

Matheus Medeiros Anacleto

**UM MODELO BASEADO EM ANÁLISE DE SENTIMENTOS
COMO SUPORTE À SISTEMAS DE RECOMENDAÇÃO**

Dissertação submetida ao Programa de
Pós-Graduação em Tecnologias de
Informação e Comunicação da
Universidade Federal de Santa
Catarina para a obtenção do Grau de
Mestre em Tecnologias de Informação
e Comunicação.

Orientador: Prof. Dr. Alexandre
Leopoldo Gonçalves.

Coorientador: Prof. Dr. Robson
Rodrigues Lemos

Araranguá
2017

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Anacleto, Matheus Medeiros

Um modelo baseado em análise de sentimentos como suporte à sistemas de recomendação / Matheus Medeiros Anacleto ; orientador, Alexandre Leopoldo Gonçalves, coorientador, Robson Rodrigues Lemos, 2017.

131 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Campus Araranguá, Programa de Pós-Graduação em Tecnologias da Informação e Comunicação, Araranguá, 2017.

Inclui referências.

1. Tecnologias da Informação e Comunicação. 2. Tecnologias da Informação e Comunicação. 3. Sistemas de Recomendação. 4. Análise de Sentimentos. 5. Gestão de ideias. I. Gonçalves, Alexandre Leopoldo. II. Lemos, Robson Rodrigues. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Tecnologias da Informação e Comunicação. IV. Título.

Matheus Medeiros Anacleto

UM MODELO BASEADO EM ANÁLISE DE SENTIMENTOS COMO SUPORTE À SISTEMAS DE RECOMENDAÇÃO

Esta Dissertação foi julgada adequada para obtenção do Título de “Mestre em Tecnologias da Informação e Comunicação”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Tecnologias da Informação e Comunicação.

Araranguá, 01 de Junho de 2017.

Prof.^a Andréa Cristina Trierweiller, Dr.^a
Coordenadora do Curso

Banca Examinadora:

Prof. Alexandre Leopoldo Gonçalves, Dr.
Orientador
Universidade Federal de Santa Catarina

Prof. Robson Rodrigues Lemos, Dr.
Coorientador
Universidade Federal de Santa Catarina

Prof.^a Olga Yevseyeva, Dr.^a
Universidade Federal de Santa Catarina

Prof. Roderval Marcelino, Dr.
Universidade Federal de Santa Catarina

Prof.^a Cristiane Raquel Woszezenki, Dr.^a
Instituto Federal de Santa Catarina

Este trabalho é dedicado à minha família, em especial minha esposa Heloisa e minha filha Alice.

AGRADECIMENTOS

Esta dissertação é o resultado de muito trabalho em conjunto. A todos que deram sua contribuição direta e indireta, quero utilizar este espaço para demonstrar minha gratidão.

Primeiramente agradeço a Deus a quem devo a vida. Obrigado pela saúde, pela força em todos os momentos, e por colocar no meu caminho pessoas fundamentais para a realização deste trabalho. Agradeço aos meus pais Amarildo Anacleto e Dilma Medeiros por terem me ensinado a batalhar honestamente pelos meus objetivos.

À minha esposa Heloisa, que durante toda minha formação academia tem assumido a postura firme de apoio irrestrito e total. Obrigado pelo amor, carinho e atenção sempre demonstrados nos momentos exatos. Obrigado por cuidar de mim e ser parte fundamental de quem eu sou. Obrigado pela filha linda que você me deu durante a construção desta dissertação. Amo você, amo nossa filha Alice.

Ao meu grande amigo e orientador, Professor Dr. Alexandre Leopoldo Gonçalves. Uma parte fundamental para a realização deste trabalho e de quem eu cultivo profunda admiração e sempre me serve de inspiração nos momentos difíceis. Obrigado pela sua interminável paciência, pela sua dedicação e pelas longas e inspiradoras horas de aconselhamento. Obrigado por me orientar não só neste trabalho, mas na minha vida como um todo.

Agradeço ao Programa de Pós-Graduação em Tecnologias da Informação e Comunicação pela oportunidade de participar deste curso e a todos os professores e demais profissionais do PPGTIC.

Aos meus pastores Vladimir e Vандir, aos demais amigos e colegas de turma que contribuíram com apoio e trocando experiências.

“Mediocridade é falta de capricho. Capricho é fazer o melhor na condição que se tem, até que se tenha condição de fazer melhor ainda. “

(Mario Sérgio Cortella)

RESUMO

Com a evolução da Web, seus usuários que outrora eram simples consumidores de conteúdo, absorveram também o papel de produtores de conteúdo. Esta alteração foi crucial para o aumento exponencial na produção de dados que é responsável pela sobrecarga de informação, gerando desafios na tomada de decisão, tornando uma simples escolha em um trabalho por vezes difícil para usuários pouco experientes. Este cenário serviu como um campo fértil para o surgimento dos Sistemas de Recomendação, que objetivam auxiliarem pessoas em suas escolhas. Ainda no contexto atual da Web, comentários efetuados por usuários a determinado item podem ajudar durante uma escolha. Para promover tal suporte surge a área de Análise de Sentimentos que objetiva extrair, de opiniões expressas via texto, informações que possibilitem qualificar uma opinião, por exemplo, como positiva ou negativa. A junção destas duas áreas fornece um importante ferramental que pode ser aplicado em uma ampla variedade de cenários. A partir disso, este trabalho propõe um modelo baseado em Análise de Sentimentos que promova suporte aos Sistemas de Recomendação. O modelo proposto foi avaliado utilizando o domínio da Gestão de Ideias visto a importância desta área no processo de inovação e competitividade para as organizações. Para demonstração de viabilidade do modelo proposto, foi desenvolvido um protótipo para suportar as fases de indexação, extração, avaliação e recomendação de ideias. O protótipo foi aplicado sobre um conjunto de ideias públicas. A partir das análises pode-se concluir que o modelo atinge seus objetivos uma vez que a análise de sentimentos demonstrou impacto positivo nas recomendações realizadas. Por fim, destaca-se que o modelo proposto através da integração entre as áreas de Análise de Sentimentos e os Sistemas de Recomendação produziu resultados capazes de auxiliar na tomada de decisão.

Palavras-chave: Sistemas de Recomendação. Análise de Sentimentos. Gestão de ideias.

ABSTRACT

With the evolution of the Web, its users who were once simple consumers of content also become producers. This change was crucial to the exponential increase in data production that is responsible for information overload phenomenon, generating challenges in decision making in which a simple choice sometimes becomes a difficult job for inexperienced users. This scenario served as a fertile field for the emergence of the Recommendation Systems, which aim to assist people in their choices. Still in the current context of the Web, comments made by users to a given item can help during a choice. To promote such support the Sentiment Analysis area arises aiming to extract, from opinions expressed via text, information that qualify an opinion as positive or negative, for instance. The combination of these two areas provides an important tool which can be applied in a wide variety of scenarios. From this, this work proposes a model based on Sentiment Analysis that promotes support to the Recommendation Systems. The proposed model was evaluated using the Idea Management domain, given its importance in the process of innovation and competitiveness for organizations. To demonstrate the viability of the proposed model a prototype was developed to support the phases of indexing, extracting, evaluating and recommending ideas. The prototype was applied on a set of public ideas. From the analysis, it can be concluded that the model achieves its objectives once the analysis of sentiments showed positive impact in the recommendations. Finally, it is highlighted that the proposed model through the integration between Sentiment Analysis and Recommendation Systems areas produced results capable of assisting decision making.

Keywords: Recommender Systems. Sentiment Analysis. Idea Management.

LISTA DE FIGURAS

Figura 1 - Sequência de etapas para a realização da dissertação.	37
Figura 2 - Comparação entre as classificações do usuário ativo e outros usuários.	50
Figura 3 - Filtragem Híbrida.	54
Figura 4 - Benefícios da aplicação do processo de Gestão de Ideias. ...	69
Figura 5 - Modelo Stage Gate	73
Figura 6 - Estágios da Inovação	73
Figura 7 - Túnel de ideias.....	74
Figura 8 - Metodologia de criação de ideias	75
Figura 9 - Modelo de planejamento da inovação.....	76
Figura 10 - Modelo lógico do sistema proposto.	79
Figura 11 - Modelo físico do sistema proposto.	82
Figura 12 - Retorno de uma busca textual.	85
Figura 13 - Retorno de recomendação.	86
Figura 14 - Review positivo de uma câmera digital extraído da Amazon®.....	89
Figura 15 - Léxico com suas colunas de palavras polarizadas.	90
Figura 16 - Ideia estruturada no formato XML.	91

LISTA DE QUADROS

Quadro 1 - Exemplo de busca por ideias que contenham a palavra Beverages.	101
Quadro 2 - Exemplo de recomendação de ideias com base em uma ideia de interesse.	102
Quadro 3 - Exemplo de busca por ideias que contenham a palavra Holiday.	103
Quadro 4 - Exemplo de recomendação de ideias com base na ideia eleita pelo usuário.	105
Quadro 5 - Relação das ideias implementadas pela Starbucks®.....	106
Quadro 6 - Detalhamento da ideia negativa implementada.	107

LISTA DE TABELAS

Tabela 1 - Base de dados de classificações.....	49
Tabela 2 - Banco de dados de avaliações médias ajustadas.	52
Tabela 3 - Matriz de contingência.....	92
Tabela 4 - Tabela com os testes iniciais utilizando o analisador base.	95
Tabela 5 - Tabela com os testes iniciais utilizando o analisador aprimorado.....	96
Tabela 6 - Tabela comparando o aprimorado com métodos estatísticos.	96
Tabela 7 - Significância estatística entre o método proposto e os demais.	97
Tabela 8 - Tabela com os testes iniciais utilizando o analisador base.	98
Tabela 9 - Tabela com os testes iniciais utilizando o analisador aprimorado.....	98
Tabela 10 - Tabela comparando o aprimorado com métodos estatísticos.	99
Tabela 11 - Significância estatística entre o método proposto e os demais.....	99

LISTA DE ABREVIATURAS E SIGLAS

API - *Application Programming Interface*

CSV - *Comma-Separated Values*

IBOPE – Instituto Brasileiro de Opinião Pública e Estatística

HTTP - *Hypertext Transfer Protocol*

JSON – *JavaScript Object Notation*

KNN – *k-Nearest Neighbors*

LSA – *Latent Semantic Analysis*

NB – *Naïve Bayes*

PLN – Processamento de Linguagem Natural

PMI – *Pointwise Mutual Information*

POS – *Part of Speech*

PPGTIC – Pós-Graduação em Tecnologias da Informação e Comunicação

RI – Recuperação de Informação

SR – Sistemas de Recomendação

SVM – *Support Vector Machines*

SML - *Supervised Machine Learning*

TF-IDF – *Term Frequency–Inverse Document Frequency*

UFSC – Universidade Federal de Santa Catarina

XML – *eXtensible Markup Language*

IMS – *Idea Management System*

SUMÁRIO

1	INTRODUÇÃO.....	27
1.1	PROBLEMÁTICA.....	30
1.2	OBJETIVOS	32
1.2.1	Objetivo geral.....	32
1.2.2	Objetivos específicos	33
1.3	JUSTIFICATIVA.....	33
1.4	ESCOPO DO TRABALHO.....	35
1.5	ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO	35
1.6	PROCEDIMENTOS METODOLÓGICOS	36
1.7	ESTRUTURA DO TRABALHO	37
2	SISTEMAS DE RECOMENDAÇÃO.....	39
2.1	ABORDAGENS	43
2.1.1	Filtragem Baseada em Conteúdo.....	43
2.1.1.1	Modelo Vetorial.....	44
2.1.1.2	Similaridades entre Vetores	45
2.1.1.3	Limitações	46
2.1.2	Filtragem Colaborativa	47
2.1.2.1	Recomendações de vizinhos próximos baseado em usuários	48
2.1.2.2	Recomendações de vizinhos próximos baseado em itens	51
2.1.2.3	Limitações	53
2.1.3	Filtragem Híbrida	53
2.2	APLICAÇÕES	54
3	ANÁLISE DE SENTIMENTOS.....	57
3.1	ABORDAGENS E MÉTODOS	59
3.1.1	Machine Learning Approach.....	60
3.1.1.1	Supervised Machine Learning (SML)	61
3.1.1.1.1	<i>Support Vector Machine (SVM)</i>	62
3.1.1.1.2	<i>Naive Bayes (NB)</i>	62

3.1.1.1.3	<i>K-Nearest Neighbors (KNN)</i>	62
3.1.1.2	Unsupervised Learning	63
3.1.1.2.1	<i>Pointwise Mutual Information (PMI)</i>	63
3.1.2	Lexicon Based Approach	63
3.1.2.1	Corpus-Based Approach.....	64
3.1.2.2	Dictionary-Based Approach	65
3.2	APLICAÇÕES	65
4	GESTÃO DE IDEIAS	67
4.1	SISTEMAS DE GESTÃO DE IDEIAS	71
4.2	MODELOS PARA GESTÃO DE IDEIAS	72
4.2.1	Modelo Stage Gate	72
4.2.2	Estágios da inovação	73
4.2.3	Metodologia de criação de ideias	74
4.2.4	Modelo de Desenvolvimento de Novo Conceito.	75
4.3	INICIATIVAS COM GESTÃO DE IDEIAS.....	76
5	MODELO PROPOSTO	79
5.1	MODELO LÓGICO	79
5.2	MODELO FÍSICO.....	81
5.2.1	Processo de Análise de Sentimentos	83
5.2.1.1	Analizador aprimorado	84
5.2.2	Processo de geração das recomendações	84
6	APLICAÇÃO DO MODELO E ANÁLISE DOS	
RESULTADOS		87
6.1	PROJETO DOS EXPERIMENTOS.....	87
6.1.1	Análise de Sentimento	88
6.1.2	Bases de Ideias	90
6.1.3	Métricas de Avaliação	92
6.1.3.1	Significância estatísticas e métricas de avaliação	93
6.1.3.2	Similaridades entre vetores	94
6.2	AVALIAÇÃO DOS EXPERIMENTOS	94

6.2.1	Análise de Sentimento.....	95
6.2.1.1	Conjunto de Dados de Câmeras.....	95
6.2.1.2	Conjunto de Dados de Filmes	97
6.2.1.3	Considerações sobre as Análises dos Conjuntos de dados..	100
6.2.2	Recomendações Baseadas em Conteúdo.....	100
7	CONSIDERAÇÕES FINAIS	109
7.1	CONSIDERAÇÕES FINAIS	109
7.2	SUGESTÕES DE TRABALHOS FUTUROS	111
	REFERÊNCIAS	113

1 INTRODUÇÃO

A *World Wide Web* ou apenas *Web* como é popularmente conhecida, foi criada em 1989 no CERN (*Conseil Européene pour la Recherche Nucléaire*) por Tim Berners-Lee. A *Web* surgiu com o objetivo de gerenciar a perda de informações ocasionada pela rotatividade de colaboradores do CERN (BERNERS-LEE, 1989). Ao expandir os limites do CERN, inicialmente a *Web* foi usada principalmente para disponibilizar informações aos leitores, semelhante à maneira como jornais e livros eram utilizados. Um grupo de pessoas especializadas era responsável por escrever, editar e publicar o conteúdo dos sites. Deste modo se consolidava o conceito de propriedade onde os sites eram proprietários e responsáveis pelo conteúdo publicado. Neste cenário pouco interativo, os usuários apenas tinham o papel de consumidores de informações (SEO; LEE, 2016).

