Universidade Federal do Paraná
Setor de Ciências Exatas
Departamento de Estatística
Programa de Especialização em *Data Science* e *Big Data*

Marcio de Liz

Consumer Review - Foco na Experiência do Consumidor

Curitiba
2019

Marcio de Liz

Consumer Review - Foco na Experiência do Consumidor

Monografia apresentada ao Programa de Especialização em *Data Science* e *Big Data* da Universidade Federal do Paraná como requisito parcial para a obtenção do grau de especialista.

Orientador: Prof. Luis Carlos Erpen de Bona

Curitiba
2019

Consumer Review - Foco na Experiência do Consumidor

Consumer Review - Consumer Ownership Experience

Marcio de Liz[1]

[1]Departamento de Estatistica, Universidade Federal do Paraná Centro Politécnico - Jardim das Américas, Curitiba , PR, Brasil*

Resumo

Nos últimos anos, o crescimento do volume de dados disponíveis na web modificou a forma como os indivíduos buscam informações, além do conteúdo disponibilizado pela mídia, os usuários expandiram sua função na navegação ou na pesquisa de conteúdo e começaram a compartilhar seus conhecimentos, experiências, críticas e opiniões em blogs pessoais, redes sociais, e-commerce e outras mídias. Com isso, o uso da web para criação e compartilhamento de dados gerou possibilidades de análise e mineração de dados. Uma das principais aplicações é estabelecer a opinião que os consumidores apresentam sobre um determinado produto ou serviço, uma vez que permitem que o relacionamento com o cliente se torne tão importante quanto os serviços e produtos oferecidos. Outro ponto é interagir permanentemente com seus clientes, tentando entender o que eles querem e o que pensam sobre a sua marca é uma informação muito valiosa, pois permite que você esteja ciente dos pontos fortes e fracos do negócio nos olhos dos clientes e na melhoria contínua do produto ou serviço. Desta forma, a marca é capaz de usar essas informações como base para suas estratégias futuras. Ter revisões de consumidores credita confiança a outros compradores em potencial, o que, por sua vez, pode deixar avaliações positivas e gerar credibilidade para outros consumidores, e assim por diante. Esse ciclo aumenta as vendas junto com a credibilidade da marca e melhora a experiência do consumidor.O objetivo do trabalho acadêmico é criar uma ferramenta de coleta e processamento de dados de sites específicos relacionados ao comportamento do consumidor em relação ao produto e a influência de sua opinião sobre as compras futuras. Essas avaliações do consumidor são padronizadas em uma escala de 1 a 5 estrelas e são acompanhadas de opinião sobre o produto.

Palavras-chave: Avaliação do consumidor, experiência do consumidor, mineração de dados, mineração de texto , nuvem de palavras

Abstract

In recent years the data amount growth available on web has modified the way that individuals search for information, in addition to the content made available by the media, users have expanded their role in browsing or searching for content and have started to share their knowledge, experiences , reviews and opinions on personal blogs, social networks, e-commerce and other media. herewith the web use for creation and sharing of data has generated analysis possibilities and datamining . One of the mainly applications is to establish the opinion that consumers present about a particular product or service, as they allow effective manner customer relationship has become as important as the services and products it offers , another point is to interact permanently with their clients, trying to understand what they want and what they think about their brand is very valuable information, because it allows you to be aware of the business strengths and weaknesses points in the customers eyes and continuous improvement of the product or service. In this way, the brand is able to use such information as the basis for its future strategies. Having consumer reviews credits trust to other potential buyers which in turn can leave positive ratings and generate credibility for other consumers and so on, this cycle causes sales increase along with brand credibility and improving consumer experience.The academic paper objective is to create a collecting tool and processing data from specific sites related to consumer behavior towards the product and the influence of their opinion on future purchases. These consumer reviews are standardized on a scale of 1 to 5 stars and it is accompanied by opinion about product.

Keywords: consumer review, consumer experience, data mining, text mining, wordcloud

