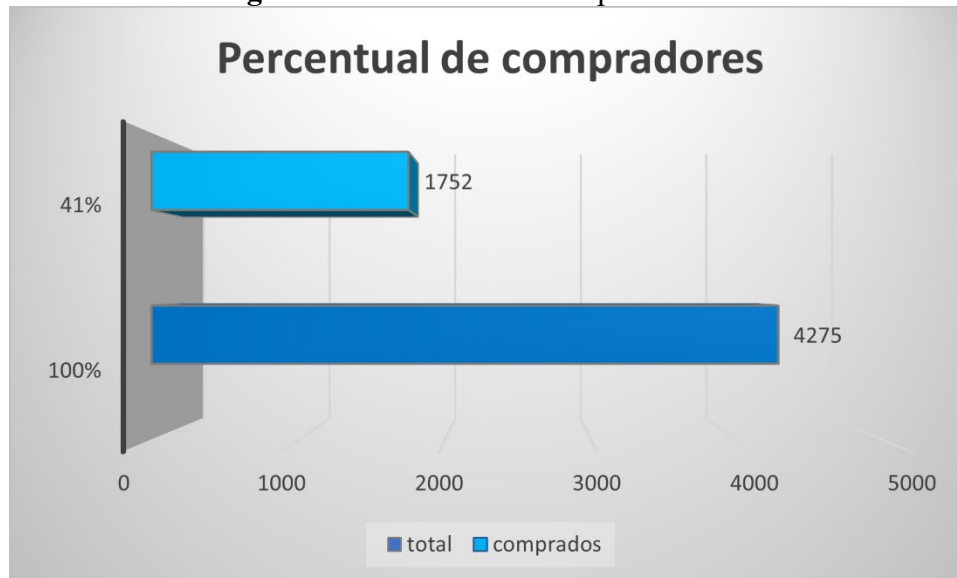
Fonte: O autor

2.1.4 Quarta Verificação

Refere-se ao consumo de livros dos leitores de livros inteiros.



Figura 6 - Percentual de compradores



Fonte: Retratos da leitura no Brasil (2019, p. 127)

Segundo a pesquisa, do total de 4275 leitores de livros, 41% compraram pelo menos um livro nos últimos três meses, resultando na distribuição normal, ilustrada na **Figura 7:**

Figura 7 - Livros lidos e comprados



Fonte: o autor

Encerram-se, aqui, os dados coletados do censo; adiante estarão expostos os dados do cenário dois (Livraria Rebelião Intelectual), em conjunto com a comparação diagnostico.



2.2 Análise e diagnóstico

Com a finalidade de encontrar a causa do problema que acomete os vários panoramas do livro e da leitura, no Brasil, tais como, os baixos índices de leitura da população em geral, ou os reduzidos percentuais de venda de livros, o presente trabalho decidiu relacionar, ambos os cenários, aqui tratados, com base em uma análise descritiva dos seus dados.

Para isso, usaremos o Teorema do Limite Central, que é assim definido por Bussab e Morettin (Bussab e Morettin, 2013) “Para amostras aleatórias simples (x_1, \dots, x_n), retiradas de uma população com média μ e variância θ^2 finita, a distribuição amostral da média \bar{x} aproxima-se, para n grande, de uma distribuição normal, com média μ e variância $\theta^2/2$.”

Já esclarecida a componente teórica da estratégia, em busca da causa do problema, são agora descritos, os dados coletados no segundo cenário.

De natureza direta, a coleta possui dados primários, extraídos de relatórios da empresa, aqui estudada, do ano de 2019, quando um dos pesquisadores deste trabalho, era colaborador e prestava contas à respectiva gerência da livraria Rebelião Intelectual.

Nas tabelas 3 e 4, estão apontados os índices de vendas da livraria Rebelião Intelectual, correspondentes aos meses de julho, agosto e setembro de 2019

Percebe-se, que houve 414 clientes, cujas compras somaram 460 pedidos, resultando num valor total de R\$ 39.844,29, uma média de 1,11 livros por comprador, além de um desvio padrão de 0,15.

Tabela 3 - Compras de julho, agosto e setembro de 2019

Compras de julho, agosto e setembro de 2019				
Mês	Compradores	Pedidos	Valor total	Ticket médio
Julho	122	149	R\$ 11.888,54	R\$ 79,79
Agosto	149	155	R\$ 12.602,22	R\$ 81,30
Setembro	143	156	R\$ 15.353,53	R\$ 107,37

Fonte: o autor

Tabela 4 - Os valores da dispersão de dados dos compradores



Os valores da dispersão de dados dos compradores				
Mês	Compradores	Pedidos	Valor total	Ticket médio
Total	414	460	R\$ 39.844,29	R\$ 268,46
Média	1	1,11	R\$ 86,62	R\$ 89,48
Desvio Padrão	0	0.15	R\$ 49,85	-
Nível de confiança		95%	95%	
Erro padrão		1.5%	1.5%	

Fonte: o autor

A partir dessas informações é possível responder:

2.2.1 Diagnóstico da Primeira Verificação

De acordo com os dados coletados, tais como os da figura 2, evidenciam um desempenho de leitura ineficiente da população brasileira, com apenas 52% de leitores.

No cenário 2, também há evidência de um baixo desempenho literal, pois nota-se baixos índices de compra de livros por leitor, com média de 1,11 livros/leitor, exposta na **Tabela 4**.

Esses números dizem muito, afinal não haverá muitas alternativas de desenvolvimento humano à uma sociedade que não cumpra com este requisito fundamental.

2.2.2 Diagnóstico da Segunda Verificação

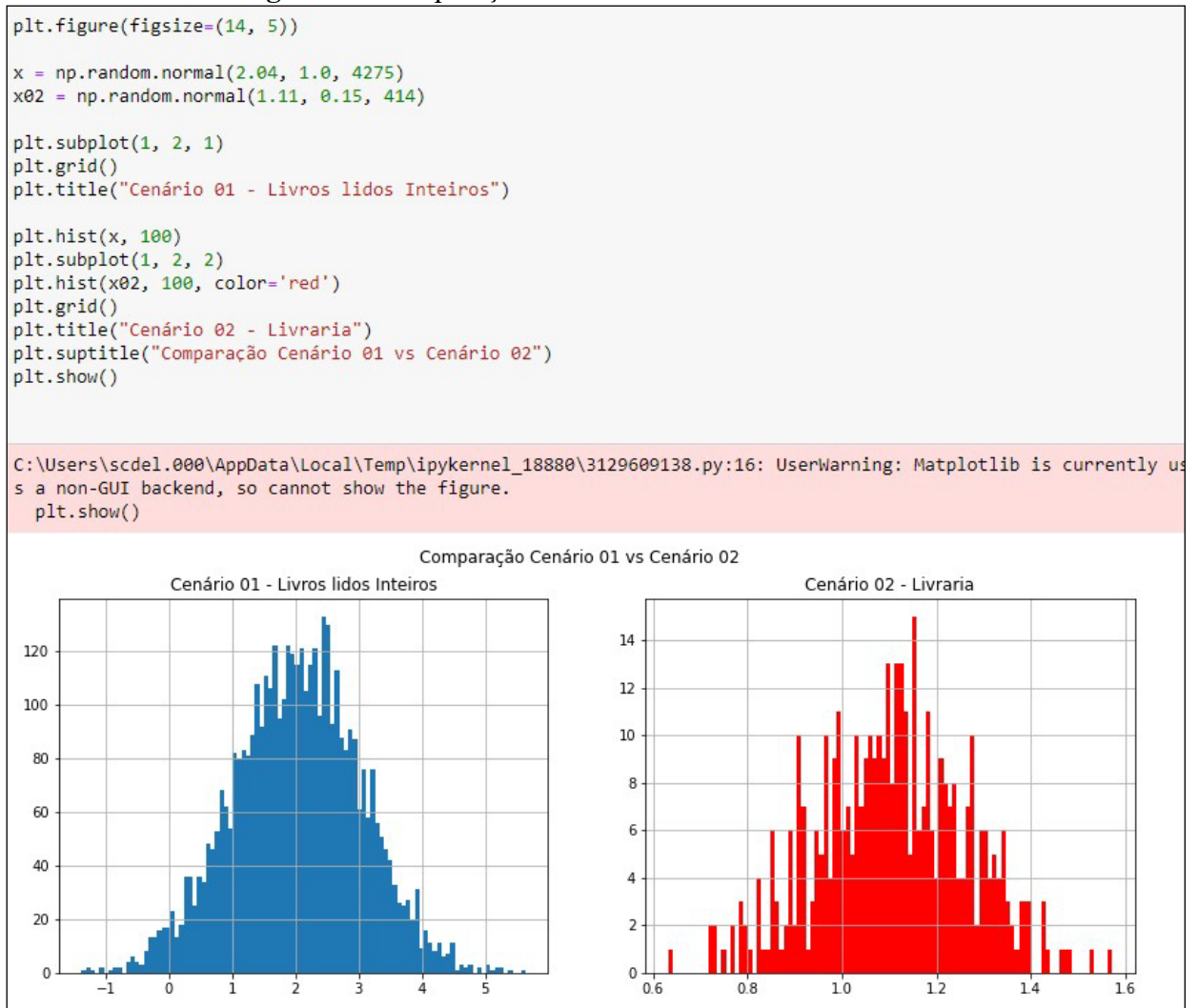
Em consequência da primeira verificação, concatenada com a incerteza de que a solução fosse angariar mais clientes para comprar livros, decidiu-se trabalhar somente com os dados de clientes leitores, pois os tornam menos obtusos, além de ponderar que é mais fácil vender quantidades maiores a clientes já existentes, do que tentar encontrar novos clientes.