Pode-se discernir diferentes estágios na evolução da *Web*, neste primeiro momento conhecido como *Web 1.0*. No início de 1990 os padrões estavam focados em como os documentos seriam processados por um navegador e como estes documentos seriam disponibilizados aos usuários. Dentre os padrões utilizados, o protocolo HTTP (*Hypertext Transfer Protocol*) e a linguagem HTML (*Hypertext Markup Language*) são considerados como importantes agentes no desenvolvimento da *Web* (HALL; TIROPANIS, 2012).

Com o passar do tempo a *Web* foi gradativamente evoluindo e em meados dos anos 2000, durante uma conferência que tratava da participação dos usuários da *Web*, Tim O'Reilly apresentou ao mundo o termo *Web 2.0* que marcaria o início do segundo estágio da *Web*, termo até hoje utilizado pela literatura (O'REILLY, 2005). Krotzfleisch et al. (2008) entendem que a *Web 2.0* representa um salto tecnológico, caracterizando uma nova configuração de tecnologias voltadas à internet, focada no usuário e no conteúdo e não somente em protocolos contrastando com sua versão anterior.

Na *Web 2.0* os usuários da Internet absorvem o papel de provedores de conteúdo. Este cenário é identificado através do comportamento dos usuários da *Web* que todos os dias ativa e voluntariamente geram conteúdos em portais (JANNACH et al, 2011, KROTZFLEISCH et al, 2008). Este participante híbrido que não só consome, mas também produz conteúdo é identificado na literatura através do termo “*prosumer*”. Este termo foi cunhado por Alvin Toffler em 1980 ao realizar a junção das palavras inglesas “*producer*” (produtor) e “*consumer*” (consumidor) (TOFFLER, 1980).

Para Eccleston e Griseri (2008), a *Web 2.0* foi introduzida como uma tecnologia inovadora que permitiu a comunicação mais interativa e personalizada entre as pessoas. Kim et al. (2009), resumem a *Web 2.0* como uma ferramenta que auxilia as pessoas a publicarem suas opiniões em sites, e também interagir e colaborar com outras pessoas em tópicos que sejam de seu interesse.

A *Web 2.0* despertou a atenção de muitas empresas devido a sua capacidade de melhorar a relação entre organizações e consumidores (ANDRIOLE, 2010). O grande potencial para construir relacionamentos mais fortes com os consumidores, levou a adoção rápida e inteligente de aplicações *Web 2.0* nas organizações. Atualmente, a maioria das principais organizações possuem contas ativas em redes sociais e interagem com seus consumidores através de várias aplicações, tais como Facebook®, Twitter® e YouTube® (SEO; LEE, 2016). Voigt e Ernst (2010) também destacam a importância da *Web 2.0* como uma parte fundamental no cotidiano de milhares de empresas, independente do ramo de negócios que estas atuam.

A cada dia mais a internet está inserida na vida das pessoas. Stephen (2016) faz um alerta para a taxa atual de uso da internet entre os adultos americanos que se aproxima a 87%; entre os jovens os níveis são igualmente elevados. O autor ainda destaca que no Reino Unido, durante a última década, o número de horas on-line por adultos atingiu as médias de 20,5 horas por semana. As redes sociais contribuem para esse crescimento: em todo mundo mais de 2 bilhões de pessoas usam as redes sociais e Facebook® sozinho tem cerca de 1 bilhão de usuários ativos por dia.

No Brasil os dados também se mostram elevados. Pesquisas demonstram que no segundo trimestre de 2012 o número de brasileiros com acesso à internet atingiu a marca de 83,4 milhões de pessoas (IBOPE,2012) e em Janeiro de 2013 o tempo utilizado pelos internautas brasileiros conectados a *Web* chegou a média de 10 horas e 26 minutos (IBOPE,2013).

A massificação do acesso à internet tem promovido a geração exponencial de conteúdos por parte dos usuários, resultando em uma explosão na produção diária de dados, alcançando o montante de 2.5 *exabytes* (1 EB = 1.000.000 *Terabytes*) (BELLO-ORGAZ; JUNG; CAMACHO, 2016). Em Julho de 2014, o ACI *Information Group* apresentou uma pesquisa contendo dados impressionantes a respeito da produtividade dos usuários da internet. Dentre outras informações, a pesquisa aponta, por exemplo, que a cada minuto o Twitter® recebe 300.000 novas postagens, usuários do Facebook® compartilham 2.5

milhões de publicações, o YouTube® recebe o upload de 72 horas de novos vídeos e a Amazon® realizou mais de \$80.000 em vendas online (ACI, 2014).

Diante deste quadro crescente na produção de dados, grande parte dos dados sem o devido tratamento são direcionados aos usuários gerando dificuldades na hora de realizar escolhas até mesmo para usuários experientes. Este cenário serviu como inspiração para o desenvolvimento de sistemas computacionais capazes de auxiliar na tomada de decisão, dentre eles destacam-se os Sistemas de Recomendação (RESNICK et. al., 1994, SHARDANAND; MAES, 1995).

Os Sistemas de Recomendação realizam a coleta de dados de um grupo de usuários e baseado nas informações geradas por esses usuários, objetivam produzir recomendações significativas de itens que sejam de interesse, tais como, livros, músicas, filmes, artigos, entre outros (BOBADILLA et al. 2013, MELVILLE; SINDHWANI, 2010).

Mulholland et al. (2015) defendem que os Sistemas de Recomendação provaram que são capazes de melhorar o processo de tomada de decisão de seus usuários em situações que muitas vezes envolvem grandes quantidades de informação nas mais diversas áreas. Dentre alguns exemplos bem conhecidos, cita-se a área de *e-commerce* com a Amazon® (GHAZANFAR; PRÜGEL-BENNETT, 2014) e área de filmes através da Netflix® (HERNANDO et al., 2013).

Para obter sucesso, os Sistemas de Recomendação trabalham com a avaliação dos usuários sobre determinado item, podendo ser através de diversas formas. A mais comum é por nível de satisfação do usuário a respeito do produto (GENA et al., 2011; POMMERANZ et al. 2012).

Porém, classificar um item apenas como interessante ou desinteressante, ainda é vago. Diariamente milhares de pessoas expressam suas opiniões através de textos em diversos portais e redes sociais (PELEJA et al. 2013). Estas opiniões podem ser utilizadas para aprimorar a recomendação de um produto antes classificados apenas em grau de satisfação. Visando este objetivo, a Análise de Sentimentos surge como uma ferramenta com grande potencial de sucesso. (GARCÍA-CUMBRERAS et al., 2013).

O termo Análise de Sentimentos é encontrado na literatura como um sinônimo de Mineração de Opinião, assim como Pang e Lee (2008), outros autores como Li e Li (2013), Tang et al. (2015), Balazs e Velásquez (2016), Gutiérrez, Vázquez e Montoyo (2016) interpretam que os dois termos tratam de uma mesma área de pesquisa que busca identificar opiniões, sentimentos, avaliações, atitudes e emoções

expressadas através de texto e direcionadas a um produto ou serviço. Para tal, este campo de pesquisa utiliza técnicas de processamento de linguagem natural (LIU, 2012; TANG et al., 2015) e técnicas de aprendizado de máquina (MULHOLLAND et al., 2015).

1.1 PROBLEMÁTICA

O termo *Web 2.0* representa um marco na inovação da internet. Anteriormente limitada a páginas com textos estáticos a internet agora possui um perfil dinâmico contendo aplicações onde seus usuários possuem uma grande contribuição na geração de conteúdo. Inserido neste conteúdo encontra-se a opinião e o compartilhamento de conhecimento destes usuários. Dentre muitas aplicações com estas características, citam-se exemplos como o Twitter®, Facebook® e LinkedIn® (SEO; LEE, 2016).

Esta evolução tem gerado uma sobrecarga de informação, um fenômeno que dificulta o processo de escolha e tomada de decisão por parte dos usuários (CHEN; CHENG; CHUANG, 2008; FERREIRA; OLIVEIRA, 2012). Também proporciona a criação de novos sistemas e novas possibilidades de interação por parte dos usuários. Tais cenários motivaram o surgimento dos Sistemas de Recomendação (CHEN; CHENG; CHUANG, 2008) que objetivam oferecer aos usuários recomendações de itens.

Os Sistemas de Recomendação evoluíram e passaram a utilizar diferentes abordagens para realizar recomendações, visando resultados cada vez mais precisos. Entre as abordagens destacam-se a filtragem colaborativa (recomendação colaborativa), a recomendação baseada em conteúdo e as abordagens híbridas (ADOMAVICIUS; TUZHILIN, 2005).

Posteriormente, outras abordagens foram desenvolvidas por pesquisadores em seus trabalhos, como por exemplo, a recomendação baseada em conhecimento (JANNACH et al. 2011; CARRER-NETO et al. 2012) e a recomendação com filtragem demográfica (MONTANER et al. 2003).

Ainda no contexto da *Web 2.0* comentários efetuados pelos usuários passam a representar informações importantes no auxílio a determina escolha. A partir da necessidade da análise destes comentários surge o campo de pesquisa denominado Análise de Sentimento (GARCÍA-CUMBRERAS et al., 2013).

Autores como Peleja et al. (2013), sugerem que as opiniões dos usuários sejam consideradas ao se realizar recomendações. Já García-

Cumbreras et al., 2013, afirmam que a análise do sentimento expresso em textos demonstra grande potencial para atingir este objetivo.

Embora os Sistemas de Recomendação tenham sido avaliados por muitas áreas de aplicação, ainda existem desafios importantes que precisam ser realizados para que estes atinjam seu pleno potencial (HE; PARRA; VERBERT, 2016).

Dentre os pontos a serem melhorados nos Sistemas de Recomendação, García-Cumbreras et al. (2013) destacam a minimização da partida fria (necessidade de alguma informação para que recomendações possam ser efetuadas), a escalabilidade e a implementação de abordagens que evitem classificações tendenciosas. Sun et al. (2015) complementam com a informação de que os modelos de recomendações tradicionais são extremamente dependentes das classificações dos usuários em forma de *score*, o que na maioria das vezes acarreta no problema da dispersão de dados. Concordando com esta interpretação, Colace et al. (2015) afirmam que um Sistema de Recomendação pode melhorar seu desempenho com o uso da Análise de Sentimento, pois este tipo de análise aplicada em comentários pode melhorar o *ranking* ao recomendar itens.

Outra área afetada pelo surgimento da *Web 2.0* foi a gestão de ideias. Este fato é identificado em grandes empresas que tornaram a decisão de criar comunidades virtuais para colher, analisar e executar ideias que tem grande potencial para se tornar um produto ou serviço. Estas ideias são oferecidas deliberadamente por seus clientes que possuem livre acesso a esta comunidade (LEE et al., 2013).

Bothos et al. (2012), entendem que a gestão de ideias está localizada no *Front End* da inovação e pode ser o núcleo da gestão da inovação. Sendo assim, a Gestão de Ideias tem um papel fundamental no processo de inovação (BARBIERI, ÁLVARES, CAJAZEIRA, 2009), uma vez que as ideias são a matéria-prima para o processo de criação de novos produtos (COOPER; EDGETT; KLEINSCHMIDT, 2002) e a filtragem de ideias é um ponto crítico do processo de inovação, sendo determinante para a falha ou sucesso da organização (ADAMS, BESSANT, PHELPS, 2006).

Para auxiliar o gerenciamento do banco de ideias, foram criados sistemas computacionais para o domínio de Gestão de Ideias denominados Sistemas de Gestão de Ideias. Estes sistemas auxiliam na geração, avaliação e seleção de ideias inovadoras através de processos e critérios diferentes (LI; LI; CHEN, 2014).

No entanto, os Sistemas de Gestão de Ideias enfrentam importantes desafios relacionados com a grande quantidade de esforço

humano, necessitando de um especialista durante o processo de gestão de ideias. As principais origens destes problemas são: a) o excesso de informação através do grande volume de ideias submetidas; b) picos súbitos de submissões de ideias; c) a redundância das ideias, e d) grandes quantidades de ideias triviais (WESTERSKI; DALAMAGAS; IGLESIAS, 2013; MARTINEZ-TORRES; OLMEDILLA, 2016).

Outro ponto importante nos Sistemas de Gestão de ideias aborda o fato de uma ideia ser encontrada diversas vezes na base de dados com classificações diferentes, porém relacionadas. Estas ideias quando analisadas individualmente não despertam interesse, mas quando agrupadas revelam grande potencial (MAGNUSSON; NETZ; WÄSTLUND, 2014).

Sendo assim, seria de grande relevância para as organizações o uso de um Sistema de Recomendação de ideias servindo de suporte no auxílio a tomada de decisão direcionada à inovação (LEE et al., 2013).