1. Introdução

Vou começar com uma afirmação óbvia: somos todos consumidores. Professores, profissionais de marketing, de tecnologia, analistas de Big Data, programadores digitais. Eu e você. Somos todos consumidores. E todos, sem exceção, somos impactados e interagimos seja curtimos páginas, seguimos empresas, aplaudimos e recomendamos seus serviços, compartilhamos críticas e as abandonamos. Entre experiências boas e ruins, o que muita gente ainda não se deu conta, porém, é a dimensão do prejuízo que certas experiências negativas acarretam para as marcas. Nos últimos anos estamos cada vez mais conectados e exigentes, os consumidores formam julgamentos sobre as empresas com base na forma como ocorre esse relacionamento nos múltiplos pontos de contato. Como vivemos a era dos negócios baseados na experiência, proporcionar uma ótima experiência do cliente, ou Customer Experience (CX, em inglês), é essencial para qualquer empresa , e torná-la cada vez melhor, indo além das expectativas e necessidades do seu público, deve ser sempre uma meta.Esse é como uma filosofia ou um valor empresarial, que deve ser incorporado por toda a organização em que o foco é gerar uma experiência positiva para o cliente com a marca considerando todas as etapas da jornada do cliente, vale ressaltar que quando falamos na experiência do cliente, devemos levar em consideração a sua percepção racional, física e emocional ao entrar em contato com uma empresa, independentemente do meio de comunicação utilizado ou canal escolhido. Quanto mais fácil, intuitiva e agradável for a jornada do cliente com a marca, melhor será a experiência que ele terá , isso leva á compreender a exata da jornada dos clientes (customer lifecycle journey).

As principais etapas da jornada do cliente são:

- Reconhece o problema, necessidade e desejo que tem;
- Pesquisa por informações para solucionar essas questões;
- Encontra as soluções oferecidas no mercado e seleciona seus preferidos;
- Realiza a compra efetivamnete.;
- Recebe o produto;
- Usa o produto;
- Mantém o produto;
- **Recomenda** o produto/empresa.

Figura 1: Diagrama do ciclo de vida do consumidor. Com consumidores exigentes é essencial prestar a atenção em cada etapa da jornada do comsumidor e, que é a trajetória pela qual o cliente passa entre a descoberta de um problema até a compra de um produto ou serviço.

Esse trabalho tem o intuito de entender melhor o que o consumidor gosta ou não e o que precisa ou não por meio de **monitoramento de reviews (avaliações online)** , assim como responder adequadamente aos reviews , ajudam a construir a credibilidade e faz com que a companhia transforme o consumidor em um embaixador da marca que, por consequência, tendem a impactar de forma positiva outros consumidores no momento de decisão de compra. Resumindo , quanto melhor for a experiência geral do cliente, mais provável será que ele promova a marca e defenda a empresa em discussões. O resultado será uma atração maior, mais fácil e barata de novos clientes.

2. A importância dos dados

A experiência do consumidor, em breve, será o principal diferencial competitivo entre as empresas e para guiar uma boa experiência nada melhor que a utilização dos dados.

Como a maioria das organizações diz ser a chave para enfrentar os desafios da experiência do consumidor?

Tornando-se mais orientado a dados. Por conta disso, trazemos 16 estatísticas mapeadas pelo Business 2 Community para evidenciar a importância dos dados na experiência do consumidor [**?**].

1. 73% dos líderes de negócios dizem que fornecer uma experiência de cliente relevante e confiável é fundamental para o desempenho geral dos negócios da empresa hoje. 93% concordam que será ainda mais daqui a dois anos. – Harvard Business Review
2. As organizações orientadas por dados têm 23 vezes mais chances de adquirir clientes, seis vezes mais chances de reter clientes e 19 vezes mais