Assim a amostra escolhida até aqui, representa os leitores.

2.2.3 Diagnóstico da Terceira Verificação

Dos três apontadores, pareceu demasiadamente duvidoso, o conceito de livros lidos em partes, pois é de difícil mensuração analisar uma média de livros parcialmente lidos, além de serem, na sua maioria, livros obrigatórios, tal fato criaria um viés na pesquisa que prioriza a espontaneidade como fator importante no combate à causa do problema, aqui descrito. De tal modo que suas curvas de densidade teriam essa forma de comparação (**Figura 8**):

Figura 8 - Comparação Cenário 01 vs Cenário 02



Fonte: o autor

A análise da Figura 8, transparece a sutil tendência de ambas as curvas



convergiem à uma recíproca simetria, resultando:

$$K_m - \text{Razão entre as médias} \rightarrow K_m = \text{cenário01/cenário02} \rightarrow K_m = 1.83$$

$$K_{dp} - \text{Razão entre os desvios padrão} \rightarrow K_{dp} = \text{cenário01/cenário02} \rightarrow K_{dp} = 6.66$$

$$K_n - \text{Razão entre o tamanho} \rightarrow K_n = \text{cenário01/cenário02} \rightarrow K_n = 10.32$$

Contudo, a paridade acima está longe do ideal; é preciso garimpar ainda mais os dados, para que eles tornem a investigação verossímil. Procede-se, então ao próximo passo.

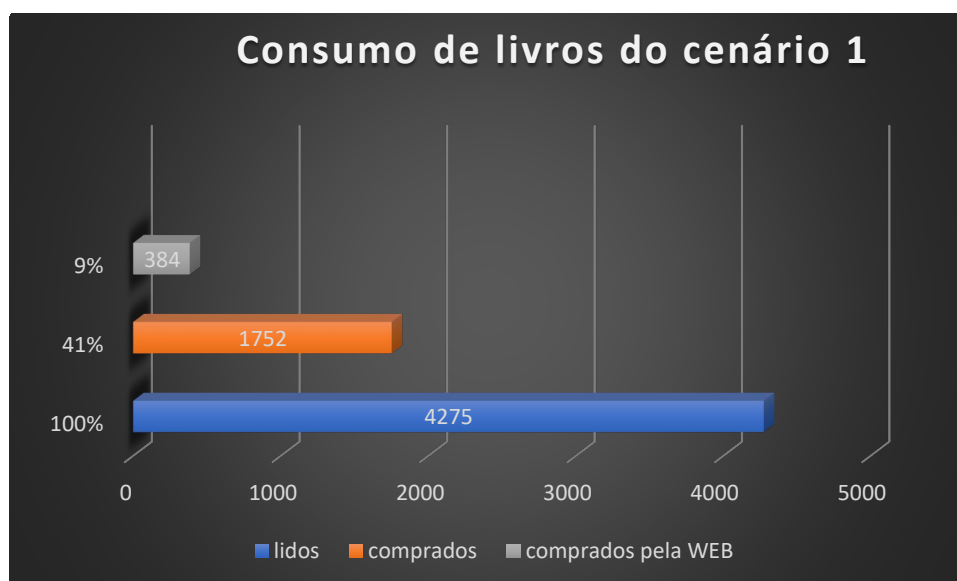
2.2.4 Diagnóstico da Quarta Verificação

No cenário dois, abstraiu-se os dados, tendo como premissa, clientes que comparam ao menos um livro, além de serem vendas exclusivas pela WEB.

De modo que, para aproximar os dois cenários, ao alcance do problema, aqui investigado, serão tomadas as mesmas medidas de abstração, aos dados do cenário 01.

Inclusive, esta informação está disponível no censo, que baseou a construção da primeira coleta de dados.

Figura 9 - Consumo de livros do cenário 1



Fonte: Retratos da leitura no Brasil (2019, p. 129)

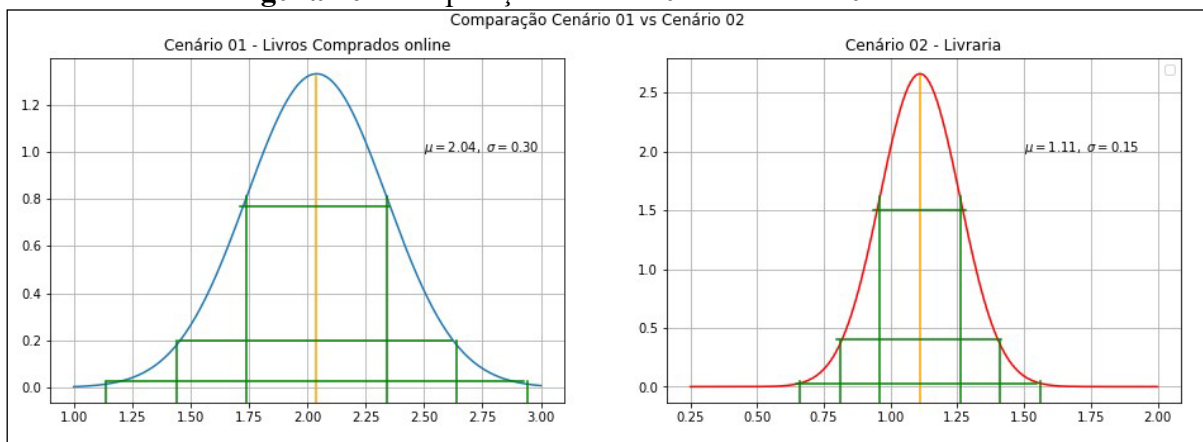
Na Figura 9, observa-se que de um total amostral de 4275 leitores, 41% (1752) compraram livros nos últimos três meses. Desses compradores, 22% deles, ou 9% (384) do total, efetuaram a compra via online.

Reformula-se, pois, a amostra do primeiro cenário, da seguinte forma:

Amostra do cenário 01: média = 2,04; desvio padrão = 0.30; amostra = 384

E, finalmente, comparam-se as amostras aptas para seguir adiante com a pesquisa:

Figura 10- Comparação Cenário 01 vs Cenário 02 II



Fonte: o autor

Aplicando novamente as razões, temos:

$$K_m = 1.83$$

$$K_{dp} = 2.0$$

$$K_n = 0.92$$

Aqui estão dados que compartilhem o problema: Poucas movimentações e aquisições de cada cliente. O foco deve ser ele, os clientes já cadastrados, e relativamente fidelizados.



3 SOLUÇÃO

O presente capítulo, apresenta, como proposta de solução ao problema aqui abordado, um modelo preditivo, com base no aprendizado de máquina, capaz de prever a próxima compra do cliente. O algoritmo preditivo, aprenderia com os dados passados de tal cliente, sugerindo a probabilidade de compra do próximo título, permitindo, deste modo, à livraria Rebelião Intelectual, antecipar a providência do volume, e ofertá-lo prontamente, resultando, por fim, no aumento de vendas per capita, dos clientes fidelizados.

Esta análise preditiva é compreensível através da linguagem de programação python.

3.1 Propostas de solução

➤ **Proposta A:** Modelo de regressão preditiva de vendas futuras

✓ **Descrição da proposta:**

Para Cezar Taurion, um modelo preditivo (TAURION, 2014) é, uma função matemática que, aplicada a uma massa de dados, consegue identificar padrões ocultos e prever o que poderá ocorrer. Tal modelo, foi concebido nos limiares do paradigma de aprendizado de máquina, cujo termo oculto, citado pelo autor, é seu desafio, e concomitante recompensa, quando o que estava oculto, emerge ao alcance da percepção sensorial do pesquisador.

Portanto, o fenômeno oculto, que impede o livro (e a leitura) de prosperar em cenários, tais como os descritos nesse trabalho, é uma possibilidade de elucidação matemática, quando submetido a um modelo de análise preditiva, sendo o modelo de regressão logística, o escolhido para tratar o problema, aqui proposto.

A análise de regressão logística, é capaz de aglutinar, classificar e agrupar características de cada cliente, com base no seu aprendizado das informações passadas de cada um, resultando na predição probabilística de específico título do livro, que todavia, não foi adquirido.



O contexto matemático é indispensável para compreender a função de um algoritmo, que foi calibrado para executar uma análise de regressão logística; assim sendo, eis seu padrão (equação):

O modelo logit (FERREIRA; ALMEIDA; SILVA,2021)

$$P_i = E\left(Y = \frac{1}{X_i}\right) = \frac{1}{1 + e^{-(B_1+B_2X_i)}}$$

Onde:

Pi = Probabilidade de aquisição do livro predito;

B1 = Constante;

B2 = Coeficiente da variável i

Xi = Variáveis explicativas (gênero, cor, escolaridade, renda, gênero textual de preferência).

✓ Possíveis impactos:

Diante da ruptura com o status quo, vigente até a virada do século, mudanças significativas na forma de compreender a realidade foram instauradas com o advento do aprendizado de máquina; isto é, os, já avançados, desenvolvimentos tecnológicos produzidos até 20 anos atrás, eram capazes de responder àquilo que a ciência sabia que não sabia, entretanto o novo paradigma, que é alicerçado no sonho dos teóricos quânticos (uma máquina com infinitas instruções, por unidade de tempo), traz consigo, a possibilidade de elucidar, aquilo que a ciência não sabia, que não sabia. É, portanto, um desafio aos colaboradores da Rebelião Intelectual, adaptarem-se à busca de resultados que já não se alinham à métrica cognitiva clássica.