Baseado nestas afirmações apresenta-se a pergunta de pesquisa deste trabalho: **Como a Análise de Sentimentos pode contribuir para aprimorar os resultados gerados pelos Sistemas de Recomendação no contexto da Gestão de Ideias?**

1.2 OBJETIVOS

Esta seção apresenta com detalhes os objetivos que direcionam este trabalho.

1.2.1 Objetivo geral

O objetivo geral deste trabalho é desenvolver um modelo baseado em Análise de Sentimentos que promova suporte aos Sistemas de Recomendação no Contexto da Gestão de Ideias.

1.2.2 Objetivos específicos

Para que o objetivo geral seja alcançado, torna-se necessário a realização de objetivos específicos, a saber:

- Definir uma abordagem de Análise de Sentimento que possa ser incorporada aos Sistemas de Recomendação;
- Propor um modelo integrando a Análise de Sentimento e Sistemas de Recomendação;
- Demonstrar a viabilidade do modelo proposto através de um cenário de estudo que possa ser aplicado aos Sistemas de Recomendação no âmbito da Gestão de Ideias;
- Promover uma discussão dos resultados obtidos, bem como apresentar possibilidades de trabalhos futuros.

1.3 JUSTIFICATIVA

A evolução da *Web* para a sua versão 2.0 trouxe consigo um salto tecnológico (KROTZFLEISCH et al., 2008) que proporcionou aos seus usuários a oportunidade de deixarem de serem meros consumidores de informações, para participarem ativamente do processo absorvendo também as funções de produtores de conteúdo (JANNACH et al. 2011). Portais como Twitter®, Facebook® e LinkedIn® são exemplos claros e atuais de produção de conteúdo. (SEO; LEE, 2016).

A explosiva produção de conteúdo gerado todos os dias somado ao surgimento dos *e-business* tem gerado uma sobrecarga de informações que induzem o usuário a tomadas de decisões pouco efetivas. Este cenário motivou a criação de sistemas computacionais capazes de auxiliar os usuários a encontrarem o conteúdo que necessitam, auxiliando-os em suas escolhas e tomadas de decisão, surgindo então os Sistemas de Recomendação (RICCI; ROKACH; SHAPIRA, 2011).

Ao longo do tempo os Sistemas de Recomendação despertaram o interesse de pesquisadores que efetivaram sua aplicação em diversas áreas, como por exemplo, no turismo (HSU; LIN; HO, 2012), no controle de poluentes emitidos por automóveis (GE et al., 2010), no tratamento ao diabetes (CHEN et al., 2012). Porém, o grande destaque destes sistemas é encontrado nos *e-commerces* como a recomendação de produtos da Amazon®. Um exemplo do grande investimento nas pesquisas sobre estes sistemas é dado pela Netflix® que ofereceu um

prêmio de 1 Milhão de dólares para quem aprimorasse seu Sistema de Recomendação de filmes (BELL; KOREN, 2007).

Ainda que os algoritmos de Sistemas de Recomendação tenham sido aperfeiçoados e aprovados em diversos testes, um longo caminho composto de muitos desafios deve ser percorrido para que estes sistemas alcancem seu pleno potencial para auxiliar seus usuários a efetuarem escolhas e tomarem decisões (HE; PARRA; VERBERT, 2016).

Seguindo o aperfeiçoamento dos Sistemas de Recomendação, entende-se que as opiniões dos usuários são uma fonte de informações adicionais que complementam as características de um item (GARCÍA-CUMBRERAS; MONTEJO-RÁEZ; DÍAZ-GALIANO, 2013). Assim, entende-se que as opiniões dos usuários sejam avaliadas ao realizar uma recomendação (PELEJA et al., 2013).

Para realizar a tarefa da extração das opiniões dos usuários, apresenta-se a Análise de Sentimentos. Uma ferramenta derivada da Recuperação de Informação (RI) que trabalha para extrair e analisar as opiniões e sentimentos dos usuários expressadas através de textos (GARCÍA-CUMBRERAS; MONTEJO-RÁEZ; DÍAZ-GALIANO, 2013).

Assim como ocorreu com os Sistemas de Recomendação, a *Web 2.0* também influenciou o campo da gestão de ideias através das comunidades virtuais que se tornaram fontes de ideias criativas com potencial inovador (LEE et al., 2013). Como as ideias atuam como matéria-prima no processo de criação de novos produtos e serviços (COOPER; EDGETT; KLEINSCHMIDT, 2002), a filtragem de ideias tornou-se um ponto crítico do processo de inovação, influenciando diretamente nos resultados da organização (ADAMS, BESSANT, PHELPS, 2006).

Assim, surgiram os Sistemas de Gestão de Ideias que auxiliam na geração, avaliação e seleção de ideias inovadoras (LI; LI; CHEN, 2014). Porém, estes sistemas enfrentam desafios como: a) o excesso de informação através do grande volume de ideias submetidas; b) picos súbitos de submissões de ideias; c) a redundância das ideias, e d) a grandes quantidades de ideias triviais (WESTERSKI; DALAMAGAS; IGLESIAS, 2013; MARTINEZ-TORRES; OLMEDILLA, 2016).

Outro fator a ser considerado é que uma ideia pode ser encontrada diversas vezes na base de dados com classificações diferentes. Quando analisadas individualmente estas ideias não despertam interesse, mas quando agrupadas revelam seu potencial (MAGNUSSON; NETZ; WÄSTLUND, 2014).

Diante deste contexto Lee et al. (2013), sugerem a aplicação de um Sistema de Recomendação de ideias com o objetivo de promover suporte no auxílio a tomada de decisão direcionada à inovação.

Tendo como base o cenário anteriormente ilustrado, a proposta de um modelo que utilize a Análise de Sentimentos como uma ferramenta para analisar a opinião dos usuários, oferece uma contribuição direta no contexto dos Sistemas de Recomendação aplicado no contexto da Gestão de Ideias. A partir destes argumentos entende-se a justificativa e relevância da presente dissertação com contribuições no âmbito organizacional e acadêmico.

1.4 ESCOPO DO TRABALHO

Esta dissertação propõe um modelo baseado em Análise de Sentimentos como suporte aos Sistemas de Recomendação. No cerne deste modelo constam:

- A implementação de um método para a análise de sentimentos (polaridades) expressos em textos que envolvam a opinião de usuários;
- A implementação da abordagem de recomendação baseada em conteúdo textual (neste contexto ideias expressas em linguagem natural) visando sugerir documentos similares a partir de determinado documento de interesse. As recomendações levam em consideração a similaridade entre as ideias e a polaridade dos documentos (positiva ou negativa). Tal polaridade é obtida a partir da utilização do analisador de sentimentos desenvolvido neste trabalho.

Este trabalho não se estende ao desenvolvimento de *software* como produto final, mas como um protótipo visando auxiliar a avaliação dos resultados obtidos. Não pertencem ao escopo a análise de desempenho e a precisão das recomendações das técnicas selecionadas. Ressalta-se ainda, que este trabalho não possui foco na proposição de novas abordagens e técnicas nos domínios de Análise de Sentimento e Sistemas de Recomendação.

1.5 ADERÊNCIA AO PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO

O Programa de Pós-Graduação em Tecnologias da Informação e Comunicação – PPGTIC possui três linhas de pesquisa: Tecnologia, Gestão e Informação, Tecnologia Educacional e Tecnologia Computacional. Esta dissertação enquadrasse na área de Tecnologia Computacional.

De acordo com o PPGTIC (2016), a linha de pesquisa a qual este trabalho pertence estuda o desenvolvimento de modelos, técnicas e ferramentas computacionais que auxiliam a resolução de problemas de natureza interdisciplinares. Em suma, esta área de pesquisa busca o desenvolvimento de novas tecnologias computacionais para serem aplicadas nas áreas de educação e gestão.

Baseado nesta definição, este trabalho atende aos requisitos solicitados pela linha de pesquisa selecionada. Esta afirmação é amparada pelo objetivo deste trabalho que busca promover a análise de sentimento em comentários que possam impactar nos resultados oferecidos pelos Sistemas de Recomendação. Tais análises podem ser aplicadas em diversos cenários, como por exemplo, na Gestão de Ideias. Neste sentido, incorporar a capacidade de analisar o sentimento presente em comentários de usuário e utilizar o resultado das análises no processo de sugestão de algum item, pode auxiliar no suporte aos processos de tomada de decisão nas mais variadas áreas.

1.6 PROCEDIMENTOS METODOLÓGICOS

Nesta seção é descrito a metodologia utilizada nesta pesquisa, bem como, a sua caracterização.

Baseado na obra literária de Gerhardt e Silveira (2009), esta pesquisa é classificada como aplicada. Segundo as autoras a pesquisa aplicada visa gerar o conhecimento para aplicações práticas, dirigidos à solução de problemas específicos envolvendo verdades e interesses locais. É também classificada como pesquisa tecnológica, pois objetiva desenvolver artefatos com base em conhecimentos científicos (CUPANI, 2011; VARGAS, 1985; BUNGE, 1985).

A abordagem aplicada neste trabalho é caracterizada como uma pesquisa exploratória. De acordo com Gil (2002) a pesquisa exploratória tem como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a constituir hipóteses.

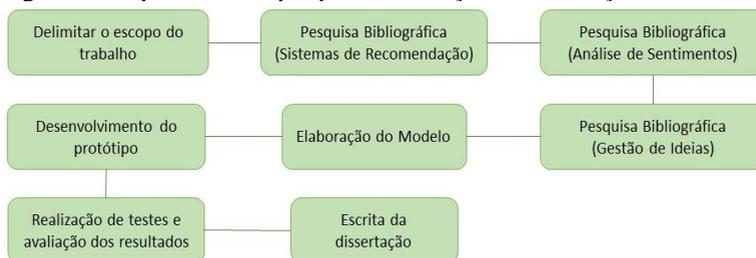
O procedimento técnico utilizado nesta dissertação foi a pesquisa bibliográfica. Para Gil (2002), a pesquisa bibliográfica é realizada baseada em materiais já elaborados, utilizando-se principalmente em livros e artigos científicos.

Visando cumprir os objetivos desta pesquisa, a execução deste trabalho ocorreu através da realização dos seguintes passos:

- Definição do escopo do trabalho;
- Realização de pesquisas bibliográficas relacionadas às áreas de Sistemas de Recomendação, Análise de Sentimentos e Gestão de Ideias. Esta é base para a construção do modelo e orienta os passos subsequentes;
- Proposição de um modelo capaz de atender aos objetivos deste trabalho;
- Desenvolvimento de um protótipo viabilizando a realização dos testes e a avaliação do modelo;
- Análise dos resultados obtidos;
- Apresentação das conclusões e perspectivas de trabalhos futuros.

A Figura 1 apresenta o procedimento utilizado para realização da presente pesquisa.

Figura 1 - Sequência de etapas para a realização da dissertação.



Fonte: Elaborado pelo Autor.

1.7 ESTRUTURA DO TRABALHO

Este trabalho é composto, além do presente capítulo que apresenta a problemática, os objetivos da pesquisa, suas características e aderência ao Programa de Pós-Graduação em Tecnologias da Informação e Comunicação, mais 6 capítulos estruturados como segue;

- Capítulo 2: O segundo capítulo apresenta um referencial teórico sobre Sistemas de Recomendação, sendo este o primeiro assunto relacionado à pesquisa.
- Capítulo 3: Este capítulo aborda o segundo tema fundamental para a pesquisa, a Análise de Sentimentos.
- Capítulo 4: O quarto capítulo apresenta a fundamentação teórica referente à Gestão de Ideias.
- Capítulo 5: Apresenta o modelo proposto por meio de uma descrição detalhada sobre as estruturas que o compõem, dividido em modelo lógico e modelo físico.
- Capítulo 6: Este capítulo apresenta a proposição de avaliação do modelo através da discussão dos resultados alcançados por meio de um cenário de estudo.
- Capítulo 7: O último capítulo apresenta as conclusões da dissertação e as sugestões de trabalhos futuros. Por fim, são disponibilizadas as referências utilizadas nesta dissertação.

2 SISTEMAS DE RECOMENDAÇÃO

A *World Wide Web* popularmente conhecida como *Web*, foi criada por Tim Berners-Lee em 1989 no CERN (*Conseil Européene pour la Recherche Nucléaire*). Esta ferramenta surgiu com o objetivo de solucionar o problema da perda de informações ocasionada pela intensa rotatividade de colaboradores do CERN (BERNERS-LEE, 1989).

Inicialmente na sua versão 1.0, a *Web* servia como uma ferramenta para disponibilizar informações aos seus usuários e era utilizada do mesmo modo das mídias impressas, como jornais e livros. As pessoas que mantinham contato com a *Web* eram separadas por dois grupos. O primeiro grupo era composto pelos produtores de conteúdo. Constituído por profissionais especializados, este grupo era responsável por escrever, editar e publicar o conteúdo dos sites. Deste modo, se consolidava o conceito de propriedade onde os sites eram proprietários e responsáveis pelo conteúdo publicado. Neste cenário pouco interativo, o segundo grupo era composto pelos consumidores de informação, usuários comuns com um comportamento totalmente passivo (SEO; LEE, 2016).

Desde sua criação a *Web* segue um processo evolutivo, e em 2005, durante uma conferência que tratava da participação dos usuários da *Web*, Tim O'Reilly apresentou ao mundo o termo *Web 2.0* que marcaria o início do segundo estágio da *Web* (O'REILLY, 2005). Este termo é até hoje utilizado pela literatura.

Para Krotzfleisch et al. (2008), o principal contraste existente entre as versões da *Web* ocorre no foco encontrado em cada uma delas. Em sua versão 1.0 o foco da *Web* era mais voltado aos protocolos e estruturas das páginas mantendo a *Web* estática. Já em sua versão 2.0, o foco passou para os usuários e conteúdos tornando a *Web* dinâmica, levando-a a um salto tecnológico utilizando novas configurações de tecnologias voltadas à internet.