*marcio.liz@electrolux.com

chances de serem lucrativas como resultado. – McKinsey Global Institute

3. Em uma pesquisa com aproximadamente 700 profissionais de negócios, apenas 15% disseram que sua organização é atualmente muito eficaz em proporcionar uma experiência relevante e confiável ao cliente (53% dizem que são um pouco eficazes e 32% dizem que não são muito eficazes). – Harvard Business Review
4. Apenas 3% dos entrevistados disseram que podem atuar em todos os dados do cliente que coletam; 21% dizem que podem agir com muito pouco. –Harvard Business Review
5. De acordo com uma pesquisa da Econsultancy com a Adobe, feita com profissionais de Customer Service de todo o mundo, 65% dos entrevistados disseram que aprimorar os recursos de análise de dados é o fator interno mais importante para proporcionar uma excelente experiência. —Digital Intelligence Briefing: 2018 Digital Trends
6. Em uma pesquisa de 2018 sobre os principais tomadores de decisões dos EUA, Big data e Analytics foram listados como as tecnologias emergentes mais importante para melhorar a experiência do consumidor. – Verndale Customer Experience (CX) Journey Research.
7. No relatório Global Customer Experience Benchmarking Report, a análise do cliente foi classificada como fator número 2 na melhoria positiva da experiência dos clientes. Nos próximos anos será o número 1. – Dimension Data
8. Na mesma pesquisa, apenas 48% dos entrevistados disseram que suas organizações atualmente têm sistemas analíticos e apenas 36% possuem grandes soluções analíticas de dados. – Dimension Data
9. 42% dos entrevistados dizem que seus sistemas analíticos não atendem às necessidades atuais. –Dimension Data
10. 92% dos profissionais de marketing veem a personalização como um elemento “crucial” da experiência do cliente, mas 51% dizem que sua organização não pode oferecer a personalização que seus clientes anseiam. –Verndale Solving for CX Survey.
11. As principais necessidades para melhorar a personalização da experiência do cliente são insights em tempo real (46%), reunindo mais dados do cliente (40%) e uma análise maior dos dados do cliente (38%). –Verndale Solving for CX Survey.
12. Profissionais de marketing em todo o mundo dizem que seu principal desafio na execução de uma estratégia de experiência do cliente baseada em dados é um sistema fragmentado para fornecer uma visão unificada da experiência do cliente em pontos de contato (38%). Seguido por silos de dados de clientes que permanecem inacessíveis por toda a organização. (30%) – CMO Council, Empowering the Data-Driven Customer Strategy
13. Apenas 7% dos profissionais de marketing pesquisados relatam que estão efetivamente em condições de oferecer engajamentos de marketing em tempo real, orientados por dados, em pontos de contato físicos e digitais. CMO Council, Empowering the Data-Driven Customer Strategy
14. 63% dos profissionais de marketing dizem que a funcionalidade big data analytics seria fundamental para melhorar o CX em sua organização. –Verndale Solving for CX Survey
15. As organizações vêem o aprendizado automático como um fator chave para ajudar a processar grandes quantidade de dados de clientes (49%), fornecer análises em tempo real (49%) e criar modelos de precificação mais precisos (49%) – Verndale Solving for CX Survey
16. Em 2018, 2/3 das empresas criarão centros de experiência para os clientes – Forrester Predictions 2018: AI Hard Fact—Treat It Like a Plug-And-Play Panacea and Fail.

3. Consumer Review

Uma das principais características do consumidor online é o seu forte comportamento de pesquisa. Com o avanço da tecnologia, tornou-se possível buscar diversas opções de produtos com apenas alguns cliques.

Qualidade, marca e preço baixo já não são características tão relevantes no momento de decidir uma compra. Afinal, é possível fazer esse comparativo em poucos minutos no Google. As empresas precisam apostar em diferenciais que possam chamar a atenção do consumidor em meio a tantas informações, e os reviews ajudam nesse momento. Para que o review traga confiança, é fundamental que ele seja real. Optar por não publicar as críticas pode ser um tiro no pé, já que os clientes “censurados” podem expressar sua indignação nas redes sociais, causando dor de cabeça e prejudi-

cando a reputação da sua empresa. A opinião de terceiros pode influenciar diretamente a decisão de compra do consumidor online. Apostar na honestidade é uma forma de melhorar sua reputação e fidelizar o cliente.

A importância do Consumer Review [**?**].

3.1. A importância do Consumer Review [?]

Figura 2: Para quase 9 em 10 consumidores, uma review é tão importante quanto uma recomendação pessoal.

- 90% dos consumidores lêem avaliações on-line antes de realizar a compra.
- É provável que os clientes gastem 31% a mais em uma empresa com avaliações “excelentes”.
- 72% dizem que críticas positivas fazem com que eles confiem em uma empresa local mais
- 92% dos usuários usarão uma empresa local se ela tiver pelo menos uma classificação de 4 estrelas
- 72% dos consumidores agem apenas depois de ler um comentário positivo
- 86% das pessoas hesitarão em comprar de uma empresa que tenha comentários on-line negativos.

A opinião dos consumidores é parte importantíssima da estratégia de qualquer empresa. Saber o que os consumidores pensam sobre sua marca é uma informação muito valiosa, pois permite que você. Tenha ciência dos pontos fortes e fracos do negócio aos olhos dos clientes e que aprimore suas ações constantemente , isso significa que esta relacionado com Qualidade durante todo o ciclo de vida do consumidor pois qualidade nos dias de hoje pode ser definida por:

Definição 1 *A Qualidade deve ser a razão que explique por que os consumidores escolhem os nossos produtos em detrimento da concorrência, quer se trate de uma primeira* ***compra, recomendação ou recompra.***

Figura 3: O que os consumidores esperam?

4. Mineração de Dados

Por ser considerada multidisciplinar, as definições acerca do termo Mineração de Dados variam com o campo de atuação dos autores. Destacamos neste trabalho três áreas que são consideradas como de maior expressão dentro da Mineração de Dados: Estatística, Aprendizadode Máquina e Banco de Dados. Em Zhou [**?**] é feita uma análise comparativa sobre as três perspectivas citadas.