✓ Prós:

A máquina, com sua atual performance de executar operações, cuja complexidade extrapola a capacidade autônoma cognitiva do ser humano, de o fazer, proporciona ao mundo, a superação e a inflação dos limites perimetrais do conhecimento, com uma velocidade exponencial.

Contudo, está entre os valores da livraria Rebelião Intelectual, perpetuar aquilo que é insubstituível, tal como o livro, tendo como realce a esse desejo a súplica de Patti



Smith (AFP, 2021), clamando a todos, que jamais abandonem o livro, por mais avançada que esteja, a ciência.

✓ **Contras:**

Como sintoma negativo à proeminência da incerteza futura, há ocorrência de certos fenômenos físicos, sobretudo os que são, dependentes de outros, que dissipam evidências, que se transformam em pistas, ora verdadeiras, ora falsas. Diante dessa contingência, são frequentes, os fracassos científicos, empresariais e orçamentários.

Não é de se estranhar que, Erwin Schroedinger (CIÊNCIA..., 2015), físico que simplesmente deduziu a equação de função de onda, além de ganhar o prêmio Nobel pelo feito, se aborreceu com as consequências de sua descoberta, negando-a, e se arrependendo profundamente, pela sua contribuição; morreu convencido de que a natureza não poderia funcionar, de acordo, com sua fórmula. Ela, a equação de onda de Schroedinger, continua sendo válida.

✓ **Recursos:**

Capital humano: dispêndio com a contratação de especialistas cientistas de dados e reformulação do marketing digital.

Financeiro: custo elevado, entretanto essencial, na contratação de serviços em nuvem, tais como: SAAS (software como serviço) para armazenamento dos dados; PAAS (plataforma como serviço) para gerenciamento dos dados; e, IAAS (infraestrutura como serviço) para escalabilidade dos dados.

✓ **Teoria de suporte e autor**

Do repertório de referências de apoio à defesa da hipótese, aqui discutida, há destaque aos conjuntos de dados, integrantes dos cenários um e dois. São, suas fontes descritas, de forma recíproca:

Primeiro cenário: diante de um panorama (livros, leitura, aquisição) pouco investigado, nos seus pormenores, e escassa disponibilidade na literatura, foi, de acordo com o autor, indispensável a utilização do censo Retratos da leitura no Brasil (2019), nobre trabalho, constituído de uma abrangente análise descritiva, dos hábitos do



brasileiro, para com o livro em geral. Não obstante, era preciso abstrair os atributos das variáveis, para que elas apontem às características chave aqui discutidas; são elas, indivíduos com hábito de leitura, frequência de livros lido, se compra livros e com que frequência. Para tal propósito, considerou-se prudente, reduzir o tamanho das amostras, à medida em que os dados fossem minerados, cuja coleta de dados, por mais reduzida que esteja, mantivesse um nível de confiança de 95%, e um erro amostral de 1,5%, índices que permitem que uma amostra, retirada de outra amostra normalizada, seja dispersa no interior de uma curva normal. Como primeira evidência preocupante, quando se discute percentual e frequência de leitura no Brasil, está a excessiva redução, de amostra em amostra, do total amostral, que cumpria as exigências da estimativa e precisão da análise estatística. Os dados de redução amostral estão expostos, na tabela abaixo. De uma população estimada em 100.000.000 de indivíduos., geraram-se, como herança, as amostras:

Tabela 5 - Redução progressiva da amostra

Total	dos Leitores	De Compradores	de Compradores (via WEB)
8076	4275	1756	384

Fonte: o autor

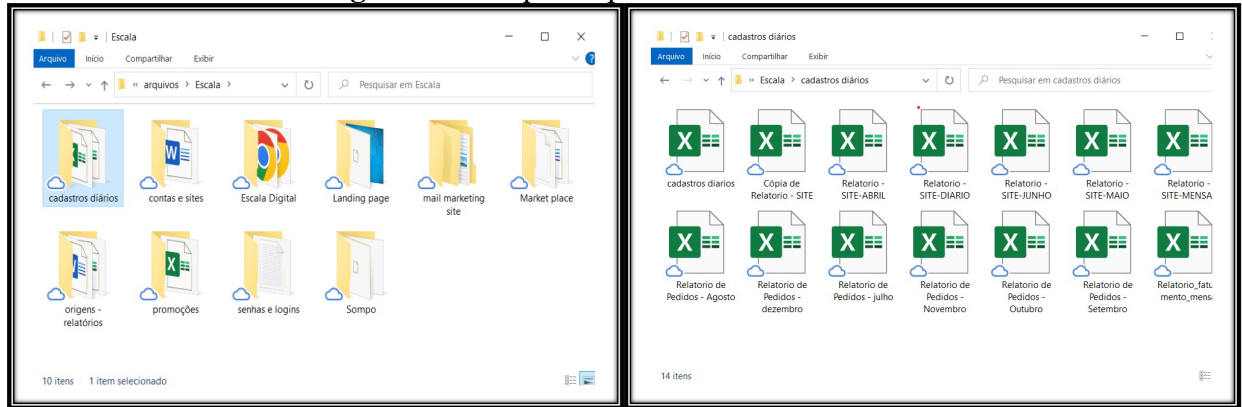
Tamanha amplitude (8076 - 384), é um claro diagnostico da fragilidade primaria dos alicerces culturais da sociedade brasileira, que a impede de aumentar seus respectivos índices de desenvolvimento humano. Enquanto for este, o cenário apresentado, nada mudará.

Segundo cenário: são os dados primários, coletados de relatórios da Rebelião Intelectual, que por bom senso, estão o mais próximo possível, de descreverem evidências do problema e como solucioná-lo.

Tais relatórios, tem origem da plataforma de e-commerce da livraria, extraídos em 2019, pelo próprio autor, que na época, era colaborador, responsável pelo gerenciamento desta plataforma, tipo SAAS (software como serviço), composta por ambas as camadas (Front end e Back end); as planilhas, tinham por finalidade, prestar contas diárias, ao gerente e diretor do departamento de sites.

Uma planilha completa, encontra-se no **Apêndice A**.

Figura 11 - Arquivos pessoais do autor



Fonte: o autor

O conjunto de dados do cenário 02, foi dispersado, no interior de uma curva normal, com as seguintes medidas de dispersão:

Tamanho da amostra: 414;

Média: 1,11;

Desvio padrão: 0,15.

A comparação, entre as coletas de dados (01 e 02), superou as expectativas, corroborando com a hipótese de que o fenômeno que restringe o aumento de aquisição de livros, se manifesta em ambos os cenários. Percebe-se, na **Tabela 6**, o nivelamento dos índices, com destaque aos tamanhos amostrais; os valores das medias são justificados pelo fato de que, o cliente que comprou na Rebelião Intelectual, pode haver comprado em outra loja online, também, no período pesquisado; e é justamente a média o foco do problema e da solução.

Tabela 6 - Comparação final de ambos os cenários

	Tamanho amostral	Média	Desvio padrão
Cenário 01	384	2,04	0,30
Cenário 02	414	1,11	0,15

Fonte: o autor



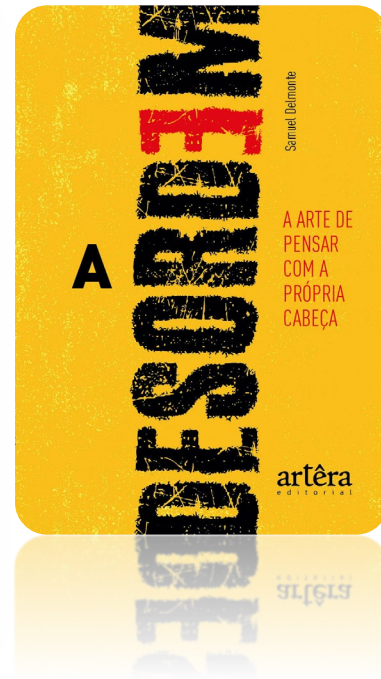
Para a solução do problema, e aumentar o número de vendas de livros, deve-se aumentar qualitativos da média que, grosso modo, presume-se, estatisticamente, que a hipótese é validada fazendo com que, prioritariamente, os clientes já existentes, comprem mais, ao invés de galgar novos clientes, em uma tentativa, de natureza quantitativa.

Atribuído, portanto, à média como, o antígeno do modelo, elucida-se, como agente de anticorpo, fidelizar o cliente, predizendo quais serão suas intenções de compra futura.

Outros autores:

- **Ferreira, Almeida e Silva (2021)** descreveram um modelo matemático, que pode mensurar os motivos que estimulam clientes optarem, pelo varejo de lojas físicas, ou virtuais; para tal, utilizaram de um modelo de regressão logística, com um algoritmo logit que aprende com a experiência (aprendizado de máquina); usaram, também, equações de séries temporais preditivas econométricas, que associam variáveis qualitativas (características humanas), às quantitativas (calibre numérico de influência dos sentimentos, hábitos, emoções etc.)
- **Rezende (2021)**, professor da disciplina de Classificação, do curso de Especialização em inteligência Artificial, da Universidade Presbiteriana Mackenzie, cursado pelo autor. São lecionados os passos necessários para modelar, um teste, em python, com base no aprendizado de máquina, e por meio de uma classificação de regressão logística
- **O Autor (2022)**, utilizou de seu repertório pessoal para atingir suas premissas iniciais, a respeito, desse estudo de caso, com o que havia incorporado na vida acadêmica. Seu portfólio online pessoal, **Delmonte (2018)** possui simuladores, programados em Javascript e python, visto na **Figura 12a**, capazes de emular, através da entrada de dados efetuada pelo usuário, uma breve apresentação das variáveis de dispersão e tendência central, por meio visual e analítico, sendo, deste modo, mais abrangente que uma calculadora. Também, foi útil, ao autor, seu livro Delmonte (2021) **A Desordem, a arte de pensar com a própria cabeça (Figura 12b)**, como referência à adoção do rigor acadêmico na formatação deste trabalho.