Monteiro e Fidencio (2013) destacam que o crescimento da *Web* tem se ampliado levando ao surgimento de novos conceitos, como por exemplo: *Web 2.0* ou *Web Social*, *Web invisível*, *Web visível*, *Web Semântica*, *Web Pragmática*.

Com a *Web 2.0*, caíram as barreiras que separavam os produtores e consumidores de conteúdo, assim os usuários comuns passaram a assumir atividades de produção de conteúdo, antes exclusivas à profissionais especializados. Sendo assim, diariamente os usuários da internet publicam e compartilham conteúdo nos mais diversos portais de

informação com extrema naturalidade (JANNACH et al, 2011, KROTZFLEISCH et al, 2008).

Este usuário híbrido que consome e produz conteúdos para a Web é identificado na literatura pelo termo “*prosumer*”, criado por Alvin Toffler em 1980. Prosumer é uma expressão originada do inglês, produto de uma junção entre as palavras *producer* (produtor) e *consumer* (consumidor) (TOFFLER, 1980).

Eccleston e Griseri (2008) destacam a Web 2.0 como uma tecnologia inovadora permitindo a comunicação interativa e personalizada entre as pessoas. Ainda neste sentido, Kim et al. (2009), definem a Web 2.0 como uma ferramenta que visa proporcionar aos seus usuários a capacidade de publicar suas opiniões em sites, e também interagir e colaborar com outras pessoas em tópicos que sejam de seu interesse. Krotzfleisch et al. (2008) entendem que este seja o principal motivo da popularização por todo planeta de portais de informação como MySpace®, Wikipedia®, Youtube®, Facebook®.

A Web 2.0 demonstrou um forte potencial para melhorar a relação entre organizações e consumidores, em que este fato despertou a atenção de muitas empresas devido a sua capacidade de (ANDRIOLE, 2010). Sendo assim, a Web 2.0 também foi naturalmente incorporada ao cotidiano de inúmeras empresas independente do ramo de atuação (VOIGT; ERNST, 2010). Atualmente, a maioria das principais organizações possuem contas ativas em redes sociais e interagem com seus consumidores através de várias aplicações Web 2.0, tais como Facebook®, Twitter® e YouTube® (SEO; LEE, 2016).

O enorme crescimento dos produtores de informações gerou um aumento exponencial na produção de dados. Em 2003, a pesquisa desenvolvida por Lyman e Varian (2003) apontou uma produção anual de novas informações mensurada em 2 *exabytes*.

Hilbert e López (2011) apontam que a capacidade tecnológica de comunicação mundial chegou a marca de 1.15 *zettabytes* em 2007 e a capacidade de armazenamento foi mensurada em 309 *exabytes*.

Dados mensurados pela IBM (2014) revelam a marca de 2,5 quintilhões de bytes de dados criados diariamente, e 90 por cento dos dados no mundo até então foram produzidos nos últimos dois anos que antecedem a realização da pesquisa.

No início do segundo semestre de 2014, o ACI *Information Group* revelou uma pesquisa informando que a cada minuto o Twitter® recebia 300.000 novas postagens, usuários do Facebook® compartilharam 2.5 milhões de publicações, o YouTube® recebeu o

upload de 72 horas de novos vídeos e a Amazon® realizou mais de \$80.000 em vendas online (ACI, 2014).

Pesquisas atuais demonstram a constância no crescimento exponencial da capacidade de geração de conteúdo dos usuários, atingindo uma produção diária de dados em 2,5 *exabytes* (1 EB = 1.000.000 *Terabytes*) (BELLO-ORGAZ; JUNG; CAMACHO, 2016).

Outra pesquisa recente realizada por Stephen (2016) aponta que a taxa atual de uso da internet entre os adultos americanos se aproxima a 87%. No Reino Unido, durante a última década, usuários adultos passaram 20,5 horas on-line semanais. O número de usuários das redes atingiu a marca de 2 bilhões de usuários e metade destes estão diariamente ativos no Facebook®.

Pesquisas voltadas ao Brasil apontam que 83,4 milhões de brasileiros possuíam acesso à internet até no segundo trimestre de 2012 (IBOPE,2012). Já em Janeiro de 2013 o tempo utilizado pelos internautas brasileiros conectados a *Web* chegou a média de 10 horas e 26 minutos (IBOPE,2013).

Este cenário contendo cada vez mais informações exigiu que os usuários da internet realizassem um grande esforço para encontrar as informações que desejavam. Sendo a sobrecarga de informação um fenômeno que dificulta o usuário no processo de escolha e tomada de decisão, fez-se necessário a inclusão de motores de busca para auxiliar os usuários. Ainda assim, as informações encontradas não eram satisfatórias às necessidades dos usuários que necessitavam de uma filtragem de informações mais aprimorada (FERREIRA; OLIVEIRA, 2012). Esta dificuldade demonstrada pelos motores de busca de retornar a informação necessária era ocasionada pela estrutura das informações, pois, estas foram construídas para serem lidas apenas por seres humanos, sendo assim, os sistemas computacionais não conseguiam processá-las, nem interpretar os dados contidos nelas (CARRER-NETO et al, 2012).

Além da exponencial crescente na produção de dados e suas variedades, outro fator sobrecarrega e induz usuários a decisões ineficazes foi a introdução dos portais denominados e-business que oferecem serviços como compra de produtos, comparação de produtos, leilões, etc. Tamanha variedade de opções não produz benefícios e diminui o bem-estar dos usuários. Assim, destacou-se a necessidade de sistemas computacionais capacitados a auxiliar as pessoas em realizar suas tomadas de decisões com maior eficiência. Este cenário foi a principal fonte de inspiração para o início de pesquisas e o surgimento dos Sistemas de Recomendação. (RESNICK et al., 1994,

SHARDANAND; MAES, 1995, CHO; KIM; KIM, 2002, MIN; HAN, 2005, CHEN; CHENG; CHUANG, 2008, RICCI; ROKACH; SHAPIRA, 2011).

Os Sistemas de Recomendação podem ser definidos como ferramentas com o objetivo de realizar uma coleta de dados de um grupo de usuários e utilizar estes dados na realização de sugestões. Sendo assim, estes sistemas têm como foco gerar recomendações significativas de itens de maneira genérica (livros, músicas, filmes, artigos, entre outros), que estejam de acordo com o interesse de um grupo de usuários (BOBADILLA et al., 2013, MELVILLE; SINDHWANI, 2010).

Mahmood e Ricci (2009) classificam os Sistemas de Recomendação como aplicações inteligentes que ajudam os usuários a encontrar as informações desejadas. Estes sistemas sugerem os itens que melhor se encaixam as necessidades e preferências dos usuários. As informações utilizadas para traçar o perfil do usuário são coletadas através das interações do usuário com o sistema. Durante este processo o usuário informa suas preferências e corresponde através feedbacks. Uma vez estipulado o perfil do usuário, itens correspondentes a este perfil serão utilizados em futuras recomendações.

Os usuários de Sistemas de Recomendação beneficiam-se pelo ganho de tempo e pela facilidade do uso, pois estes sistemas proporcionam um baixo nível de esforço e não é necessário possuir grande experiência. Atrelado a estes benefícios encontra-se o retorno eficiente para cada busca destinada a encontrar itens de seus interesses (FERREIRA; OLIVEIRA, 2012).

Os primeiros Sistemas de Recomendação surgiram em meados da década de 1990, porém, eram na época denominados Sistemas de Filtragem Colaborativa. O primeiro Sistema de Recomendação chamado *Tapestry* foi desenvolvido em 1992. Diante do desafio de encontrar o conteúdo desejado perante um grande volume de dados que aumentava exponencialmente com o uso dos e-mails, o *Tapestry* tinha como objetivo aumentar a eficácia de filtragem utilizando a participação humana neste processo (GOLDBERG et al., 1992). Em 1997 um trabalho similar ao *Tapestry* foi desenvolvido. Este novo sistema nomeado *GroupLens* classificava e compartilhava notícias de acordo com o resultado da avaliação realizada por muitos outros usuários (RESNICK; VARIAN, 1997).

2.1 ABORDAGENS

Os Sistemas de Recomendação podem ser classificados em duas categorias: sistema de recomendação pessoal e sistema de recomendação de grupo. O primeiro visa a eficácia na filtragem de informações úteis que se adequam as necessidades particulares de cada usuário. A segunda classificação se enquadra aos sistemas que visam fornecer sugestões eficazes para permitir a decisão em grupo e satisfazer as necessidades dos usuários que compõem este grupo (CHEN; CHENG; CHUANG, 2008).

Adomavicius e Tuzhilin (2005) classificam os Sistemas de Recomendação em três tipos, considerados tradicionais pela literatura, a saber: Sistemas de Recomendação com Filtragem Colaborativa, Sistemas de Recomendação com Filtragem Baseada em Conteúdo e uma terceira abordagem resultante pela união dos anteriores, denominada como Sistemas de Recomendação Híbridos.

A literatura ainda apresenta outras abordagens para Sistemas de Recomendação mencionadas pelos autores em suas obras. Cita-se como exemplo, Jannach et al. (2011) e Carrer-Neto et al. (2012) mencionam em seus trabalhos a Filtragem Baseada em Conhecimento. Nesta abordagem, após traçar o perfil do usuário, algoritmos de inferência são aplicados para identificar a correlação existente entre as suas preferências e produtos existentes, serviços ou conteúdos. Uma outra abordagem para Sistemas de Recomendação denominada como Filtragem Demográfica é citada por Montaner et al. (2003). Neste caso as descrições dos usuários são utilizadas para estabelecer uma relação para descobrir qual item é mais adequado para este determinado perfil. Estereótipos preestabelecidos são utilizados para classificar os perfis dos usuários. O próprio usuário fornece as informações que constituem seu perfil através do preenchimento de um formulário de cadastro.

Neste trabalho serão discutidas as seguintes abordagens de Sistemas de Recomendação consideradas tradicionais: Filtragem Baseada em Conteúdo, Filtragem Colaborativa e Filtragem Híbrida.

2.1.1 Filtragem Baseada em Conteúdo

A Filtragem Baseada em Conteúdo originou-se na Recuperação de Informação (RI) e nos filtros de pesquisas de informação. Esta abordagem executa recomendações de itens conforme a similaridade entre eles. A recomendação de um item é baseada em uma classificação realizada inicialmente por outros usuários com perfis semelhantes. No

caso de um sistema de recomendação de filmes, ao realizar a indicação de um filme para o usuário 1, a filtragem baseada em conteúdo busca encontrar as características semelhantes entre o filme a ser recomendado e outros filmes que o usuário 1 já avaliou. Cita-se como exemplos de características de filmes os atores, diretores, gênero, classificação indicativa dentre outras. Assim, os filmes com alto grau de similaridade e que se enquadram nas preferências do usuário 1 serão recomendados a ele (ADOMAVICIUS; TUZHILIN, 2005; CHEN; CHENG; CHUANG, 2008; LÜ et al., 2012).

De acordo com o trabalho realizado por Melville e Sindhvani (2010), muitas pesquisas relacionadas a esta abordagem visam a recomendação de itens com conteúdo textual agregado (páginas Web, livros, filmes, etc.). Muitos estudos tratam este caso como Recuperação de Informação (RI), visto que as interações entre o usuário e o item acabam por configurar o perfil deste item. O perfil é representado por um vetor de informações que é composto pelos atributos mais relevantes de um item. Os vetores de informação são utilizados para determinar as semelhanças entre os itens a fim de recomendar itens que possuam vetores com informações similares.

2.1.1.1 Modelo Vetorial

Na Recuperação de Informação (RI), o Modelo de Espaço Vetorial (*Vector Space Model* - VSM) é amplamente aplicado (MANNING; SCHÜTZE, 1999). O VSM é considerado um modelo de alta aplicabilidade e simples, onde VSM trata a proximidade semântica como proximidade espacial, o que justifica o amplo uso do VSM em RI. Na aplicação do VSM em RI, o documento é transformado em um vetor de espaço n -dimensional, sendo n o número representativo para diversos termos (RUSSEL; NORVIG, 1995). Os vetores em conjunto formam a matriz documento-termo, onde a matriz pode ser armazenada com uma estrutura de índice invertido (GONÇALVES, 2006).

Para cada elemento que compõe o vetor é atribuído um identificador e um peso que representa a importância deste elemento para o conteúdo de um documento. Este peso é definido por um método estatístico numérico chamado *TF-IDF* (*Term Frequency / Inverted Document Frequency*) (TRSTENJAK; MIKAC; DONKO, 2014). O TF-IDF é frequentemente utilizado no processamento de linguagem natural (NLP) e na recuperação de informação e mineração de texto (MASUDAA; MATSUZAKIB; TSUJIC, 2011, FRIEDMAN; RINDFLESC; CORN, 2013).

Na visão de Trstenjak; Mikac; Donko (2014), o *TF-IDF* é utilizado para aferir a frequência relativa dos termos de um determinado documento através da proporção inversa deste termo ao longo de todo conteúdo contido neste documento. Para encontrar o peso do termo o método baseia-se em dois fundamentos:

- *TF*: Representa a frequência que o termo é exibido em um documento;
- *IDF*: Representa a frequência inversa de documentos que contêm o termo;

O peso de cada termo é incrementado à medida que este termo é repetido no conteúdo do documento (TRSTENJAK; MIKAC; DONKO, 2014).

Transportando o *TF-IDF* para o universo dos Sistemas de Recomendação, identifica-se o documento como o item a ser o elemento da análise, sendo assim, os termos deste documento serão transformados nos atributos que constituem o vetor (REINEHR, 2013).

De posse do vetor base, recuperam-se outros vetores mediante suas similaridades. A literatura apresenta algumas medidas de similaridade para determinar a distância entre os vetores como, o produto interno (SALTON; BUCKLEY, 1988), o cosseno (JONES; FURNAS, 1987) e a distância Euclidiana (LEE et al., 2014). Nouali e Blache (2004) entendem que o modelo vetorial possibilita que os documentos recuperados sejam facilmente classificados e avaliados conforme sua importância, tornando-o um modelo flexível.