Definição 2 *Em Hand et al.[**?**] , a definição é dada de uma perspectiva estatística: "Mineração de Dados é a análise de grandes conjuntos de dados a fim de encontrar relacionamentos inesperados e de resumir os dados de uma forma que eles sejam tanto úteis quanto compreensíveis ao dono dos dados"*

Definição 3 *Em Cabena et al. [**?**], a definição é dada de uma perspectiva de banco de dados: "Mineração de Dados é um campo interdisciplinar que junta técnicas de aprendizado de máquinas, reconhecimento de padrões, estatísticas, banco de dados e visualização, paraconseguir extrair informações de grandes bases de dados"*

Definição 4 *Em Fayyad et al. [**?**], a definição é dada da perspectiva do aprendizado de máquina:"Mineração de Dados é um passo no processo de Descoberta de Conhecimento que consiste na realização da análise dos dados e na aplicação de algoritmos de descoberta que,sob certas limitações computacionais, produzem um conjunto de padrões de certos dados."*

A mineração de textos é uma extensão da mineração de dados, e pode ser definida como um processo

de extração de informações desconhecidas e úteis de documentos textuais escritos em linguagem natural. Como a maioria das informações são armazenadas em forma de texto, a mineração de textos possui alto valor comercial, e pode ser aplicada para melhoria de produtos e serviços.

O principal objetivo da mineração de textos é encontrar termos relevantes em documentos de texto com grande volume de dados e estabelecer padrões e relacionamentos entre eles com base na frequência e temática dos termos encontrados [**?**]

A mineração de textos pode conter várias etapas, mas cinco delas são básicas em todos os processos: coleta, pré-processamento, Indexação , Mineração e Análise. [**?**]

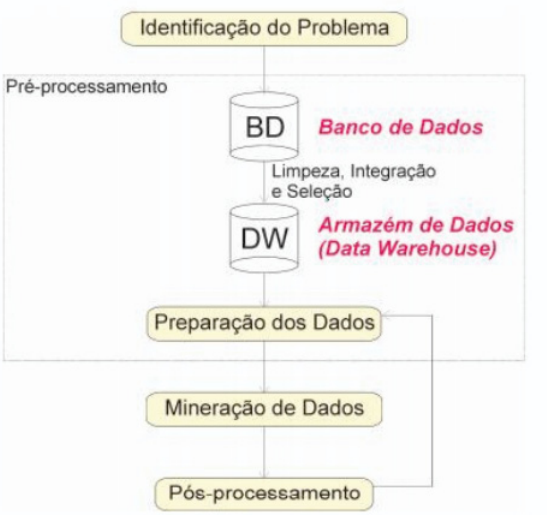

Figura 4: Etapas do processo de Mineração de Texto.

Cada uma destas etapas será aplicada no estudo de Consumer Reviews proposto e terão seus resultados apresentados Nesse trabalho.

1. Identificação do Problema: Importante coletar e analisar os reviews para melhoria de qualidade percebida pelo consumidor através de sua opinião sobre um produto ou serviço.
2. Pré-Processamento:

   Limpeza, integração e seleção dos dados: dados dos Reviews foram carregados no R, a coluna das opiniões foi separada e as opiniões duplicados foram retiradas,

   Armazenagem dos dados: a coluna que contém os textos dos reviews formou o chamado corpus, que nada mais é que a coleção de textos.

   Preparação dos dados: remover dados desnecessários do corpus, tais como pontuação, números, urls , stopwords do negócio e stopwords.
3. Mineração dos dados: uma matriz de associação é criada para a identificação de frequência dos termos e associação entre eles. Gráfico de linhas e nuvem de palavras são feitos para mostrar termos mais frequentes.
4. Pós- Processamento: conclusões sobre as análises feitas e interpretações sobre a resolução do problema

5. Material e Método

Nesta seção será apresentado o software utilizado , mostrando também como os dados foram coletados e que técnicas de mineração de texto foram utilizadas na análise.

A linguaguem utilizada para esse artigo foi a linguagem R , versão 3.5.1, sendo escolhido por ser um software que contém recursos para coleta de dados através de web scraping e análise de Mineração de Texto, por ser gratuito. O Software R pode ser baixado diretamente na internet pelo site 'https://www.r-project.org/'. Além disso, o Software R tem amplo conjunto de funções e pode ser aperfeiçoado com o uso de novos pacotes, ou seja, é um programa bastante poderoso quanto a análises estatísticas.