Figura 12 - a - Simuladores estatísticos do site; 12 -b - Livro do autor publicado



Fonte: o autor

3.2 Conexão da proposta com os resultados esperados

O teste do modelo, realizado até aqui, tem conexão com a hipótese adotada. Ou seja, optou-se pelo estudo de caso, como procedimento técnico, para investigar a ocorrência do mesmo fenômeno, cuja influência, repete-se, com simétrica significância, em diferentes cenários, com a mesma herança de classe (semelhanças) e certa polimorfia (especialização), que pode ser desprezada da quantificação.

Tal fenômeno, que, de acordo com a hipótese aqui defendida, influenciaria o índice de vendas, assim como interviria na intensidade média de leitura, de amostras inseridas neste contexto. Influência, esta, tida como déficit em ambos os cenários.

A integral da função de densidade da curva normal, dotada de um rigor matemático, que ela manifesta, com fria e austera beleza, conspirou a favor das previsões feitas, e deu indícios de que uma medida de tendencia central, aqui no caso a média, seria



o agente de influência, com significância o bastante, para contribuir à uma baixa média do índice de leitura, dos brasileiros, no cenário 01, e, concomitante baixo índice na aquisição de livros, dos clientes da livraria Rebelião Intelectual, no cenário 02.

Equação 1 - Integral da função densidade normal

$$F(x, \mu) = \frac{1}{\sqrt{2\pi\theta^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\theta^2}}$$

Estabelecida a média, como agente de mudança, faz-se necessário, aumentar sua intensidade numérica, para atingir o objetivo maior de aumentar a quantidade de vendas, e conseqüente lucro.

E, para tal, decidiu-se romper com as teorias de mercado neoliberal, vistas pelo autor, como aspirações fantasiosas de uma sociedade feliz, rica e abastada, através do livre mercado altruísta. Fato que, como já argumentado no início, simplesmente não acontece.

Por isso que, o foco não será aumentar o número de clientes, mas sim, nas vendas per capita, aumentando a média de aquisições de cada cliente. A estratégia é ousada, afinal, pretende-se, utilizar como modus operandi, a predição, do interesse ou desejo, de aquisição futura de cada cliente, podendo antecipar a edição e a própria oferta do título, ao cliente em questão.

A predição será construída por uma rede neural, cuja classificação será feita, com base no paradigma de regressão logística, um algoritmo de entrada e saída binária, clássico do aprendizado de máquina.

Grosso modo, o algoritmo logístico, aprenderá, colhendo características específicas de cada cliente, com o passar do tempo, sugerindo, deste modo, quais serão os interesses de compra futura de livros de cada cliente.

O modelo matemático se desenvolverá, através da inspiração causada pelo esboço de código, visualizado nas **Figuras 13 e 14**, que já na sua gestação, calibra promissoras predições de acurácia, com valores que superam os 70 % de precisão. O código completo do protótipo, encontra-se no **Apêndice B**.

Figura 13- Protótipo regressão logística 01

```
In [1]: import pandas as pd # dataframe manipulation
import numpy as np # linear algebra

# data visualization
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
print('Seaborn version', sns.__version__)
sns.set_style('whitegrid')

# text data
import string
import re
df = pd.read_csv('bestsellers with categories.csv')
```

Seaborn version 0.11.2

```
In [2]: df02 = pd.read_excel('clientes.xlsx')
df03=df.head(50)
df02=df02.head(50)
df03.client=df02
df03.head()
```

C:\Users\scdel.000\AppData\Local\Temp\ipykernel_6764\3501404910.py:4: UserWarning: Pandas doesn't allow columns to be created via a new attribute name - see https://pandas.pydata.org/pandas-docs/stable/indexing.html#attribute-access

```
df03.client=df02
```

```
Out[2]:
```

	Name	Author	User Rating	Reviews	Price	Year	Genre
0	10-Day Green Smoothie Cleanse	JJ Smith	4.7	17350	8	2016	Non Fiction
1	11/22/63: A Novel	Stephen King	4.6	2052	22	2011	Fiction
2	12 Rules for Life: An Antidote to Chaos	Jordan B. Peterson	4.7	18979	15	2018	Non Fiction
3	1984 (Signet Classics)	George Orwell	4.7	21424	6	2017	Fiction
4	5,000 Awesome Facts (About Everything!) (Natio...	National Geographic Kids	4.8	7665	12	2019	Non Fiction

```
In [3]: n_best = 20

top_authors = df.Author.value_counts().nlargest(n_best)
no_dup = df.drop_duplicates('Name') # removes all rows with duplicate book names

fig, ax = plt.subplots(1, 3, figsize=(11,10), sharey=True)

color = sns.color_palette("hls", n_best)

ax[0].hlines(y=top_authors.index, xmin=0, xmax=top_authors.values, color=color, linestyle)
ax[0].plot(top_authors.values, top_authors.index, 'go', markersize=9)
ax[0].set_xlabel('Number of appearances')
ax[0].set_xticks(np.arange(top_authors.values.max()+1))
ax[0].set_yticklabels(top_authors.index, fontweight='semibold')
ax[0].set_title('Appearances')

book_count = []
total_reviews = []
for name, col in zip(top_authors.index, color):
    book_count.append(len(no_dup[no_dup.Author == name]['Name']))
    total_reviews.append(no_dup[no_dup.Author == name]['Reviews'].sum()/1000)
```

Fonte: o autor



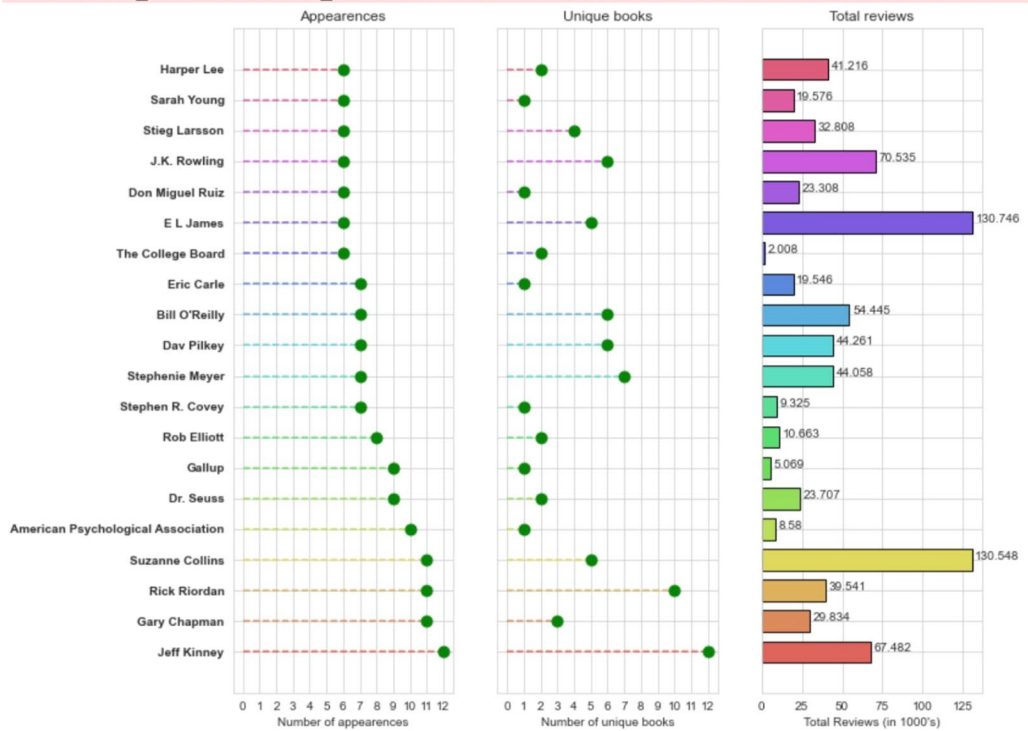
Figura 14 - Protótipo regressão logística 02

```
ax[1].plot(book_count, top_authors.index, 'go', markersize=9)
ax[1].set_xlabel('Number of unique books')
ax[1].set_xticks(np.arange(max(book_count)+1))
ax[1].set_title('Unique books')

ax[2].barh(y=top_authors.index, width=total_reviews, color=color, edgecolor='black', hei
for name, val in zip(top_authors.index, total_reviews):
    ax[2].text(val+2, name, val)
ax[2].set_xlabel("Total Reviews (in 1000's)")
ax[2].set_title('Total reviews')

plt.suptitle('Top 20 best selling Authors (from 2009 to 2019) details', fontsize=15)
plt.show()
```

C:\Users\scdel.000\AppData\Local\Temp\ipykernel_6764\1934722200.py:14: UserWarning: FixedFormatter should only be used together with FixedLocator
ax[0].set_yticklabels(top_authors.index, fontweight='semibold')



Fonte: o autor



Pode-se, portanto, verificar até aqui, concordância com as evidências e a hipótese inicial; grosso modo, há necessidade de aumentar as médias qualitativas dos clientes que são compradores de livros da editora Rebelião Intelectual, obtendo, deste modo, maior lucro; descarta-se a tentativa de aumentar o número de leitores, por se tratar de um encargo com outras variantes, de tal complexidade, que superam os horizontes e o bom senso deste trabalho; os meios que farão atingir o objetivo se referem às redes neurais que prenciarão as compras futuras de cada cliente, através de um modelo estatístico de regressão logística.