2.1.1.2 Similaridades entre Vetores

Jones e Furnas (1987) afirmam que para encontrar a similaridade entre vetores, é necessário aplicar uma medida que indique a proximidade entre estes vetores compreendidos em um universo Ω , sendo Ω conhecido como o grupo de documentos. Deste modo, um universo contendo um grupo de objetos que podem ser representados através de um vetor, é ilustrado por Ω .

Para calcular a distância entre vetores Egghe e Michel (2002) citam as seguintes equações para encontrar esta informação: o índice Jaccard, a medida *overlap* (máxima e mínima), a medida do cosseno e a medida do pseudo-cosseno.

A similaridade existente entre os vetores é obtida ao calcular o ângulo do cosseno que é formado pelos vetores que representam os documentos contendo os termos e suas frequências (ADOMAVICIUS; TUZHILIN, 2005). De acordo com Gonçalves (2006), a equação que mede o ângulo entre dois vetores pode apresentar o resultado com uma variação entre 1.0 e -1.0. A saber:

- 1.0 ($\cos(0^\circ) = 1.0$) onde os vetores apontam na mesma direção.
- 0.0 ($\cos(90^\circ) = 0.0$) quando os vetores formam um ângulo reto.
- -1.0 ($\cos(180^\circ) = -1.0$) neste caso, os vetores apontam em direções opostas.

Esta equação pode ser representada da seguinte maneira:

$$\cos \theta = \frac{\sum_{i=1}^n (t_i \times q_i)}{\sqrt{\sum_{k=1}^n (t_k)^2} \times \sqrt{\sum_{j=1}^n (q_j)^2}}$$

Onde t_i e t_k representam o peso contido nas posições as i th e k th do vetor t , do mesmo modo q_i e q_j representam os pesos encontrados nas posições i th e j th do vetor q . Ao aplicar esta fórmula será obtido um resultado entre 0 e 1 que determinará o grau de similaridade dos vetores.

2.1.1.3 Limitações

A Filtragem Baseada em Conteúdo possui algumas limitações que são destacadas por diversos autores. A seguir, lista-se as três principais limitações encontradas:

- **Análise limitada pelo conteúdo:** A aplicação da filtragem baseada em conteúdo torna-se muito complexa quando executada sobre conteúdos multimídia como imagens e vídeos, pois é difícil analisar o conteúdo de dados pouco estruturados (CAZELLA et al., 2010, BOBADILLA et al., 2013, RICCI et al., 2011). Além disso, a Filtragem Baseada em Conteúdo não consegue identificar a qualidade do texto, logo, não pode destacar textos bem escritos de textos mal escritos (JANNACH et al., 2011).
- **Superespecialização:** A Filtragem Baseada em Conteúdo não pode recomendar itens inéditos aos usuários. Em algumas situações, os itens não serão recomendados caso sua

similaridade com algum outro item que o usuário já apreciou seja alta. Como por exemplo, uma notícia diferente descrevendo o mesmo fato (ADOMAVICIUS E TUZHILIN, 2005, CARRER-NETO et al., 2012).

- Problema com novos usuários: Um novo usuário deve avaliar uma quantidade considerável de itens para que um sistema de Filtragem Baseada em Conteúdo tenha subsídios suficientes para realizar recomendações confiáveis. Sendo assim, um novo usuário com poucas avaliações realizadas, corre um alto risco de não receber recomendações corretas.

2.1.2 Filtragem Colaborativa

As primeiras pesquisas relacionadas à Filtragem Colaborativa (FC) foram publicadas no início da década de 90. Goldberg et al. (1992), criaram um sistema denominado *Tapestry*. Este sistema utilizava filtragem colaborativa com o objetivo de filtrar *e-mails* a partir de várias listas de discussão, tendo como base a opinião dos usuários sobre suas leituras. Logo em seguida, Resnick et al. (1994), apresentaram outro sistema considerado pioneiro na Filtragem Colaborativa. Denominado *GroupLens*, o sistema oferecia aos usuários a opção de avaliar artigos lidos dentro de uma escala que variava entre 1 e 5 pontos e realizar sugestões (TAKÁCS; PILÁSZ; NÉMETH; TIKK, 2009).

Grande parte das abordagens de Filtragem Colaborativa visa encontrar usuários com perfis semelhantes, com o objetivo de compartilhar recomendações entre eles (CARRER-NETO et al., 2012). Assim, elementos bem qualificados por um grupo de usuários, serão sugeridos para outro usuário que possui preferências semelhantes ao grupo (CHEN; CHENG; CHUANG, 2008).

Neste mesmo sentido, Daddio e Manzato (2014) destacam que na abordagem de Filtragem Colaborativa (FC), a recomendação é feita por item ou associação de usuários contando com classificações fornecidas por estes usuários. Wang, Liu e Yu (2012) definem Filtragem Colaborativa como uma técnica amplamente utilizada em sistemas de recomendação, que recomendam itens para um determinado usuário baseado em avaliações de outros usuários. Estas classificações, muitas vezes sob a forma de um escalar, por exemplo, de 1 a 10 estrelas, representam opiniões gerais das pessoas sobre itens. Ainda em concordância com os autores anteriores Fukumoto, Motegi e Matsuyoshi (2012), analisam a Filtragem Colaborativa como um processo de filtragem de informações ou padrões usando técnicas que envolvem a

colaboração entre os vários agentes, pontos de vistas, fontes de dados e assim por diante.

A Filtragem Colaborativa é o método amplamente aplicado para uso de recomendação pessoal. Um exemplo de sua aplicação é encontrado no portal de vendas Amazon®, conhecido mundialmente (LINDEN; SMITH; YORK, 2003).

Alguns autores classificam a Filtragem Colaborativa em dois métodos: elicitación explícita e implícita, onde, na elicitación explícita, os usuários avaliam os itens através de uma escala, por exemplo, atribuindo de uma a cinco estrelas. Já na elicitación implícita, as preferências são derivadas de uma análise do comportamento de navegação e seleção de usuários (GENA et al., 2011; POMMERANZ et al., 2012; KNIJNENBURG et al., 2012). De maneira semelhante, porém utilizando outras denominações, Yang et al. (2014) entendem as elicitaciones como *feedbacks* de usuários, polarizados em *feedback* explícito e implícito. No primeiro termo, o usuário atribui uma classificação a um item. O segundo caso é caracterizado quando o usuário interage com um item, por exemplo, clicar em um *link*, ouvir uma música, ou comprar um item. Baseados em pesquisas, Koren, Bell e Volinsky (2009) alertam para o acréscimo obtido na precisão da recomendação ao unir os *feedbacks* explícitos e implícitos.

Neste trabalho serão exploradas duas abordagens de recomendações aplicadas na Filtragem Colaborativa: Recomendações da vizinhança baseada em usuários e recomendações da vizinhança baseada em itens. Estas abordagens e seus exemplos, foram baseadas na obra literária de Jannach et al. (2011).

2.1.2.1 Recomendações de vizinhos próximos baseado em usuários

A aplicação desta abordagem necessita a base de dados que contém as avaliações. Os pares que definem os vizinhos mais próximos são formados utilizando as preferências semelhantes entre usuário ativo e os vizinhos existentes. Para os itens que o usuário ativo ainda não conhece aplica-se uma previsão calculada com base na avaliação dos vizinhos mais próximos.

A Tabela 1 ilustra as avaliações promovidas pelos usuários aos itens. O peso das avaliações varia entre 1 e 5, sendo 1 para itens que o usuário não gostou e 5 para itens que ele gostou muito. Neste caso, o usuário ativo desconhece o item 5, logo, o sistema de recomendação

deve prever se o usuário ativo gostará ou não deste item baseando-se na avaliação dos usuários semelhantes.

Tabela 1 - Base de dados de classificações.

	Item 1	Item 2	Item 3	Item 4	Item 5
Ativo	5	3	4	4	?
Usuário 1	3	1	2	3	3
Usuário 2	4	3	4	3	5
Usuário 3	3	3	1	5	4
Usuário 4	1	5	5	2	1
Usuário 5	1	2	3	4	5

Fonte: Adaptado de JANNACH et al., (2011).

A seguir abordam-se os cálculos necessários para realizar a recomendação. A identificação dos usuários é dada pelo conjunto $U = \{u_1, \dots, u_n\}$, para identificar os produtos o conjunto $P = \{p_1, \dots, p_m\}$ e R sendo uma matriz $n \times m$ de classificações $r_{i,j}$. As posições da matriz $r_{i,j}$ estarão vagas caso não exista avaliações do usuário i para o item j .

A semelhança entre usuários é determinada através do coeficiente de Pearson, onde a similaridade entre os usuários a e b é representada por $sim(a, b)$. Uma vez ciente que alguns usuários possuem a tendência de aplicar avaliações altas, enquanto outros são pré-dispostos a aplicar baixas classificações, o coeficiente de Pearson utiliza a média das classificações para promover um nivelamento entre os usuários. A classificação média do usuário é representada pelo símbolo \bar{r}_a .

O preenchimento da matriz de classificação R é realizado através da fórmula abaixo:

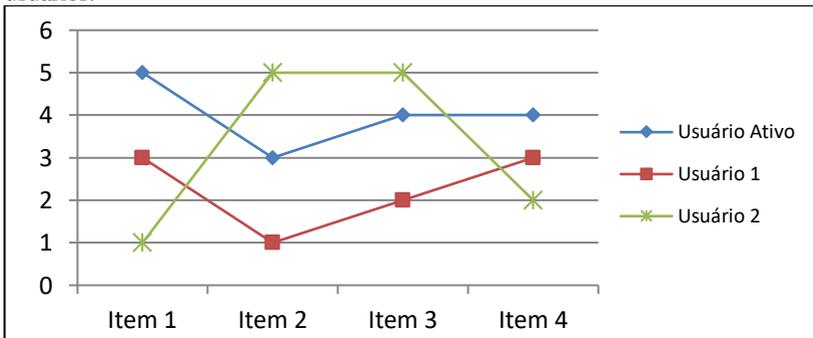
$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

A semelhança entre o usuário ativo e o usuário 1 é, portanto, como se segue $(\bar{r}_{Ativo}) = (\bar{r}_a) = 4$, $(\bar{r}_{Usuário1}) = (\bar{r}_b) = 2.4$:

$$\frac{(5 - \bar{r}_a) * (3 - \bar{r}_b) + (3 - \bar{r}_a) * (1 - \bar{r}_b) + \dots + (3 - \bar{r}_a) * (4 - \bar{r}_b)}{\sqrt{(5 - \bar{r}_a)^2 + (3 - \bar{r}_a)^2 + \dots} \sqrt{(3 - \bar{r}_b)^2 + (1 - \bar{r}_b)^2 + \dots}} = 0.85$$

O coeficiente de correlação de Pearson apresenta valores compreendidos entre +1 (forte correlação positiva) e -1 (forte correlação negativa). Aplicando o coeficiente de Pearson, percebe-se que existe uma semelhança entre as classificações do usuário ativo e do usuário 1. Analisando a Figura 2 percebe-se uma correlação linear clara entre as classificações.

Figura 2 - Comparação entre as classificações do usuário ativo e outros usuários.



Fonte: Adaptado de JANNACH et al., (2011).

Para realizar a previsão de classificação do usuário ativo para o Item 5, é necessário definir qual das classificações vizinhas serão consideradas. Uma possível maneira para calcular a previsão de classificação de um usuário qualquer (a) para um determinado item (p) é considerando a média de classificação \bar{r}_a em que os vizinhos mais próximos (N) atribuíram a este item (p), como na fórmula a seguir:

$$prev(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

Neste exemplo a previsão da classificação que o usuário ativo fará para o **item 5**, baseada nas classificações de seus vizinhos próximos Usuário1 e Usuário 2, será a seguinte:

$$4 + 1 / (0.85 + 0.7) * (0.85 * (3 - 2.4) + 0.70 * (5 - 3.8)) = 4.87$$

Através destes cálculos, podem-se aplicar previsões de classificações para itens desconhecidos por um usuário. É necessário

considerar que as bases de dados de classificações reais podem conter milhões de usuários e itens, sendo necessário um expressivo poder computacional para atender a esta demanda.

2.1.2.2 Recomendações de vizinhos próximos baseado em itens

A Filtragem Colaborativa já provou sua qualidade em vários domínios de aplicações. Mas sua funcionalidade fica extremamente comprometida quando se trata de uma aplicação com um grande volume de dados como em um *e-commerce* onde existem milhares de itens e usuários. Neste cenário, é praticamente impossível calcular recomendações em tempo real, pois o número de usuários pode crescer em larga escala. Criou-se então a recomendação de vizinhos mais próximos com base em itens, pois o número de itens não acompanha este ritmo de expansão. Neste caso, a ideia principal é recomendar itens conforme a similaridade entre eles.

Observando a Tabela 1, percebe-se que as avaliações dadas ao item 5 e ao item 1 são semelhantes. As recomendações baseadas em itens simplificam a visão das classificações do usuário ativo para os itens similares. Para gerar uma previsão de classificação para o item 5, será aplicado a média ponderada com base nas classificações anteriores do usuário ativo. Visando encontrar itens semelhantes é necessário definir uma medida de similaridade. Para abordagens de recomendação baseadas em itens é comum estabelecer a similaridade do cosseno como métrica padrão, uma vez que foi demonstrado que produz os resultados mais precisos.

A similaridade entre dois itens a e b pode ser encontrada através da fórmula abaixo, onde os itens são representados pelos vetores de classificação correspondentes \vec{a} e \vec{b} , assim formalmente definidos:

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}$$

A seguir, encontra-se a aplicação da fórmula aos dados propostos para o item 1 e o item 5 na Tabela 1 com o intuito de encontrar a similaridade através do cosseno.

$$sim(I5, I1) = \frac{3 * 3 + 5 * 4 + 4 * 3 + 1 * 1}{\sqrt{3^2 + 5^2 + 4^2 + 1^2} * \sqrt{3^2 + 4^2 + 3^2 + 1^2}} = 0.99$$

Na próxima etapa, aplicar-se a medida cosseno ajustada, assim como na medida Pearson abordada na seção anterior, onde U torna-se o conjunto de usuários que avaliaram os itens a e b . A medida do cosseno ajustado é calculada pela formula abaixo:

$$sim(a, b) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$$

Portanto, pode-se transformar a base de dados de classificações originais e substituir os valores de classificação originais com os desvios em relação às avaliações médias, como mostrado na Tabela 2. Por exemplo, para o Item 1 do usuário Ativo o valor é obtido subtraindo o valor da avaliação, neste caso 5, da média de avaliações do usuário que possui o valor 4. Isto se repete para as demais avaliações da tabela.