5.1. Principais Etapas - Mineração de Texto

5.1.1. Coleta de Dados

Tem como objetivo formar a base textual que irá ser processada , que foi adquirida utilizando a Técnica de WEB SCRAPING(DATA SCRAPING) , que é uma técnica na qual um programa de computador extrai dados de saída legível para humanos, proveniente de um outro programa, e disponibiliza esses dados de modo que se tornem legíveis para outros programas de computador.Scraping é a atividade de extrair dados de sites e transportá-los para um formato mais simples e maleável para que possam ser analisados e cruzados com mais facilidade. Muitas vezes a informação necessária para reforçar uma história está disponível, mas em sites de navegabilidade ruim ou em bancos de dados difíceis de manipular. Para coletar automaticamente e visualizar essas informações, recorre-se a softwares conhecidos como scrapers.[**?**]

Definição 5 *Em Zeviani[**?**] "Web Scrap é a ação ou conjunto de técnicas usadas para fazer consumo de infor-*

mações de web sites. Também chamada de web harvest, com web scraping é possível capturar dados de texto e não textuais que estão abundantes na internet. A análise de dados da Web, seja de texto ou não, tem sido utilizada para otimizar diversas decisões de mercado que incluem: design de produto, análise de sentimento, monitoramento de doenças, modelagem preditiva, dentre várias outras."

A raspagem de dados da Web é uma técnica de software de computador para extrair informações de sitese consiste principalmente na transformação de dados não estruturados(formato HTML) na web em dados estruturados(banco de dados ou planilha eletrônica). Alguns exemplos de ferramentas de Web Scraping: Scrapy, Web Scraper, Apache Camel, Import io, entre outras. Também é comum a raspagem de dados utilizando linguagem de programação como python usando a biblioteca BeautifulSoup e com a linguagem R usando o pacote rvest.[**?**]

A linguaguem R possui os pacotes httr, xml2 e rvest, Esses são os três pacotes mais modernos do R utilizados para fazer web scraping. O pacote xml2 tem a finalidade de estruturar arquivos HTML ou XML de forma eficiente, tornando possível a obtenção de tags e seus atributos dentro de um arquivo. Já o pacote httr é responsável por realizar requisições web para obtenção das páginas de interesse, buscando reduzir ao máximo a complexidade da programação. O pacote rvest é escrito sobre os dois anteriores e por isso eleva ainda mais o nível de especialização para raspagem de dados.

As características dos pacotes implicam na seguinte regra de bolso. Para trabalhar com páginas simples, basta carregar o rvest e utilizar suas funcionalidades. Caso o acesso à página exija ações mais complexas e/ou artifícios de ferramentas web, será necessário utilizar o httr. O xml2 só será usado explicitamente nos casos raros em que a página está em XML, que pode ser visto como uma generalização do HTML.Esses pacotes não são suficientes para acessar todo tipo de conteúdo da web. Um exemplo claro disso são páginas em que o conteúdo é produzido por javascript, o que acontece em muitos sites modernos. Para trabalhar com esses sites, é necessário realmente "simular" um navegador que acessa a página web. Uma das melhores ferramentas para isso é o Rselenium.

A coleta de dados foi realizada nos sites da B2W - Companhia Digital, que correspondem a três sites especificos,**americanas, submarino e shoptime** , através de Web scraping utilizando a linguaguem R e os pacotes Rvest e Rselenium, foram coletados produtos da categoria de Eletroportateis e das sub categorias abaixo:

- Aspirador de Pó
- Batedeira
- Bebedouro e Purificador de Água
- Cafeteira
- Centrífuga e Espremedor de Fruta
- Chaleira Elétrica
- Churrasqueira Elétrica
- Ferro de Passar
- Forno Elétrico
- Fritadeira Elétrica
- Grill, Sanduicheira e Torradeira
- Liquidificador
- Máquina de Costura
- Processador de Alimentos
- Mixer
- Panela Elétrica
- Vaporizador e Higienizador
- Passadeira a Vapor

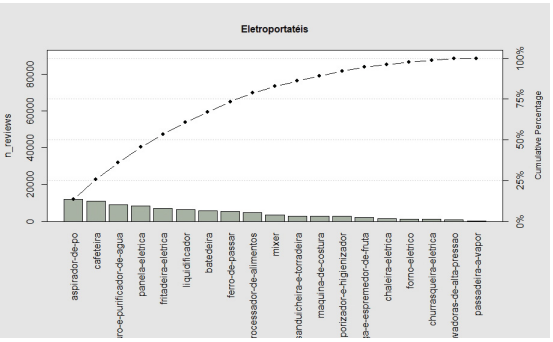

Figura 5: Pareto- Número de reviews por categoria.