Sem mais o que dizer sobre a solução, prossegue-se com a realização do plano de ação, responsável em mensurar um cronograma coeso com as exigências, elucidadas até aqui, por este trabalho.



4 PLANEJAMENTO

A afirmação de Franz Kafka (CARONE,2009) “A partir de certo ponto não há mais retorno”, é verdadeira nesta etapa do desenvolvimento do presente trabalho, pois uma vez que, o plano de ações esteja em atividade, submeterá a Editora Rebelião Intelectual, à uma ruptura digital com precedentes permanentes e irreversíveis, tratando-se, é claro de um novo paradigma de estratégia e execução de suas atividades. E, para tanto, é primordial transformar suas informações, sobre os clientes, tanto já armazenadas, como as futuras em um *big data*, termo que se refere, grosso modo, a um repositório de dados, previamente estruturado, capaz de operar, em tempo real, a síntese dos 5 *V's* (Velocidade, Veracidade, Volume, Valor e Variedade). Daqui a diante, no decorrer das atividades expostas no **Quadro 1**, emerge a máquina programada para aprender de acordo com sua própria experiência.

4.1 Planos de ações

Quadro 1 – Proposta de solução A: Modelo de regressão preditiva de vendas futuras.

Objetivo: Estruturar as informações, a respeito dos clientes, armazenadas nos bancos de dados da editora, em um <i>big data</i> , cuja complexa rede de relacionamento entre os atributos das suas estruturas de dados, desencadeia a rede neural programável.		
Ação Detalhada	Prazo para Finalização/Implantação	Responsável (área/função)
Primeira Etapa: Coleta de Dados		
Definição dos atributos e comportamentos das entidades a ser estruturadas.	1 semana	Área: Tecnologia Função: Gerente de Projetos
Criação dos diagramas de Entidade-Relacionamento.	1 semana	Área: Tecnologia Função: Analistas de Sistemas



Formatação de dados Incompletos	2 semanas	Área: Tecnologia Função: Desenvolvedores Front End e Cientista de Dados
Filtragem e eliminação de erros dos dados coletados	1 semana	Área: Tecnologia Função: Analistas de Sistemas
Segunda Etapa: Integração dos Dados		
Catálogo dos livros e exemplares disponíveis	1 semana	Área: Tecnologia Função: Analistas de Sistemas
Especificação das classes de gêneros literários	1 semana	Área: Tecnologia Função: Analistas de Sistemas
Associação específica das referências e citações de cada obra	2 semanas	Área: Tecnologia Função: Desenvolvedores Front End e Cientista de Dados
Especificar condições de relacionamento entre entidades clientes e gêneros de leitura	2 semanas	Área: Tecnologia Função: Desenvolvedores Front End e Cientista de Dados
Terceira Etapa: Análise e Modelagem de Dados		
Preparação e seleção de plataformas e compiladores ideais para classificação por regressão	1 semana	Área: Tecnologia Função: Gerente de Projetos
Desenvolver módulos de códigos de classificação regressiva e logística em Python	3 semanas	Área: Tecnologia Função: Cientista de Dados
Calibrar os dados de entrada e saída das recém-criadas redes neurais	2 semanas	Área: Tecnologia Função: Cientista de Dados



Quarta Etapa: Análise e Interpretação de Dados		
Fase de testes da rede neural programada	2 semanas	Área: Tecnologia Função: Cientista de Dados
Fase de treinamento e ajustes do calibre dos neurônios com base no seu poder de mensurar acurácia preditiva	2 semanas	Área: Tecnologia Função: Cientista de Dados
Análise descritiva dos dados testados e treinados nas últimas 4 semanas	1 semana	Área: Tecnologia Função: Analistas de Sistemas e Cientista de Dados
Última Etapa: Testes finais e implantação do sistema		
Realizar testes com <i>stakeholders</i>	1 semana	Área: Tecnologia Função: Analistas de Sistemas e Cientista de Dados
Implantação do sistema	2 semanas	Área: Tecnologia Função: Analistas de Sistemas e Cientista de Dados
Iniciar auditorias de desempenho do aprendizado de máquina instaurado	Permanente	Área: Tecnologia Função: Analistas de Sistemas e Cientista de Dados

Todas as cinco etapas descritas no plano de ação, consideram que a editora Rebelião Intelectual, já havia, previamente, equiparado o setor de tecnologia com a contratação dos profissionais adequados à proposta, definido um gerente de projetos e providenciado a infraestrutura necessária para tal.



5 CONSIDERAÇÕES FINAIS

A proposta estabelecida pelo presente trabalho, objetivou investigar, diagnosticar e analisar o mercado editorial de livros brasileiro, no que diz respeito aos decrescentes índices de vendas do volume de exemplares e fracasso na consolidação de um perfil exclusivo de mercado, e se há relação direta com os baixíssimos índices médios de aquisição e leitura dos indivíduos da população; ademais, e como componente principal do propósito, aqui defendido, é a hipótese segundo a qual um modelo estatístico de classificação retroalimentável (neurônios binários), fosse capaz de se opor ao baixo rendimento dos já referidos indicadores, em uma relação de causa e efeito, análogo ao anticorpo que captura o antígeno. Os meios utilizados para tais fins, foram alicerçados em uma metodologia descritiva, com base no estudo de casos.

Na etapa subsequente, foram coletados dados referentes a dois modelos abstraídos de acordo com o Teorema do Limite Central, cuja premissa permitiu comparar a ambos, a um mesmo fenômeno, gerador dos resultados insatisfatórios. De tal modo que os dados do primeiro modelo, de natureza secundária e de indicadores dos índices médios de leitura do brasileiro em geral, quando pareados com o modelo dois, de dados primários, e de medidas referentes ao controle de pesquisa e vendas dos clientes da editora Rebelião Intelectual, apresentaram indícios promissores na simetria das proporções, de ordem direta, das medidas de dispersão dos seus dados, normalizados sob uma curva normal, tendo a média da amostra como destaque de evidencia; não obstante, inexistiram evidências que corroborassem com qualquer estratégia modificadora dos índices de leitura e, logicamente, níveis educacionais tupiniquins.

Modelou-se uma rede neural, em python, com paradigma preditivo, concebido com base no modelo estatístico de classificação em regressão logística, treinado e testado com calibre ajustado para auferir a acurácia de previsão do próximo livro de dado cliente.

Os resultados foram condizentes com a hipótese, com acurácia mensurada acima dos 70%, valores que tornam a possibilidade preditiva.

É, portanto, sugerido, a continuidade das pesquisas e testes da modelagem preditiva, com base no aprendizado de máquina. Não poderia ser encerado este trabalho, sem antes reverenciar a sublime e indiferente manifestação da natureza, através da função integral de densidade da curva normal.



REFERÊNCIAS

- AFP AGÊNCIA. **O povo**, 2021. Disponível em: <<https://www.opovo.com.br/vidaarte/2022/05/23/uma-grande-alegria-icone-punk-patti-smith-recebe-honraria-da-franca.html>>. Acesso em: 26 de agosto de 2022.
- ALMEIDA, D. F.; FERREIRA, M. M.; SILVA, J. G. O perfil dos consumidores de *e-commerce*: um estudo de caso para o município de Dourados (MS). **Research, Society and Development**, 2022. Disponível em: <[The profile of e-commerce consumers: a case of study for the city of Dourados \(MS\) | Research, Society and Development \(rsdjournal.org\)](https://www.rsdjournal.org/)>. Acesso em: 19 de agosto de 2022.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. São Paulo: Saraiva, 2013
- CARONE, M. **Os Aforismos Reunidos de Franz Kafka**; Instituto Moreira Salles, 2009. Disponível em <<https://www.revistaserrote.com.br/2011/06/os-aforismos-reunidos-de-franz-kafka>>. Acesso em 05 de setembro de 2022.
- CIÊNCIA E HUMANISMO: a visão da ciência de Erwin Schroedinger. **A Natureza das Coisas**, 2015. Disponível em <<https://dererummundi.blogspot.com/2015/10/ciencia-e-humanismo-visao-da-ciencia-de.html>>. Acesso em: 29 de agosto de 2022.
- DELMONTE, S. **A Desordem – a arte de pensar com a própria cabeça**. Curitiba, Appris, 2021.
- DELMONTE, S. **Natureza Elegante – simulador de física**. São Paulo, 2015. Disponível em: <<https://naturezaelegante.com/>>. Acesso em: 2 de setembro de 2022.
- FIPE – Fundação Instituto de Pesquisas Econômicas. **Produção e vendas do setor editorial brasileiro**. 2018. Disponível em: <https://cerlalc.org/pt-br/publicaciones/producao-e-vendas-do-setor-editorial-brasileiro-ano-base-2018/>.. Acesso em: 25 agosto de 2022.
- FRASES ESTATÍSTICAS, **Aprender Estatística Fácil**, 2021. Disponível em: <<https://estatisticafacil.org/2021/09/30/frases-estatistica>>. Acesso em 05 de setembro de 2022.
- IBOPE INTELIGÊNCIA. **Retrato da Leitura no Brasil, Edição 5**. 2017



OLIVEIRA, Sílvio L. **Metodologia científica aplicada ao direito**. São Paulo: Thomson Learning, 2002.

REZENDE, M, N. **Classificação**. Curso de Extensão em Inteligência Artificial, 2021. Universidade Presbiteriana Mackenzie.

SILVA, E. L.; MENEZES, E. M. (2000). **Metodologia da pesquisa e elaboração de dissertação**. Revisada, UFSC, Florianópolis, SC.