Tabela 2 - Banco de dados de avaliações médias ajustadas.

	Item 1	Item 2	Item 3	Item 4	Item 5
Ativo	1.00	-1.00	0.00	0.00	?
Usuário 1	0.60	-1.40	-0.40	0.60	0.60
Usuário 2	0.20	-0.80	0.20	-0.80	1.20
Usuário 3	-0.20	-0.20	-2.20	2.80	0.80
Usuário 4	-1.80	2.20	2.20	-0.80	-1.80

Fonte: Adaptado de JANNACH et al., (2011).

O valor ajustado de similaridade do cosseno para os itens 5 e 1 para o exemplo é ilustrado abaixo:

$$\frac{0.6 * 0.6 + 0.2 * 1.2 + (-0.2) * 0.80 + (-1.8) * (-1.8)}{\sqrt{(0.6)^2 + 0.2^2 + (-0.2)^2 + (-1.8)^2} * \sqrt{0.6^2 + 1.2^2 + 0.8^2 + (-1.8)^2}} = 0.80$$

Uma vez encontradas as semelhanças entre os itens, é possível prever a classificação do usuário ativo para Item 5 através da soma ponderada das classificações do usuário ativo para os itens semelhantes ao Item 5. A previsão da classificação de um usuário U para um produto P é representada formalmente da seguinte maneira:

$$perd(u, p) = \frac{\sum_{i \in ClassificacaoItem(u)} sim(i, p) * r_{u,i}}{\sum_{i \in ClassificacaoItem(u)} sim(i, p)}$$

Destaca-se que assim como na abordagem com base no usuário, o tamanho da vizinhança é normalmente considerado limitado a um tamanho específico, logo, alguns vizinhos serão desconsiderados para a previsão.

2.1.2.3 Limitações

Assim como ocorre na Filtragem Baseada em Conteúdo, a Filtragem Colaborativa também possui suas limitações (WU; CHANG; LIU, 2014, CAZELLA et al., 2010, ADOMAVICIUS; TUZHILIN, 2005), entre elas:

- Problema do primeiro avaliador: Não é possível recomendar um item cadastrado recentemente até que ele receba avaliações de outros usuários.
- Problema de avaliações esparsas: Quando a quantidade de usuários for pequena comparada ao volume de informações, há uma probabilidade considerável de que as avaliações se tornarem muito esparsas.
- Similaridade: Quando um usuário possuir uma ampla variedade de preferências ao ponto de se tornar um padrão único, será difícil encontrar perfis semelhantes para basear as recomendações, aumentando o risco de realizar recomendações fracas.

2.1.3 Filtragem Híbrida

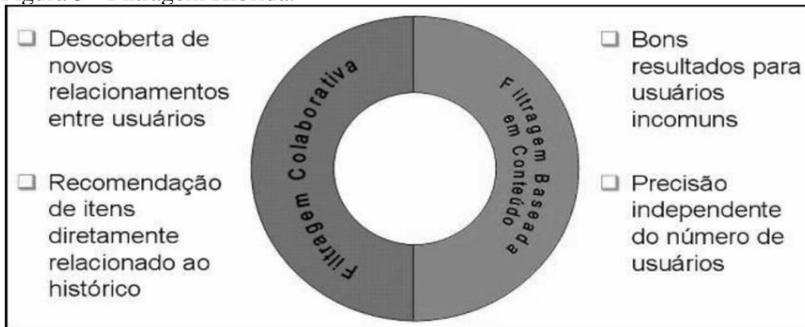
Diversos Sistemas de Recomendação utilizam uma abordagem híbrida, onde as abordagens de filtragem colaborativa e baseada em conteúdo são aplicadas complementando-se uma a outra. Esta técnica visa prevenir as limitações existentes em cada abordagem. Conforme Adomavicius e Tuzhilin (2005), é possível combinar métodos colaborativos e baseados em conteúdo em um sistema de recomendação híbrido das seguintes maneiras:

- Implementação de métodos colaborativos e baseados em conteúdo separadamente e combinar as suas previsões;
- Incorporando algumas características baseadas em conteúdo em uma abordagem colaborativa;

- Incorporando algumas características colaborativas em uma abordagem baseada em conteúdo;
- Construção de um modelo de unificação geral que incorpora características das filtragens baseadas em conteúdo e filtragem colaborativa.

De um modo mais simples, Cazella et al. (2011) apresentam a abordagem de Filtragem Híbrida. Os autores focam na combinação dos pontos fortes existentes nas abordagens baseada em conteúdo e filtragem colaborativa, visando criar um sistema que atenda melhor as necessidades de seus usuários. Assim, neutralizando os pontos fracos de cada abordagem. A concepção dos autores sobre a abordagem híbrida é ilustrada na Figura 3.

Figura 3 - Filtragem Híbrida.



Fonte: (CAZELLA et al., 2010).

2.2 APLICAÇÕES

Outras abordagens são mencionadas por autores em suas obras, por exemplo, Jannach et al. (2011) e Carrer-Neto et al. (2012) citam em seus trabalhos a Filtragem Baseada em Conhecimento e Montaner et al. (2003), mencionam uma outra abordagem denominada como Filtragem Demográfica. Ao longo do tempo esta área de pesquisa teve sua aplicação muito diversificada, sendo encontrada em aplicações com foco na avaliação de satisfação de clientes (JIANG; SHANG; LIU, 2010), na sugestão de filmes através do MovieLens® (MILLER et al., 2003) e mais recentemente com a empresa Netflix® (HERNANDO et al., 2013).

O setor de turismo também vem recebendo a aplicação dos Sistemas de Recomendação (HSU; LIN; HO, 2012; GAVALAS et al., 2014, GARCÍA-CUMBRERAS et al., 2013). No trabalho de Borràs, Moreno e Valls (2014), um sistema de recomendação de viagem orienta os turistas, combinando atrações turísticas e de lazer conforme os interesses dos viajantes. Já Colomo-palacios et al. (2017) desenvolveram uma plataforma que utiliza opiniões de usuários postadas nas redes sociais para realizar as recomendações turísticas.

Na gestão ambiental, os Sistemas de Recomendação são aplicados visando a redução da emissão de poluentes. Isto é possível aprimorando o gerenciamento da frota de táxis com o uso da recomendação de rotas mais produtivas abreviando o tempo utilizado em cada atendimento (GE et al., 2010).

Chen *et al.* (2012) mencionam em seu trabalho a aplicação dos Sistemas de Recomendação na área da saúde auxiliando pacientes que passam por tratamento de doenças como o diabetes. Neste caso, este sistema é capaz de analisar os sintomas da diabetes, bem como recomendar os medicamentos mais adequados a partir de drogas relacionadas.

Os Sistemas de Recomendação também têm contribuído com a transferência de conhecimento. Empresas que buscam capacitar seus colaboradores têm investido em sistemas especialistas que funcionam como mentores informais. Esta ferramenta oferece recomendações de conteúdos com base nas competências pessoais do aprendiz e seus interesses (VALENCIA-GARCÍA et al., 2012, Colomo-palacios et al., 2012).

Casas inteligentes também têm utilizado Sistemas de Recomendação e após passar por um período de aprendizagem sobre o comportamento do usuário, o sistema recomenda ações que se encaixam aos hábitos do morador. Como por exemplo, ligar o sistema de som com determinada música enquanto o usuário prepara o jantar (KATHARINA, 2014).

O setor financeiro também pode ser mencionado com o uso de ferramentas que realizam recomendações de aplicações e tendências de mercado, visando auxiliar analistas financeiros sobre investimentos das organizações (RODRÍGUEZ-GONZÁLEZ et al., 2011, GONZALEZ-CARRASCO et al., 2012).

Contudo, seu maior foco concentra-se ainda no *e-commerce*, tendo como um grande exemplo de sucesso a sugestão de produtos da Amazon® (GHAZANFAR; PRÜGEL-BENNETT, 2014).

3 ANÁLISE DE SENTIMENTOS

Atualmente na *Web 2.0*, frequentemente encontram-se sites que permitem aos usuários avaliarem um item ou serviço atribuindo uma nota em um placar e/ou através de um comentário. O uso da avaliação por placar proporciona a aplicação de soluções voltadas à Sistemas de Recomendação, porém, esta forma de avaliação nem sempre é o melhor caminho para mensurar a afinidade entre um usuário e um item. Este método de avaliação também, nem sempre é suportado por algumas plataformas, como por exemplo, em uma blogosfera. Este cenário dificulta muito a aplicação de tarefas voltadas aos Sistemas de Recomendação, ainda que as avaliações estejam expostas com clareza (GARCÍA-CUMBRERAS; MONTEJO-RÁEZ; DÍAZ-GALIANO, 2013).

Entender o que as pessoas pensam sempre foi uma peça importante no processo de tomada de decisão. Sendo assim, extrair opiniões expostas através de texto tem se tornado uma valiosa ferramenta que não está restrita ao comércio em que o foco é utilizar a opinião de clientes sobre itens avaliados. Esta ferramenta também é utilizada em outros cenários, como por exemplo, entender o pensamento de um público relacionado a um candidato em uma disputa política (PANG; LEE, 2008).

Para Serrano-Guerrero et al. (2015), este é um fato que impulsiona o início de muitos estudos, por exemplo, a mineração de opinião, a análise da subjetividade, a detecção de emoção ou detecção de opinião, a detecção de *spam*, entre outros. Os autores ainda afirmam a existência de muitas tarefas vinculadas à Análise de Sentimentos, porém, classificam as seguintes tarefas como as mais importantes: classificação de sentimentos, classificação de subjetividade, sumarização de opinião, recuperação de opinião, identificação de sarcasmo e ironia, entre outros.

De acordo com Pang e Lee (2008), o termo Análise de Sentimentos surgiu com os trabalhos de Das e Chen (2001) e Tong (2001). Pang e Lee ainda destacam que as diferenças entre Análise de Sentimentos (em inglês, *Sentiment Analysis*) e Mineração de Opinião (em inglês, *Opinion Mining*) estão restritas a pequenos detalhes. Sendo assim, para os autores os dois termos tornam-se sinônimos e atuam sobre o mesmo campo de pesquisa.

Em concordância com Pang e Lee (2008), outros autores como Li e Li (2013), Tang et al. (2015), Balazs e Velásquez (2016), Gutiérrez,

Vázquez e Montoyo (2016) também entendem que Análise de Sentimentos e Mineração de Opinião sejam sinônimos que tratam da mesma área de pesquisa. Já Esuli e Sebastiani (2006), destacam Análise de Sentimentos como o resultado de uma junção envolvendo dois campos de pesquisa, a Recuperação de Informação e a Linguística Computacional.

A Análise de Sentimentos tem como objetivo identificar opiniões, sentimentos, avaliações, atitudes e emoções das pessoas relacionadas a um produto ou serviço, quando expressadas através de texto. Para tal, este campo de pesquisa utiliza técnicas de processamento de linguagem natural (LIU, 2012; TANG et al., 2015) e técnicas de aprendizado de máquina (MULHOLLAND et al., 2015).

Outros autores como Agarwal et al. (2011), Gutiérrez, Vázquez e Montoyo (2016), entendem que Análise de Sentimentos é parte de uma tarefa existente no Processamento de Linguagem Natural (*Natural Language Processing* – NPL), tornando-se uma disciplina popular devido à sua grande relação a estudos de comportamento de mídia social sendo comumente usada para analisar os comentários que as pessoas postam em redes sociais. Além disso, ela permite identificar as preferências e os critérios dos usuários sobre situações, eventos, produtos, marcas, etc.

A literatura apresenta um consenso sobre a definição de Análise de Sentimentos como uma técnica utilizada para identificar a forma como os sentimentos são expressos no texto (NASUKAWA, YI, 2003; LIU et al., 2013; NAKATSUJI et al., 2014). Muitos estudos envolvendo Análise de Sentimentos avaliam as revisões dos usuários dentro de uma bipolaridade, ou seja, se representam sentimentos positivos ou negativos. Estudos recentes estendem a Análise de Sentimento para classificar comentários utilizando escalas de avaliação multiponto, uma tarefa conhecida como classificação por inferência (LEUNG et al., 2011).

Para analisar as revisões dos usuários, Kang e Park (2014) propõem a montagem de um dicionário de palavras e frases estruturado em três grupos: polaridade positiva que contém palavras como maravilhoso, ótimo, perfeito; polaridade negativa, por exemplo, péssimo, ruim, horrível, e o último grupo, denominado polaridade contextual ou neutra. Este grupo contém palavras e frases que podem transportar significados diferentes obedecendo o contexto de aplicação (KANG; PARK, 2014).

Em geral, a Análise de Sentimento é realizada em diferentes níveis de unidades de texto que incluem o nível de palavra, o nível de aspecto, o nível de frase e o nível de documento (SUN et al., 2015). Por aspecto, entende-se como um conceito ou característica extraída a partir de determinado texto.

3.1 ABORDAGENS E MÉTODOS

Muitos trabalhos buscam mostrar as diferentes técnicas aplicadas na Análise de Sentimento. Para Rosa, Rodriguez e Bressan (2015), a Análise de Sentimentos pode ser classificada em três abordagens, a saber: abordagem baseada em léxico (*lexicon based approach*), abordagem baseada em corpus (*corpus based approach*) e abordagem híbrida (*hybrid-based approach*).

Já na visão de outros autores como Serrano-Guerrero et al., (2015) e também Ravi e Ravi (2015), existe uma pequena divergência nesta classificação. Embora exista uma harmonia relacionada às abordagens *lexicon based approach* e *hybrid-based approach*, estes autores defendem que a terceira abordagem mais adequada seria a abordagem de aprendizado de máquina (*machine learning approach*), colocando a abordagem baseada em corpus (*corpus based approach*) como uma subdivisão da abordagem baseada em léxico (*lexicon based approach*).