Foram coletados aproximadamente 90000 reviews de 250 modelos de produtos diferentes e 30 marcas distintas. , o banco de dados foi criado com as informações abaixo:

- Data
- Categoria
- Marca
- Modelo
- Descrição do modelo
- Preço
- Nota média
- Percentual de Recomendação
- Nome do autor do review
- Titulo do review
- Opinião (Review)
- Nota (Quantidade de estrelas)

5.1.2. Análise descritiva

As avaliações do consumidor são normalizadas para uma escala de 1 a 5 estrelas, isso oferece várias opções para analisar os dados , como tendências , detalhamento em diferentes parâmetros, comparação de avaliações com concorrentes,

Utilizando a quantidade de estrelas foram criados Indicadores para monitoramento de performance dos produtos e serviços, como média das avaliações e o percentual de notas baixas((1, 2 ou 3 estrelas)

Definição 6 *Consumer Reviews Average (CRA) - O CRA é a média das notas das avaliações (1-5 estrelas).*

Definição 7 *Consumer Reviews Rate (CRR) que é porcentagem de avaliações baixas (1, 2 ou 3 estrelas) sobre o total de Avaliações.*

Os dados foram carregados no R através da função read.csv ,como apresentado a seguir.

```
cr_dsbd <- read.csv("eletroport.txt",
                    sep="|" ,
                    encoding="UTF-8")
```

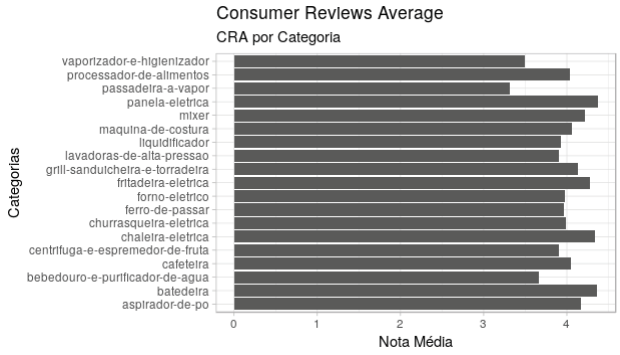

Figura 6: Gráfico CRA - Por Categoria de Eletroportáteis

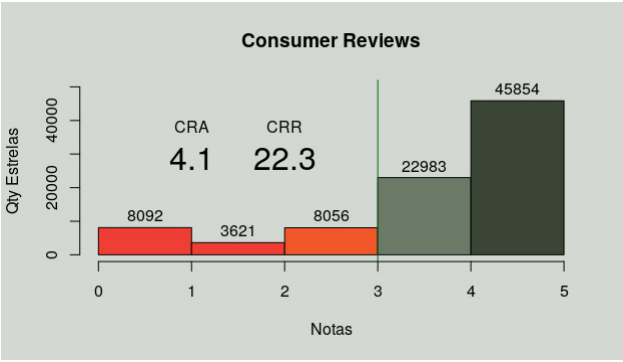

Figura 7: Histograma - Resumo das Avaliações

É comum que a gestão de uma empresa não consiga acompanhar todos os setores diariamente, Por isso, analisar apenas os números de cada departamento ajuda a ter uma visão do andamento dos resultados, e nesse cenário, os gráficos auxiliam muito. Sem uma boa analise de dados, não é possível medir, mensurar e melhorar a performance da companhia. Para observar os indicadores de modo mais visual, nada melhor que um conjunto de gráficos.

5.1.3. Limpeza, integração e seleção dos dados

Para esse estudo utilizarei apenas a categoria de Aspirador de Pó com aproximadamente 11 mil reviews e Como algumas colunas que contém informações, como, por exemplo, a da nota da pelo usuário, não serão utilizadas, foi criado um objeto chamado de 'crtext' que contém somente a coluna com as opiniões dos consumidores

```
category <- 'aspirador-de-po'
cr_dsbd <- subset(cr_dsbd,
          (cr_dsbd$categoria == category))

crtext <- cr_dsbd$opiniao
```

Com o objeto **crtext** formado, o próximo passo é preparar os dados, excluindo palavras sem significado, antes de as análises serem feitas. As letras foram transformadas todas para minúsculas, foram retiradas as pontuações, os números e os espaços duplicados e também foi necessária a conversão dos dados textuais para o padrão internacional. A função abaixo foi extraida de:
www.leg.ufpr.br/ walmes/ensino/mintex/tutorials/ word-embedding.html[**?**]

```
preprocess <- function(x) {
  x <- tolower(x)
  x <- gsub(pattern = "[[:punct:]]+",
            replacement = " ", x = x)
  x <- removeWords(x,
       words = stopwords("portuguese"))
  x <- removeWords(x,
       words = stopbussines)
  x <- removeNumbers(x)
  x <- gsub(pattern = "[[:space:]]+",
       replacement = " ", x = x)
  x <- iconv(x, to = "ASCII//TRANSLIT")
  x <- trimws(x)
  return(x)
}

cr_clean <- preprocess(crtext)
```