SOUSA, Rainer Gonçalves. **Origem do Capitalismo**; Brasil Escola. Disponível em: <https://brasilecola.uol.com.br/historiag/origem-capitalismo.htm>, Acesso em 20 de agosto de 2022.

TAURION, C. **Big Data**. São Paulo: Brasport, 2013.



APÊNDICES

APÊNDICE A

Tabela A – Planilha completa do Cenário 02.

Dias	Percentual	Vis. Únicos	Nº de Visitas	Page Views
01/07/2019	1,50%	381	1746	2111
02/07/2019	1,40%	398	1666	2424
03/07/2019	1,40%	401	1664	2294
04/07/2019	1,30%	377	1564	1903
05/07/2019	1,00%	303	1169	1616
06/07/2019	0,60%	179	685	984
07/07/2019	0,70%	214	871	986
08/07/2019	1,20%	306	1378	1750
09/07/2019	0,90%	314	1094	1401
10/07/2019	1,30%	370	1504	1920
11/07/2019	1,20%	348	1424	1717
12/07/2019	0,90%	271	1062	1347
13/07/2019	0,70%	209	840	1018
14/07/2019	0,80%	217	914	1068
15/07/2019	1,40%	370	1609	1897
16/07/2019	1,40%	417	1684	2262
17/07/2019	1,30%	362	1543	2004
18/07/2019	1,20%	349	1421	1954
19/07/2019	1,00%	307	1187	1597
20/07/2019	0,70%	216	859	1044
21/07/2019	0,80%	216	878	1076
22/07/2019	1,30%	353	1494	2083
23/07/2019	1,20%	383	1432	2513
24/07/2019	1,40%	378	1684	2792
25/07/2019	1,40%	392	1574	2197
26/07/2019	1,20%	315	1438	2032
27/07/2019	0,80%	226	936	1081
28/07/2019	0,80%	243	906	1114
29/07/2019	1,30%	389	1468	1882
30/07/2019	1,30%	355	1543	2086
31/07/2019	1,10%	360	1280	1708
01/08/2019	1,30%	377	1542	2057
02/08/2019	1,00%	288	1111	1590
03/08/2019	0,90%	233	1002	1185
04/08/2019	0,80%	227	877	1161
05/08/2019	1,30%	363	1498	1916
06/08/2019	1,30%	346	1458	2190
07/08/2019	1,20%	368	1383	1912



Universidade Presbiteriana Mackenzie



08/08/2019	1,10%	351	1335	1708
09/08/2019	1,00%	315	1176	1499
10/08/2019	0,60%	191	746	872
11/08/2019	0,70%	210	819	924
12/08/2019	1,30%	375	1485	2044
13/08/2019	1,40%	382	1624	2314
14/08/2019	1,30%	367	1536	1913
15/08/2019	1,10%	330	1264	2279
16/08/2019	1,00%	305	1162	1960
17/08/2019	0,90%	240	1059	1247
18/08/2019	1,00%	262	1116	1394
19/08/2019	1,90%	469	2250	3262
20/08/2019	1,90%	526	2183	3090
21/08/2019	1,40%	441	1621	2345
22/08/2019	0,90%	264	1016	1590
23/08/2019	0,70%	225	853	1477
24/08/2019	0,60%	176	683	1016
25/08/2019	0,80%	181	931	1328
26/08/2019	1,10%	298	1236	1643
27/08/2019	0,80%	238	949	1516
28/08/2019	0,90%	281	1004	1570
29/08/2019	0,80%	241	980	1466
30/08/2019	0,70%	193	761	1276
31/08/2019	0,50%	159	639	918
01/09/2019	0,70%	194	851	1037
02/09/2019	1,00%	275	1121	1965
03/09/2019	1,10%	264	1286	2038
04/09/2019	0,80%	244	924	1401
05/09/2019	0,70%	224	820	1067
06/09/2019	1,10%	301	1227	1855
07/09/2019	1,00%	269	1177	1447
08/09/2019	1,10%	293	1240	1528
09/09/2019	1,60%	419	1804	2464
10/09/2019	1,40%	401	1591	2701
11/09/2019	1,30%	384	1501	1970
12/09/2019	1,30%	393	1482	1849
13/09/2019	1,10%	353	1313	1886
14/09/2019	0,80%	217	964	1279
15/09/2019	0,80%	218	913	1071
16/09/2019	1,40%	373	1626	2431
17/09/2019	1,40%	395	1594	2120
18/09/2019	1,30%	387	1567	2240
19/09/2019	1,20%	332	1390	1828
20/09/2019	1,40%	383	1664	2297



Universidade Presbiteriana Mackenzie



21/09/2019	0,70%	231	841	1014
22/09/2019	0,90%	260	1049	1291
23/09/2019	1,20%	387	1411	1789
24/09/2019	1,30%	391	1514	2263
25/09/2019	1,40%	405	1616	2202
26/09/2019	1,30%	386	1533	2008
27/09/2019	1,00%	308	1164	1516
28/09/2019	0,80%	220	882	1081
29/09/2019	0,80%	229	963	1158
30/09/2019	1,20%	361	1400	1742
TOTAL	100,00%	28638	116244	159061

Fonte: o autor



APÊNDICE B

Figura B1 – Protótipo completo regressão logística 01

```
[1]: import pandas as pd # dataframe manipulation
import numpy as np # linear algebra
import plotly.express as px
import sklearn
from scipy import stats
# data visualization
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
print('Seaborn verion', sns.__version__)
sns.set_style('whitegrid')
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import classification_report
# test data
import string
import re
df = pd.read_csv('bestsellers with categories.csv')
```

Seaborn verion 0.11.2

```
[2]: df02 = pd.read_excel('clientes.xlsx')
df03=df.head(50)

df02=df02.head(50)
df03.Name=df02
df04=[]
df04=df02
df04=pd.DataFrame(df04)
df03=df03.drop("Year", axis='columns')
for i in range(50):
    if(df03.Genre[i]=='Fiction'):
        df03.Genre[i]=1
    else:
        df03.Genre[i]=0
```

Fonte: o autor



Figura B2 – Protótipo completo regressão logística 02

```
print(df03.to_string())
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_21552\2106220587.py:5:  
SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
df03.Name=df02
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_21552\2106220587.py:14:  
SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
df03.Genre[i]=0
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_21552\2106220587.py:12:  
SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
df03.Genre[i]=1
```

	User_Rating	Reviews	Price	Genre	Name	Author
0	4.7	17350	8	0	FRANCISCO ELSON CANPELO DE CASTRO	JJ Smith
1	4.6	2052	22	1	Gerson Delgado Barros	Stephen King
2	4.7	18979	15	0	kelly macedo do vale	Jordan B. Peterson
3	4.7	21424	6	1	regina canola sales	George Orwell
4	4.8	7665	12	0	FELIX MIGUEL DE OLIVEIRA JUNIOR	National Geographic Kids
5	4.4	12643	11	1	carolina santana pacheco	George R. R. Martin
6	4.7	19735	30	1	Mônica Cordeiro Maciel de Mello Braga	George R. R. Martin
7	4.7	19699	15	1	Mona alcantara cici	Amor Towles
8	4.7	5983	3	0	Lourival da Rocha	James Comey
9	4.6	23848	8	1	Simony Alperstedt Lofiego	Fredrik Backman

2

Fonte: o autor



Figura B3 – Protótipo completo regressão logística 03

10		Valquiria Rita Geraldini		Fredrik Backman
4.6	23848	8	1	
11		LUCIANA DE MARQUES		Larry Schweikart
4.6	460	2	0	
12		Sandra Magalhães Uchoa Pinheiro		Jaycee Dugard
4.6	4149	32	0	
13		Renato de Freitas Soares Loureiro		Madeleine L'Engle
4.5	5153	5	1	
14		Caio Henrique Rodrigues		Steve Harvey
4.6	5013	17	0	
15		Juliana Ramos Dos Santos Oliveira		Adult Coloring Book Designs
4.5	2313	4	0	
16		Kelly Bregondi Rostirola		Blue Star Coloring
4.6	2925	6	0	
17		ROSA MARIA DOS SANTOS NACARINI		Blue Star Coloring
4.4	2951	6	0	
18		THIAGO DIAS DE MELO		Coloring Books for Adults
4.5	2426	8	0	
19		vanessa barreto		Ron Chernow
4.8	9198	13	0	
20		Clodoaldo da Luz		Anthony Doerr
4.6	36348	14	1	
21		ADILENE NUNES DA SILVA		Anthony Doerr
4.6	36348	14	1	
22		Maria Emilia Leitão da Costa		Veronica Roth
3.9	6310	13	1	
23		ALESSANDRA VALENTE		Chris Kyle
4.6	15921	9	0	
24		GISELE MENDES		Khaled Hosseini
4.3	12159	13	1	
25		Antonio Scotta		Glenn Beck
4.6	798	5	0	
26		Fernanda Luiza Bechi		Neil deGrasse Tyson
4.7	9374	9	0	
27		ELZY DA SILVA CAETANO		Mark Twain
4.2	491	14	0	
28		Juliane Passos		DK
4.6	5360	5	0	
29		Júlia Silveira		Angie Grace
4.6	1909	11	0	
30		Priscila de Souza Vargas		Ina Garten
4.8	1296	24	0	
31		Ári Raynsford		Ina Garten
4.7	615	21	0	
32		Ana Lucia Ferreira da Silva		Michelle Obama
4.8	61133	11	0	
33		Maria Aparecida Barbosa		Michelle Obama
4.8	61133	11	0	