Outros autores também classificam a Análise de Sentimentos em abordagens baseada em léxico e abordagens baseadas em aprendizado de máquina, porém não consideram a abordagem híbrida (KOUKOURIKOS; STOITSIS; KARAMPIPERIS, 2012, MARTÍNEZ-CÁMARA ET AL., 2012).

Para Ghorpade e Ragha (2012) existem basicamente dois tipos de abordagens para a análise de sentimento: método de aprendizagem de máquina e método de orientação semântica. Para os autores os algoritmos de classificação denominados classificadores são o núcleo de classificação de textos. Os algoritmos baseados na aprendizagem de máquina mais populares são *k-Nearest Neighbors* (KNN), *Naive Bayes* (NB) e *Support Vector Machine* (SVM).

De modo semelhante, no entendimento de Singh, Mukherjee e Mehta (2011), a Análise de Sentimentos usualmente emprega uma das duas abordagens a seguir: (a) o uso de classificadores de texto, como NB, SVM e KNN para categorizar documentos em dois grupos (positivo/negativo); (b) uma abordagem não supervisionada de

orientação semântica para calcular o sentimento contido nos documentos baseado em *tags* atribuídas para cada palavra no documento.

Para Lunardi et al. (2016), a Análise de Sentimento pode ser realizada utilizando algoritmos de aprendizagem de máquina, análise léxica, palavras-chave ou ontologias.

De uma forma geral, trabalhos na área da Análise de Sentimentos focam em duas direções, sendo, as abordagens léxicas e os métodos supervisionados como técnicas de aprendizado de máquina (KARAMPIPERIS; KOUKOURIKOS; STOITSIS, 2014).

3.1.1 Machine Learning Approach

A construção de características é uma parte importante nos métodos de aprendizado de máquina para a análise de sentimento. A aprendizagem de máquina é útil para definir uma regra que restringe as características, buscando incluir apenas as informações relevantes (SUN et al., 2015).

As técnicas de aprendizado de máquina concentram-se na seleção de vetores de características para promover suporte aos classificadores (algoritmos que realizam a classificação). O caminho mais comum para a escolha dos vetores de características são a inclusão de unigramas ou *n*-gramas, o somatório do número de palavras positivas/negativas, o tamanho do documento, etc. (KARAMPIPERIS; KOUKOURIKOS; STOITSIS, 2014). Alguns fatores como linguística, estatística e *n*-gramas influenciam no processo de aprendizado de máquina (Zhang et al., 2010).

Nesta tarefa, as técnicas de Processamento de Linguagem Natural desempenham um papel importante. Entre as características mais utilizadas estão, por exemplo: (1) os termos (palavras ou *n*-gramas) e suas frequências; (2) as informações parte de um texto (*Part Of Speech - POS*) onde os adjetivos desempenham um papel importante e os substantivos podem ser significativos; (3) os termos negativos uma vez que estes podem alterar o significado de qualquer frase; e (4) as dependências sintáticas que podem determinar o significado da frase (SERRANO-GUERRERO et al.; 2015).

O sucesso da aprendizagem de máquina depende principalmente da seleção e extração do conjunto apropriado de recursos usados para detectar sentimentos. Esta abordagem requer uma grande quantidade de dados para obter um resultado de sentimento confiável, porque uma frase incomum pode causar ruído na aferição do sentimento (ROSA; RODRIGUEZ; BRESSAN, 2015). Ou seja, é extremamente dependente

da disponibilidade de conjuntos de dados rotulados (MARTÍNEZ-CÁMARA ET AL., 2012).

Alguns classificadores frequentemente implementados são, por exemplo, o *Naive Bayes (NB)*, o *Support Vector Machine (SVM)* e o *K-Nearest Neighbors (KNN)* (KARAMPIPERIS; KOUKOURIKOS; STOITSIS, 2014, GHORPADE; RAGHA, 2012). Estas técnicas podem ser agrupadas em duas categorias principais: supervisionadas e não supervisionadas (SERRANO-GUERRERO et al., 2015, SCHWENKER; TRENTIN, 2014).

3.1.1.1 *Supervised Machine Learning (SML)*

No SML o computador tenta replicar as decisões de codificação dos codificadores humanos. Assim, preservam-se os pontos fortes de codificação humana juntamente com o seu rigor sistemático e sensibilidade contextual na direção do significado latente e as sutilezas da linguagem humana, maximizando a capacidade e precisão dos métodos computacionais (LEWIS; ZAMITH; HERMIDA, 2013).

Para van Zoonen e Van der Meer (2016), SML oferece vantagens sobre procedimentos de codificação manual ou abordagens baseadas em dicionários. A principal vantagem é que SML permite aos pesquisadores expandirem o âmbito da sua análise. Adotando SML para pesquisa em comunicação organizacional permite uma pesquisa com mais nuances, condicional e comparativa, o que é relevante visto que a informação *online* se torna cada vez mais disponível. Além disso, uma vez que o classificador é treinado para codificar as categorias de conteúdo pode ser aplicado a grandes conjuntos de dados.

Zhang et al. (2010) alertam que embora indicações como a de Pang, Lee e Vaithyanathan (2002) de que os métodos supervisionados tenham um desempenho superior aos métodos não supervisionados, devido aos grandes conjuntos de treinos que o constroem, é necessário ter em mente o contraponto feito por Qiu et al. (2009) onde é destacado que o processo de construção de conjuntos de treinos pode ser desnecessariamente demorado, podendo tornar inviável a sua aplicação em larga escala. Já Zhang et al. (2010), alertam que o desempenho dos métodos supervisionados sofre um decréscimo quando os dados para treino da aprendizagem são insuficientes. Nas seções seguintes são apresentados alguns classificadores supervisionados.

3.1.1.1.1 Support Vector Machine (SVM)

O *Support Vector Machine* (SVM) foi proposto por Vladimir N. Vapnik, primeiramente usando classificadores lineares e depois classificadores não lineares. Esta técnica tem o objetivo de encontrar o limite para decidir entre a classificação em duas ou mais classes, baseando-se em treinamento de dados (CECI, 2015).

Sun et al., (2015) analisam SVM como uma técnica de mineração de dados com capacidade já aprovada em muitas aplicações. Amplamente mencionada na literatura por sua excelente capacidade de desempenho de aprendizagem e generalização, esta técnica pode resultar em modelos matemáticos complexos, porém, tende a capturar as características inerentes dos dados de forma mais concisa.

3.1.1.1.2 Naive Bayes (NB)

Naive Bayes tem sido bem conhecido por sua simplicidade e suas aplicações bem-sucedidas para muitos problemas de classificação. Este algoritmo também tem sido aplicado para análise de sentimento binário e geralmente apresenta bons resultados apesar da sua simplicidade (LEUNG et al., 2011).

Capdevila, Arias e Arratia (2016) e Ceci (2015), definem um classificador *Naive Bayes* como um modelo probabilístico que prevê as chances de um elemento (documento ou sentença) fazer parte de uma determinada classe.

3.1.1.1.3 K-Nearest Neighbors (KNN)

Em particular, a pesquisa de similaridade representa o núcleo de muitos sistemas de Recuperação de Informação. Permite realizar consultas em conjuntos de documentos, apresentando ao usuário uma lista de documentos ordenados pelo documento com maior pontuação. O algoritmo *KNN* é comumente utilizado para esta função, recuperando os documentos com mais semelhanças para cada consulta considerando um determinado valor (fator k). No entanto, *KNN* é conhecido por ter baixo desempenho (tempo de execução) em tarefas de classificação em comparação com outros métodos supervisionados. Normalmente não é a melhor escolha para lidar com as classificações *on-the-fly* (CANUTO et al., 2015).

3.1.1.2 *Unsupervised Learning*

Ao contrário das abordagens supervisionadas, as abordagens sem supervisão fazem a suposição de que existem certas palavras que as pessoas tendem a usar para expressar seus sentimentos, e isto tende a ser suficiente para, no contexto da Análise de Sentimento, classificar textos (Zhang et al., 2010, ZHANG et al., 2013).

Entre as técnicas aplicadas com mais frequência na abordagem de aprendizado não supervisionado está a *Pointwise Mutual Information (PMI)* (PELEJA et al., 2013).

3.1.1.2.1 *Pointwise Mutual Information (PMI)*

Esta técnica é conhecida pela sua capacidade de medir a força de associações semânticas. *Pointwise Mutual Information (PMI)* visa estimar o grau de dependência estatística entre palavras, observando a probabilidade de coocorrerem em conjunto, e individualmente. (PELEJA; DIAS; MAGALHAES, 2012).

O grau de dependência estatística será realizado entre a palavra a ser aferida e duas palavras de referência sendo uma palavra de referência positiva e uma palavra de referência negativa. Logo, a alta coocorrência entre a palavra a ser aferida e a palavra positiva indica um sentido positivo. Por exemplo, a palavra sorvete será classificada como positiva houve uma alta coocorrência entre "sorvete" e a palavra de referência "excelente" (PELEJA et al., 2013).

3.1.2 *Lexicon Based Approach*

A aplicação da abordagem léxica depende da criação de dicionários apropriados. Os termos presentes no dicionário são marcados respeitando a sua polaridade. O sentimento geral de um texto é calculado através da comparação dos termos positivos e negativos contidos nele com os termos que compõem o dicionário. Apesar de sua aparente simplicidade esta abordagem já apresentou resultados significativos (KARAMPIPERIS; KOUKOURIKOS; STOITSIS, 2014).

Em concordância Serrano-Guerrero et al., (2015) afirmam que esta abordagem depende principalmente de um léxico de sentimento, ou seja, uma coleção de termos conhecidos e pré-compilados, frases e expressões idiomáticas desenvolvidas para gêneros tradicionais de comunicação. Os autores ainda acrescentam que estruturas mais

complexas como ontologias também podem ser usadas para essa finalidade.

Na visão de Martínez-Cámara et al., (2012), o fato da indicação do sentimento contido na mensagem depender da presença destas palavras no léxico, gera um problema frequente nesta abordagem que o autor define como baixos valores de recordação.

Para Lin et al., (2015), os métodos utilizados pela abordagem léxica são intuitivos e fáceis de implementar. Porém estes métodos podem não funcionar adequadamente em alguns casos, pois muitas palavras são polissêmicas e o seu sentimento é dependente do domínio.

Serrano-Guerrero et al., (2015) ainda subdividem esta abordagem em duas linhas: baseado em dicionário (*Dictionary-Based Approach*) e baseado em corpus (*Corpus-Based Approach*). Conhecidas também como abordagens manuais estas subdivisões podem ser altamente demoradas, por isso, é muito comum encontrar sua aplicação em conjunto com métodos automatizados (PELEJA; DIAS; MAGALHAES, 2012).

Nesta abordagem, cabe ressaltar a contribuição de Pröllochs, Feuerriegel e Neumann (2016), onde os autores alertam que o tratamento de negação tem uma considerável participação para a melhorar a acurácia do processo. Como tal, identificar e prever palavras de negação no texto são atividades cruciais para fornecer um suporte de decisão preciso, uma vez que as sentenças são suscetíveis a serem classificados erroneamente.

3.1.2.1 *Corpus-Based Approach*

Normalmente, uma abordagem baseada em corpus baseia-se na identificação de padrões de coocorrência (PELEJA; DIAS; MAGALHAES, 2012). Esta abordagem surge com o objetivo de fornecer dicionários relacionados a um domínio específico. Esses dicionários são gerados a partir de um conjunto de termos (sementes) que cresce através da busca de palavras relacionadas por meio da utilização de técnicas estatísticas ou semânticas. Métodos baseados em estatísticas, como análise semântica latente (*Latent Semantic Analysis - LSA*) ou simplesmente a frequência de ocorrência das palavras dentro de uma coleção de documentos também podem ser utilizados, por outro lado, métodos semânticos, tais como o uso de sinônimos e antônimos também podem ser uma solução interessante (SERRANO-GUERRERO et al.; 2015).

3.1.2.2 Dictionary-Based Approach

Esta abordagem geralmente é baseada na utilização de um conjunto inicial de termos denominados como sementes que normalmente são recolhidos e anotados de uma forma manual (PELEJA et al., 2013). Este conjunto cresce, pesquisando os sinônimos e antônimos de um dicionário. Porém, a principal desvantagem deste tipo de abordagem é a incapacidade de lidar com orientações específicas de domínio ou contexto. Mesmo assim, pode ser uma solução interessante, dependendo do problema (SERRANO-GUERRERO et al.; 2015).

Rosa, Rodriguez e Bressan (2015), utilizam o dicionário manual para definir o nível de intensidade de sentimento. Neste dicionário cada palavra tem uma respectiva classificação, por exemplo, numa escala positivo de 1 a 5 ou uma escala negativa de -1 a -5. Uma vez definido o dicionário, a métrica de intensidade de sentimento pode ser modelada. A métrica básica para obter o sentimento de uma frase é comumente obtida por uma soma aritmética de cada palavra encontrada no dicionário.

3.2 APLICAÇÕES

Encontrar o humor e as preferências dos usuários por meio de técnicas de análise de texto é uma área de pesquisa que tem sido bastante ativa na última década. Esse campo tem progredido muito desde suas primeiras técnicas, não atuando apenas como um classificador de opinião positiva e negativa, mas também com escalas de avaliação multiclasse respeitando o domínio de aplicação (PELEJA et al., 2013).

Assim, as opiniões dos usuários surgem com uma valiosa informação que é aplicada com sucesso pelos Sistemas de Recomendação (GARCÍA-CUMBRERAS et al., 2013). Os Sistemas de Recomendação que incorporam tarefas de Análise de Sentimentos são aplicados em diversos domínios. Alguns exemplos serão apresentados a seguir.