5.1.4. Armazenagem dos dados

A parte textual foi extraída a partir do comando 'Corpus' para a construção da coleção de textos que nesse trabalho

foi chamado de **crcps**, A partir dessa parte o pacote de Text Mining 'tm' [**?**] será utilizado.

```
crcps <- VCorpus(VectorSource(x = cr_clean),
         readerControl = list(lang = "pt",
         load = TRUE))
```

5.1.5. Mineração dos dados

Após a preparação dos dados Uma matriz de associação é criada para a identificação de frequência dos termos e associação entre eles.

```
crcps <- VCorpus(VectorSource(x = cr_clean),
         readerControl = list(lang = "pt",
         load = TRUE))
```

Uma matriz de associação é criada para a identificação de frequência dos termos e associação entre eles. Gráfico com a nuvem de palavras foram construídos para mostrar termos mais frequentes.

5.1.6. Frequência de Palavras

As palavras que são as mais apresentadas no texto e podem aparecer com grande frequência foram contadas, para obter essa freqência o conjunto de textos é transformado em uma matriz de termos, contendo as palavras em cada linha e os documentos (cada review) em cada coluna. Sendo assim, obtém-se quantas vezes cada palavra apareceu em cada documento. Para esse artigo o conceito de bigrama será utilizado devido à sua simplicidade de obtenção através de uma abordagem estatística para extração de termos. Para a obtenção de bigramas basta combinar palavras duas a duas e iterar o procedimento para todos os termos de um texto.

Definição 8 *N-gramas podem ser classificados como uma seqüência de caracteres de um texto agrupados conforme a aplicação. Existem nominações que representam o número de palavras que compõem um ngrama, a saber: unigrama, bigrama e trigrama, que respectivamente são formados por um, dois e três termos.*

```
tdm = TermDocumentMatrix(crcps)
```

Com a matriz de termos 'tdm', é possível inspecionar as palavras mais frequentes. Usando uma frequência mínima de 100 (lowfreq=100) pode-se usar a função 'findFreqTerms'para mostrar as palavras (também chamadas de termos) que aparecem nessa condição.

```
mft <- findFreqTerms(tdm, lowfreq = 100)
```

Com essa informação de palavras mais frequentes é possível saber os assuntos que aparecem nos reviews. Além de saber o assunto, é uma forma de saber se a análise está seguindo o caminho correto, pois mostra as palavras que seriam supostamente esperadas. Caso aparecessem palavras que não tenham nenhum sentido ou relação com o produto ou serviço que está sendo avaliado, isso seria sinal de que algo estaria errado na análise. Nesse caso temos palavras condizentes com o produto, o que indica que a análise está seguindo o caminho correto. Para observar a frequência dos termos de modo mais visual, um gráfico de barras foi feito, mostrando os termos com suas frequências

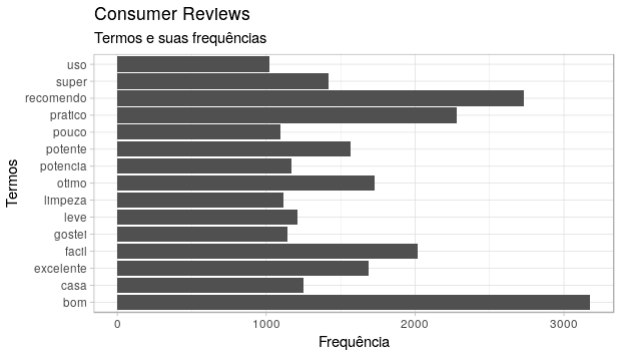

Figura 8: Gráfico de termos e suas frequências

A nuvem de palavras é feita com as palavras de maior frequência nos reviews. Quanto maior a frequência da palavra, maior o tamanho da fonte da palavra que é apresentada. Para ser feita a nuvem de palavras é necessária instalação do pacote denominado 'wordcloud'[**?**]. Outro pacote que também foi usado foi o 'RColorBrewer'[**?**], para a colocação de cores diversas na nuvem de palavras, fazendo com que visualmente fique mais fácil localizar os termos.

Figura 9: Unigrama - Nuvem de palavras com reviews coletados

6. Considerações Finais

A análise textual cresceu bastante nos últimos anos, pois cada vez mais as pessoas usam a internet para transmitir informações em texto, e isso, muitas vezes, pode ser uma informação muito valiosa para os gestores. Atualmente, um grande nicho de informação pode ser encontrado nas diver-

sas redes sociais, sendo gratuita e de acesso público. Desta forma, obter esses dados e utilizar o Mineração de Texto resultará em informação consistente para propor novas ideias, tomar decisões e ajustar os negócios de acordo com o público alvo. A utilização da Análise de Mineração de texto é de grande importância para obter informações sobre textos sem uma leitura prévia. Sendo assim, é possível obter informação dos textos e utilizá-los como auxílio na procura de melhorias para produtos e serviços, melhorando a Experiência do Consumidor

Referências

[1] F; Ingo,Hornik; Kurt , *tm: Text Mining Package ,Version: 0.7-6*, (https://CRAN.R-project.org/package=tm, 2018).