Fonte: o autor



Figura B4 – Protótipo completo regressão logística 04

```
34          JHEINE ALVES DE SOUZA          Atul Gawande
4.8  11113  15  0
35          Wagner Oliveira Pacheco        Ta-Nehisi Coates
4.7  10070  13  0
36          Regiani Janaina da costa      Ta-Nehisi Coates
4.7  10070  13  0
37          Luiz Sebastiao Pereira Teixeira Bruce Springsteen
4.7  3729   18  0
38          Rodrigo de Assis Moraes       Stephenie Meyer
4.6  9769   13  1
39          Maria Cecília Poncio Gasparoto Glenn Beck
4.5   471   8  0
40          JULIANE Frerix Barja          Bill Martin Jr.
4.9  14344   5  1
41          Lúcia Maria Diniz Siciliano Campos Bill Martin Jr.
4.9  14344   5  1
42          CRISTINA RODRIGUES DE OLIVEIRA Jeff Kinney
4.8  4505   0  1
43          wANCLEICY CORREIA            Sasha O'Hara
4.6  10369   4  0
44          GIANNE SOMENSI                David Goggins
4.8  16244  18  0
45          Cristina Massoco              Thomas Piketty
4.5  2884   28  0
46          Alice Costa de Souza          Suzanne Collins
4.7  22614  11  1
47          CELIA CHAYB AMERICANO DO BRASIL Suzanne Collins
4.7  22614  11  1
48          Maria juliana de souza       Suzanne Collins
4.7  22614  11  1
49          Angélica Rosana Bier         Chrissy Teigen
4.7  4761   16  0
```

```
[3]: tips=df03['Genre']>0
tips02=df03[tips]
tips02
```

```
[3]:
```

	Name	Author	User_Rating	\
1	Gerson Delgado Barros	Stephen King	4.6	
3	regina canola sales	George Orwell	4.7	
5	carolina santana pacheco	George R. R. Martin	4.4	
6	Mônica Cordeiro Maciel de Mello Braga	George R. R. Martin	4.7	
7	Mona alcantara cici	Amor Towles	4.7	
9	Simony Alperstedt Lofiego	Fredrik Backman	4.6	
10	Valquiria Rita Geraldini	Fredrik Backman	4.6	
13	Renato de Freitas Soares Loureiro	Madeleine L'Engle	4.5	
20	Clodoaldo da Luz	Anthony Doerr	4.6	



Figura B5 – Protótipo completo regressão logística 05

21	ADILENE NUNES DA SILVA	Anthony Doerr	4.6
22	Maria Emilia Leitão da Costa	Veronica Roth	3.9
24	GISELE MENDES	Khaled Hosseini	4.3
38	Rodrigo de Assis Moraes	Stephenie Meyer	4.6
40	JULIANE Frerix Barja	Bill Martin Jr.	4.9
41	Lúcia Maria Diniz Siciliano Campos	Bill Martin Jr.	4.9
42	CRISTINA RODRIGUES DE OLIVEIRA	Jeff Kinney	4.8
46	Alice Costa de Souza	Suzanne Collins	4.7
47	CELIA CHAYB AMERICANO DO BRASIL	Suzanne Collins	4.7
48	Maria juliana de souza	Suzanne Collins	4.7

	Reviews	Price	Genre
1	2052	22	1
3	21424	6	1
5	12643	11	1
6	19735	30	1
7	19699	15	1
9	23848	8	1
10	23848	8	1
13	5153	5	1
20	36348	14	1
21	36348	14	1
22	6310	13	1
24	12159	13	1
38	9769	13	1
40	14344	5	1
41	14344	5	1
42	4505	0	1
46	22614	11	1
47	22614	11	1
48	22614	11	1

```
[4]: tips02.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 19 entries, 1 to 48
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Name         19 non-null    object
1   Author       19 non-null    object
2   User_Rating  19 non-null    float64
3   Reviews      19 non-null    int64
4   Price        19 non-null    int64
5   Genre        19 non-null    object
dtypes: float64(1), int64(2), object(3)
memory usage: 1.0+ KB
```

Fonte: o autor



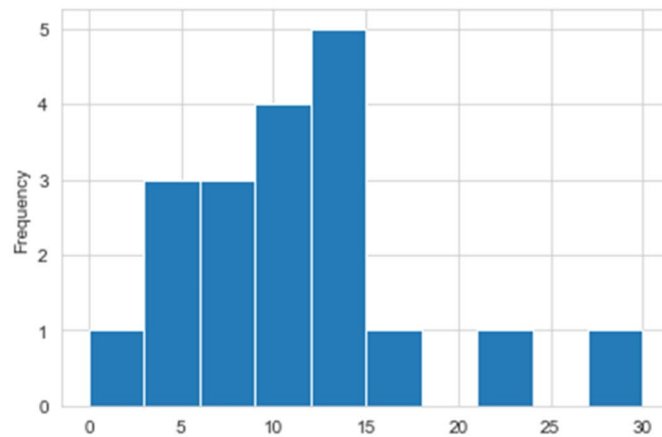
Figura B6 – Protótipo completo regressão logística 06

```
[5]: tips02.describe()
```

```
[5]:
```

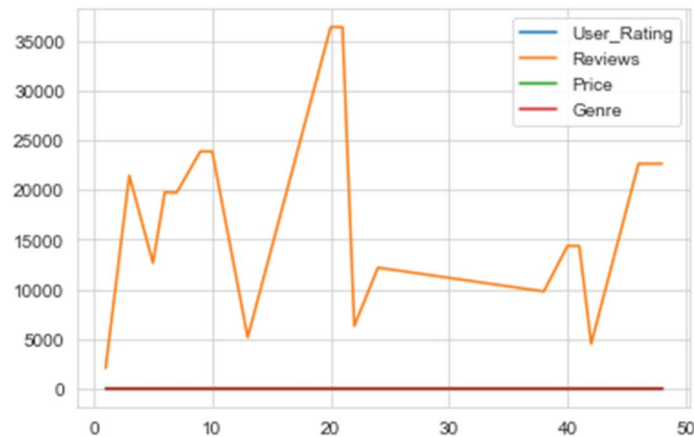
	User_Rating	Reviews	Price
count	19.000000	19.000000	19.000000
mean	4.605263	17387.947368	11.315789
std	0.224781	9751.120867	6.658767
min	3.900000	2052.000000	0.000000
25%	4.600000	10964.000000	7.000000
50%	4.600000	19699.000000	11.000000
75%	4.700000	22614.000000	13.500000
max	4.900000	36348.000000	30.000000

```
[6]: tips02["Price"].plot(kind = 'hist')  
plt.show()
```



```
[7]: tips02.plot()  
plt.show()
```


Figura B7 – Protótipo completo regressão logística 07



```
[8]: n_best = 20

top_authors = df.Author.value_counts().nlargest(n_best)
no_dup = df.drop_duplicates('Name') # removes all rows with duplicate book names
fig, ax = plt.subplots(1, 3, figsize=(11,10), sharey=True)

color = sns.color_palette("hls", n_best)

ax[0].hlines(y=top_authors.index, xmin=0, xmax=top_authors.values,
            color=color, linestyle='dashed')
ax[0].plot(top_authors.values, top_authors.index, 'go', markersize=9)
ax[0].set_xlabel('Number of appearances')
ax[0].set_xticks(np.arange(top_authors.values.max()+1))
ax[0].set_yticklabels(top_authors.index, fontweight='semibold')
ax[0].set_title('Appearances')

book_count = []
total_reviews = []
for name, col in zip(top_authors.index, color):
    book_count.append(len(no_dup[no_dup.Author == name]['Name']))
    total_reviews.append(no_dup[no_dup.Author == name]['Reviews'].sum()/1000)
ax[1].hlines(y=top_authors.index, xmin=0, xmax=book_count, color=color,
            linestyle='dashed')
ax[1].plot(book_count, top_authors.index, 'go', markersize=9)
ax[1].set_xlabel('Number of unique books')
```

7

Fonte: o autor

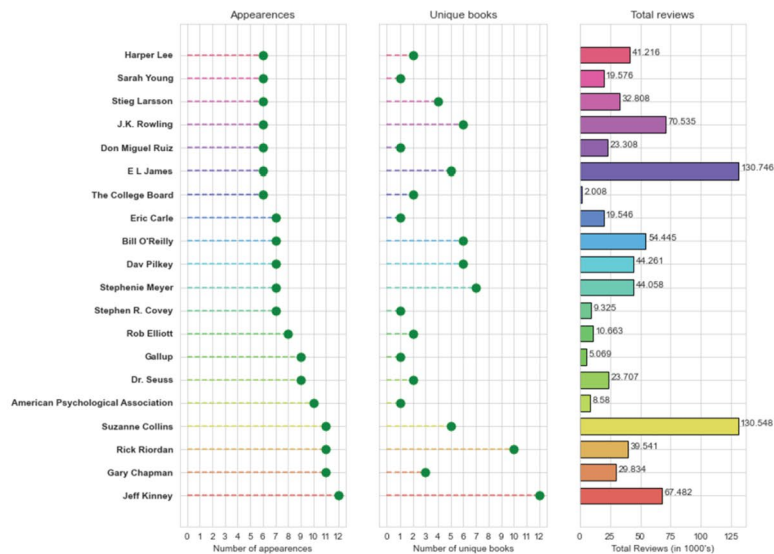
Figura B8 – Protótipo completo regressão logística 08

```
ax[1].set_xticks(np.arange(max(book_count)+1))
ax[1].set_title('Unique books')

ax[2].barh(y=top_authors.index, width=total_reviews, color=color,
           edgecolor='black', height=0.7)
for name, val in zip(top_authors.index, total_reviews):
    ax[2].text(val+2, name, val)
ax[2].set_xlabel("Total Reviews (in 1000's)")
ax[2].set_title('Total reviews')
plt.suptitle('Top 20 best selling Authors (from 2009 to 2019) details',
             fontsize=15)
plt.show()
```