[2] Fellows; Ian, *wordcloud2: Create Word Cloud by 'htmlwidget' ,Version:0.2.1*, (https://CRAN.R-project.org/package=wordcloud , 2018).

[3] Lang; Dawei, Chien; Guan-tin, *wordcloud: Word Clouds ,Version: 2.6*, (https://CRAN.R-project.org/package=wordcloud2 , 2018).

[4] Neuwirth; Erich , *RColorBrewer: ColorBrewer Palettes, Version:1.1-2*, (https://CRAN.R-project.org/package=RColorBrewer , 2014).

[5] Wickham Hadley ; Chang Winston;Pedersen; Thomas Lin , Takahashi;Kohske , Wilke ;Claus ,Woo ;Kara , Yutani; Hiroaki , *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics, Version:3.2.0*, (https://CRAN.R-project.org/package=ggplot2 , 2019).

[6] Scrucca ; Luca , Snow Greg, Bloomfield;Peter , *qcc: Quality Control Charts, Version:2.7*, (https://CRAN.R-project.org/package=qcc , 2017).

[7] Wickham;Hadley , François; Romain , Henry ;Lionel , Müller , Kirill, , *dplyr: A Grammar of Data Manipulation, Version:0.8.1*, (https://CRAN.R-project.org/package=dplyr , 2019).

[8] Wickham; Hadley , *tidyverse: Easily Install and Load the 'Tidyverse', Version:1.2.1*, (https://CRAN.R-project.org/package=tidyverse , 2017)

[9] Wickham; Hadley , *rvest: Easily Harvest (Scrape) Web Pages, Version:0.3.4*, (https://CRAN.R-project.org/package=rvest , 2019)

[10] Harrison ;John ,Kim ,; Ju Yeong , *RSelenium: R Bindings for 'Selenium WebDriver', Version:1.7.5*, (https://CRAN.R-project.org/package=RSelenium , 2019)

[11] ZHOU, Z.-H, *Three perspectives of data mining*, (Artificial Intelligence Journal, p.139–146, 2003).

[12] HAND, D; MANNILA, H; SMYTH, *Principles of Data Mining*, (MIT Press, London, 2001).

[13] CABENA, P; HADJINIAN, P; STADLER, R; JAAP-VERHEES; ZANASI, P. , *From Data Mining to Knowledge Discovery in Databases*, (MIT Press, 2001).

[14] Morais, Edison Andrade M.; AMBROSIO, Ana Paula L. , *Mineração de Texto - Relatório Técnico*, (Universidade Federal de Goiás, 2007).

[15] Zeviani ,Walmes; *Mineração de Texto:http://www.leg.ufpr.br/ walmes/ensino/mintex/index.html*, (Acesso em 16 06 2019).

[16] Geneze ,Pedro; *https://blog.neoassist.com/experiencia-do-consumidor-dados/*, (Acesso em 16 06 2019).

[17] Saleh ,Khalid; *https://www.invespcro.com/blog/the-importance-of-online-customer-reviews-infographic/*, (Acesso em 14 06 2019).

[18] SERAPIÃO, Paulo Roberto Barbosa; *Uso de mineração de texto como ferramenta de avaliação da qualidade informacional em laudos eletrônicos de mamografia. 2010. Disponível em:http://www.scielo.br/pdf/rb/v43n2/a10v43n2.pdf*, (Acesso em 17 06 2019).

[19] Usama Fayyad ; Gregory Piatetsky-shapiro ; Padhraic Smyth, *From Data Mining to Knowl-edge Discovery in Databases*, (American Association for Artificial Intelligence, 1996).

[20] Andriolo, Eric , *Desvendando 'Data Scraping': Entenda como raspar dados pode facilitar o trabalho jornalístico. [Online] Available at: https://knightcenter.utexas.edu/pt-br/blog/00-9586-desvendando-o-data-scraping-entenda-como-raspar-dados-pode-facilitar-o-trabalho-jornali*, (Acesso em 16 06 2019).

[21] Ray ,Sunil ; *Beginner's guide to Web Scraping in Python using BeautifulSoup. [Online] Available at: https://www.analyticsvidhya.com/blog/2015/10/beginner-guide-web-scraping-beautiful-soup-python/*, (Acesso em 16 06 2019).