C:\Users\Dell\AppData\Local\Temp\ipykernel_21552\3380732487.py:13: UserWarning: FixedFormatter should only be used together with FixedLocator

```
ax[0].set_yticklabels(top_authors.index, fontweight='semibold')
```



```
[9]: best_nf_authors = df.groupby(['Author', 'Genre']).agg({'Name': 'count'}).
      unstack()['Name', 'Non Fiction'].sort_values(ascending=False)[:11]
      best_f_authors = df.groupby(['Author', 'Genre']).agg({'Name': 'count'}).
      unstack()['Name', 'Fiction'].sort_values(ascending=False)[:11]
```



Figura B9 – Protótipo completo regressão logística 09

```
with plt.style.context('Solarize_Light2'):
    fig, ax = plt.subplots(1, 2, figsize=(8,8))

    ax[0].barh(y=best_nf_authors.index, width=best_nf_authors.values,
              color='blue')
    ax[0].invert_xaxis()
    ax[0].yaxis.tick_left()
    ax[0].set_xticks(np.arange(max(best_f_authors.values)+1))
    ax[0].set_yticklabels(best_nf_authors.index, fontsize=12,
->fontweight='semibold')
    ax[0].set_xlabel('Number of appearances')
    ax[0].set_title('Non Fiction Authors')

    ax[1].barh(y=best_f_authors.index, width=best_f_authors.values,
              color='red')
    ax[1].yaxis.tick_right()
    ax[1].set_xticks(np.arange(max(best_f_authors.values)+1))
    ax[1].set_yticklabels(best_f_authors.index, fontsize=12,
->fontweight='semibold')
    ax[1].set_title('Fiction Authors')
    ax[1].set_xlabel('Number of appearances')

    fig.legend(['Non Fiction', 'Fiction'], fontsize=12)

plt.show()
```

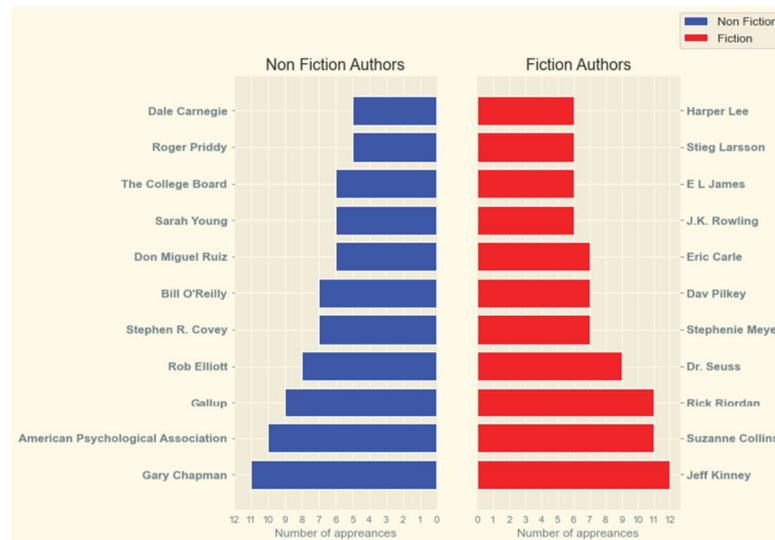
```
C:\Users\Dell\AppData\Local\Temp\ipykernel_21552\1934067301.py:12: UserWarning:
FixedFormatter should only be used together with FixedLocator
```

```
    ax[0].set_yticklabels(best_nf_authors.index, fontsize=12,
fontWeight='semibold')
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_21552\1934067301.py:20: UserWarning:
FixedFormatter should only be used together with FixedLocator
```

```
    ax[1].set_yticklabels(best_f_authors.index, fontsize=12,
fontWeight='semibold')
```

Figura B10 – Protótipo completo regressão logística 10



```
[10]: cleanDF = df.drop('User_Rating', axis=1)
      #Convert all the non-numeric columns to numeric
      for column in cleanDF.columns:
          if cleanDF[column].dtype == np.number:
              continue
          cleanDF[column] = LabelEncoder().fit_transform(cleanDF[column])
      x = cleanDF.drop('Genre', axis=1)
      y = cleanDF['Genre']
      x = StandardScaler().fit_transform(x)

      X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.33,
      random_state=42)

      model = LogisticRegression()
      # Train the model
      model.fit(X_train, y_train)

      C:\Users\Dell\AppData\Local\Temp\ipykernel_21552\3461102990.py:4:
      DeprecationWarning: Converting `np.inexact` or `np.floating` to a dtype is
      deprecated. The current result is `float64` which is not strictly correct.
          if cleanDF[column].dtype == np.number:

      [10]: LogisticRegression()
```



Figura B11 – Protótipo completo regressão logística 11

```
[11]: predictions = model.predict(X_test)

# print the predictions
print(predictions)

[1 0 0 1 1 0 1 0 1 1 1 1 0 1 1 1 1 0 1 1 1 1 0 0 0 1 1 1 1 1 0 0 0 0 0 1 1 1 0 1
 1 1 1 0 1 1 0 1 0 1 0 1 0 1 1 0 1 1 1 1 1 1 1 1 1 0 0 1 1 0 0 1 0 0 1 1 0 1 1
 1 1 1 0 0 1 0 1 0 1 0 0 1 0 1 0 0 1 0 1 0 1 1 1 0 1 1 1 1 0 1 0 1 1 1 1 1 0
 1 1 1 1 1 1 0 0 0 1 1 1 0 1 1 1 0 1 1 1 0 1 0 1 0 0 0 0 0 0 0 1 1 1 0 1 1 0
 1 1 0 1 1 1 0 1 1 1 0 0 1 1 0 1 1 0 0 1 1 0 1 0 0 1 1 1 1 0 1 1 1 0]
```

```
[12]: print(classification_report(y_test, predictions))
```

	precision	recall	f1-score	support
0	0.71	0.60	0.65	82
1	0.71	0.80	0.75	100
accuracy			0.71	182
macro avg	0.71	0.70	0.70	182
weighted avg	0.71	0.71	0.71	182

```
[13]: #Médias
media_1 = tips02['User_Rating'].mean()
media_2 = tips02['Reviews'].mean()
media_3 = tips02['Price'].mean()

print('Média User_Rating:', media_1, ' ', 'Média Reviews:', media_2, ' ', 'Média_
->Price:', media_3)

Média User_Rating: 4.605263157894737 Média Reviews: 17387.947368421053 Média
Price: 11.31578947368421
```

```
[14]: #Desvio Padrão

std_1 = tips02['User_Rating'].std()
std_2 = tips02['Reviews'].std()
std_3 = tips02['Price'].std()

print('Desvio User_Rating:', std_1, ' ', 'Desvio Reviews:', std_2, ' ', 'Desvio_
->Price:', std_3)

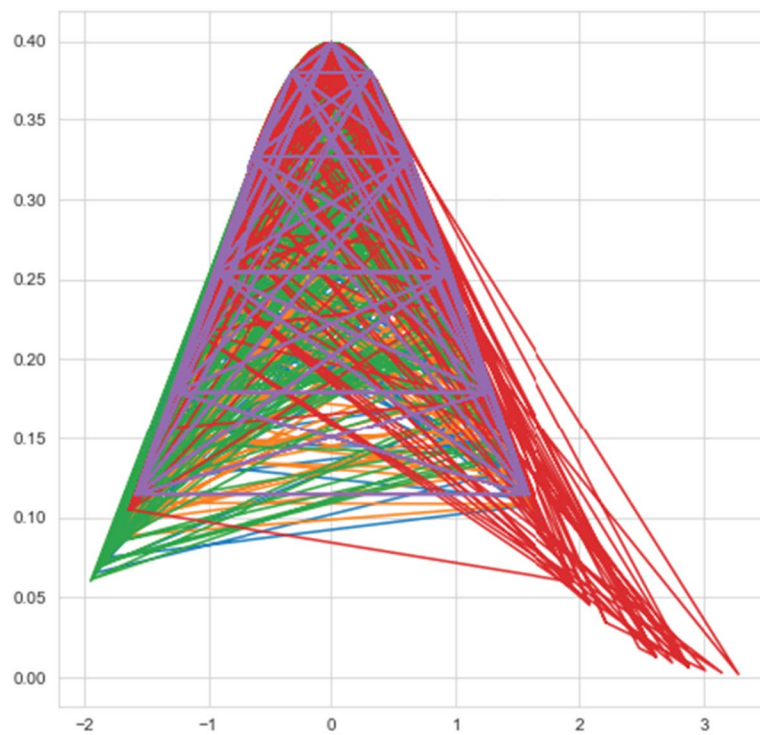
Desvio User_Rating: 0.22478059477960663 Desvio Reviews: 9751.120867046142
Desvio Price: 6.658767249765507
```

```
[15]: #Curva normal 3D
mu = 0
std = 1
```



Figura B12 – Protótipo completo regressão logística 12

```
snd = stats.norm(mu, std)
plt.figure(figsize=(7.5,7.5))
plt.plot( $\lambda$ _train, snd.pdf( $\lambda$ _train))
plt.show()
```



[]:

[]:

[]:

[]:

[]:



Universidade Presbiteriana Mackenzie