Singh et al. (2013) utilizam a Análise de Sentimentos para a recomendações de *e-books*. Na recomendação de programas de TV e filmes encontra-se o trabalho de Peleja et al. (2013). Em uma área semelhante, Mulholland et al. (2015) aplicam a análise de sentimentos na recomendação de filmes, assim como, o trabalho de Singh, Mukherjee e Mehta (2011). Ainda no entretenimento, a recomendação

de músicas é contemplada pelas pesquisas de Rosa, Rodriguez e Bressan (2015) e Colace e Casaburi (2016).

No turismo, as opiniões dos usuários são consideradas na recomendação de hotéis. Esta área de pesquisa é encontrada nos trabalhos de Lunardi et al. (2016), Levi et al. (2012), Dong, O'mahony e Smyth (2014). A gastronomia também é contemplada nas pesquisas de Kim e Song (2013) e Suresh, Roohi e Eirinaki (2014) com as recomendações de restaurantes.

Os *e-commerces* são uma área comum de atuação da Análise de Sentimentos. As opiniões dos usuários sobre os produtos também são utilizadas buscando aprimorar as recomendações para usuários que fazem compras em sites de vendas como Amazon® (XIE et al., 2012, ISSA et al., 2015).

Entender a opinião dos eleitores também é um problema que podem ser resolvidos com a Análise de Sentimentos. As postagens de eleitores em redes sociais foram uma importante fonte de informações para as pesquisas de Gurini e Gasparetti (2013) e Yang et al. (2013).

Na área da saúde, a Análise de Sentimentos é sugerida como uma ferramenta para auxiliar a identificação de surtos epidêmicos. Para tal, é utilizada como base as postagens de usuários em redes sociais como Twitter® (CULOTTA, 2010).

4 GESTÃO DE IDEIAS

Conforme o Manual de Oslo, a inovação é a implementação de um novo ou significativamente melhorado produto (bem ou serviço), ou processo, um novo método de marketing ou um novo método organizacional em práticas de negócios, organização no local de trabalho ou relações externas (OECD, 2005).

Para (QUANDT et al., 2014), num ambiente competitivo marcado por mudanças súbitas, as organizações estão cada vez mais dependentes de sua capacidade de explorar seus ativos intelectuais para conceber ideias que resultem em produtos e serviços inovadores.

Sérgio, Gonçalves e Souza (2015) também destacam que o Mercado global possui uma forte competitividade que exerce sobre as organizações a necessidade e investir na inovação de seus produtos/serviços. Porém a inovação é o resultado de um processo complexo que visa a manutenção da organização no mercado e obter alguma vantagem competitiva.

Segundo Clark e Wheelwright (1993) a inovação é fundamental para a sustentabilidade de organizações concorrentes que necessitam conquistar novos mercados e estão enquadradas em um cenário competitivo. Marzano (2005) e Moos et al. (2011) também destacam a inovação como um processo sistemático para obter vantagem competitiva.

Bautzer (2009) afirma que através da inovação agrega-se valor aos produtos e serviços, sendo fundamental que as organizações compreendam sua capacidade em gerar inovações. Para Schweisfurth e Herstatt (2014), a inovação acontece quando o conhecimento das necessidades do mercado e as soluções tecnológicas são combinadas.

Em qualquer negócio a inovação é considerada um processo com estágios distintos, desde a geração de ideias até a implementação. Dentre estes estágios o processo de geração de ideias ocupa uma parte substancial do processo de inovação e determina fundamentalmente o sucesso da inovação posterior, uma vez que o desempenho de novos produtos e serviços depende principalmente da criatividade de determinada ideia. Isso torna o processo de geração de ideias uma questão central para o processo de inovação (GEUM; PARK, 2016).

A inovação é um processo em que as empresas transformam ideias em produtos, serviços ou processos novos ou melhorados. Sendo assim, a inovação é uma atividade crítica para manter a competitividade das empresas no mercado. Ciente disto, as empresas continuam a

investir no desenvolvimento de novos produtos, serviços e processos (LI; KANKANHALLI; KIM, 2016).

De acordo com Mahroum e Al-Saleh (2013) a capacidade de gerar ideias inovadoras e implementá-las tornou-se uma estratégia poderosa para o crescimento econômico em uma economia globalizada.

A geração de ideias frequentemente é o ponto de partida da inovação, portanto, é o primeiro e mais importante estágio de desenvolvimento e inovação de novos produtos (GEUM; PARK, 2016).

Com o objetivo de estimular a criatividade dos seus colaboradores e direcioná-la às questões críticas do negócio, muitas empresas passaram a estruturar iniciativas formais de gestão de ideias, como programas ou sistemas de sugestões, visando capturar, avaliar, compartilhar e implementar eficazmente as ideias mais relevantes para a sua competitividade (QUANDT et al., 2014).

Para Yoshida et al. (2014), a gestão de ideias também pode ser usada para capturar a reação emocional e racional do cliente com o objetivo de testar e melhorar ideias propostas.

Vandenbosch, Saatcioglu e Fay, (2006) definem o gerenciamento de ideias como um processo de reconhecimento da necessidade de ideias, com a capacidade de gerar e avaliar novas ideias (VANDENBOSCH; SAATCIOGLU; FAY, 2006).

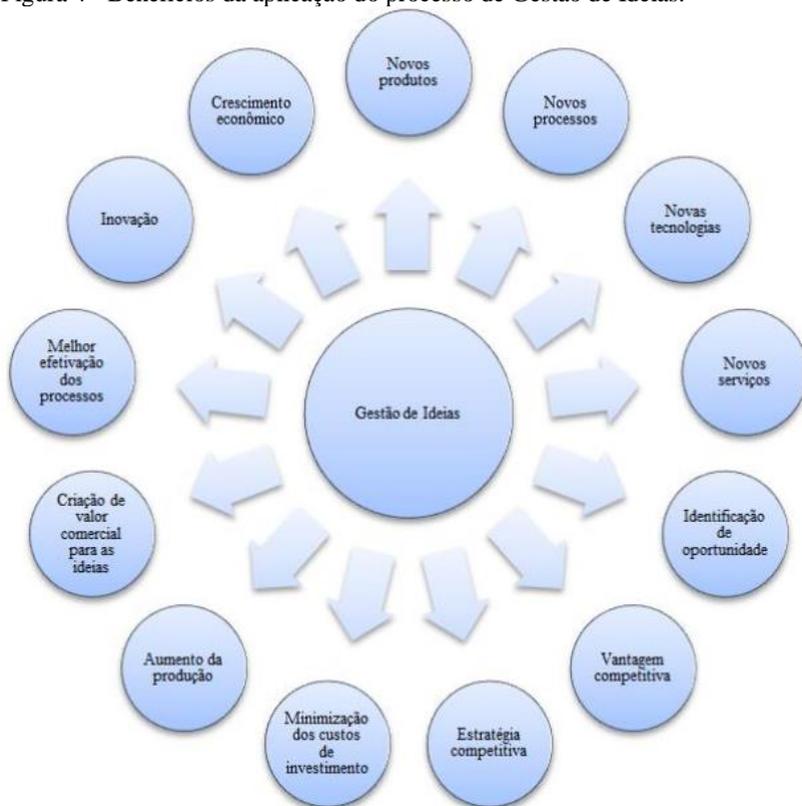
Bothos et al. (2012) entendem que a gestão de ideias está localizada no *front-end* da inovação caracterizado por Bothos, Apostolou & Mentzas (2012) como o centro da gestão da inovação e definido por Mohan, Voss e Jiménez (2017) o primeiro e o mais importante estágio no processo da inovação. Como as ideias se constituem na matéria-prima para a inovação, o gerenciamento de ideias pode ser considerado o núcleo da gestão da inovação. Sendo assim, a Gestão de Ideias tem um papel fundamental no processo de inovação (BARBIERI, ÁLVARES, CAJAZEIRA, 2009), uma vez que as ideias são a matéria-prima para o processo de criação de novos produtos (COOPER; EDGETT; KLEINSCHMIDT, 2002) e a filtragem de ideias é um ponto crítico do processo de inovação, sendo determinante para a falha ou sucesso da organização (ADAMS, BESSANT, PHELPS, 2006).

Sandström e Björk (2010) destacam que a Gestão de Ideias possui como objetivo claro de originar, avaliar e eleger ideias buscando obter produtos e serviços inovadores.

Conforme o trabalho desenvolvido por Sérgio (2016), os benefícios obtidos com a aplicação da gestão de ideias são muitos, dentre estes, cita-se com exemplo: Novos produtos, inovação,

crescimento econômico. A Figura 4 apresenta os benefícios da gestão de ideias.

Figura 4 - Benefícios da aplicação do processo de Gestão de Ideias.



Fonte: Sérgio (2016).

A implementação bem-sucedida de uma ideia criativa, que pode ser expressada na forma de conhecimento, prática ou objeto físico, é considerada uma inovação. O ciclo de vida da inovação se estende desde a concepção inicial de uma ideia até a produção completa e lançamento no mercado de um produto. Os processos de inovação dentro deste ciclo de vida incluem a triagem preliminar de ideias, a avaliação de sua viabilidade, o desenvolvimento de um caso de negócio e a formação e execução de um novo projeto, incluindo testes e validação. (BOTHOS; APOSTOLOU; MENTZAS, 2012).

Para Magnusson, Netz e Wästlund (2014), em uma abordagem geral, a triagem de ideias deve empregar um processo de avaliação racional usando a avaliação formal de critérios específicos. Isso geralmente é ilustrado como um processo de filtragem em que todas as ideias são colocadas em um funil e depois avaliadas usando uma série de critérios predefinidos

Sobre o sucesso de processo de Gestão de Ideias Brem e Voigt (2009) alertam que este processo possui uma forte dependência da estrutura de processo correta para os diferentes tipos de ideias e da correspondente implementação organizacional adequada. Já para Westerski e Iglesias (2011) a etapa mais importante e problemática do processo de Gestão de Ideias está diretamente vinculada à avaliação das ideias. Em concordância, Adams, Bessant e Phelps (2006) destacam que a filtragem e a seleção das ideias são fatores que influenciam no êxito ou falha do processo.

Porém, selecionar ideias inovadoras é um processo complexo, pois estas ideias podem ser encontradas dentro e fora da organização (MCGRATH, 2004). Preocupados em reduzir os custos e riscos no processo de invocação, sem perder a coragem de inovar de suas empresas, os gestores das organizações decidiram envolver os clientes e usuários neste processo (LI; KANKANHALLI; KIM, 2016).

Grandes empresas tem tomado a decisão de criar comunidades virtuais para colher, analisar e executar ideias que tem grande potencial para se tornar um produto ou serviço. Estas ideias são oferecidas deliberadamente por seus clientes que possuem livre acesso a esta comunidade (LEE et al., 2013).

Schweisfurth e Herstatt (2014) em seu trabalho também apontam que o envolvimento do cliente tem sido cada vez mais presente no processo de inovação. Outros autores como Huizingh (2011) Martinez-Torres e Olmedilla (2016), também alertam para a contribuição do cliente no processo de inovação.

Para Brem e Voigt (2009), a inovação é uma responsabilidade de todas as unidades de negócio e departamentos, sendo assim, sua aplicação deve ser realizada de maneira clara e bem definida. Nesse contexto, a capacidade de uma organização para identificar, adquirir e utilizar ideias (externas) pode ser vista como um fator crítico em relação ao seu sucesso no mercado.

Para auxiliar o gerenciamento das ideias, tem sido criados sistemas computacionais para o domínio de Gestão de Ideias denominados Sistemas de Gestão de Ideias. Estes sistemas auxiliam na geração, avaliação e seleção de ideias inovadoras (LI; LI; CHEN, 2014).

4.1 SISTEMAS DE GESTÃO DE IDEIAS

Ao longo do tempo, os sistemas de gerenciamento de ideias se tornaram cada vez mais sofisticados. Este desenvolvimento foi influenciado pela evolução das ferramentas de TI que permitem um tratamento sistemático e eficiente das ideias. À medida que esses sistemas se tornaram mais avançados, o uso deles também foi ampliado. Assim, os sistemas de gerenciamento de ideias são hoje mais úteis, não só para melhorias contínuas, mas também para criar novos produtos. Os sistemas de sugestões podem ser uma lucrativa fonte de inovação nas organizações (SANDSTRÖM; BJÖRK, 2010).

Originalmente na língua inglesa como *Idea Management System* (IMS), os Sistemas de Gestão de Ideias, foram desenvolvidos com o objetivo de gerenciar as ideias coletadas por uma organização. Estes sistemas promovem suporte para que as empresas gerenciem o processo de descoberta, incubação, aceleração e dimensionamento das ideias para obter valor comercial através da criação de produtos e processos inovadores (POVEDA; WESTERSKI; IGLESIAS, 2012).

Para Westerski, Dalamagas e Iglesias (2013), os sistemas informatizados de gestão de ideias, em geral, utilizam formas de solicitar, capturar e registrar ideias, ferramentas de colaboração, mecanismos de recompensas, e ferramentas para avaliação e seleção de ideias e para a geração de relatórios.

No entendimento de Quandt et al., (2014), os sistemas de gestão de ideias podem ser definidos como um subprocesso da gestão da inovação focado na geração, avaliação e seleção de ideias que podem levar à implementação de melhorias e inovações. Esses sistemas podem proporcionar diversos benefícios. Uma vez aplicados de maneira eficaz, eles facilitam o compartilhamento e transparência, permitindo que todos tenham acesso aos resultados e melhores práticas, mesmo em companhias que possuem instalações geograficamente dispersas.

Com o objetivo de estimular a inovação e melhorar seus produtos e serviços, grandes organizações como Dell®, Starbucks®, Cisco® e Canonical®, vem desenvolvendo aplicativos de Gestão de Ideias para aumentar o envolvimento de seus clientes na participação deste processo (WESTERSKI; DALAMAGAS; IGLESIAS, 2013).

No entanto, os Sistemas de Gestão de Ideias enfrentam importantes desafios relacionados com a grande quantidade de esforço humano, necessitando de um especialista durante o processo de gestão de ideias. As principais origens destes problemas são: a) o excesso de informação através do grande volume de ideias submetidas; b) picos

súbitos de submissões de ideias; c) a redundância das ideias, e d) grandes quantidades de ideias triviais (MARTINEZ-TORRES; OLMEDILLA, 2016).

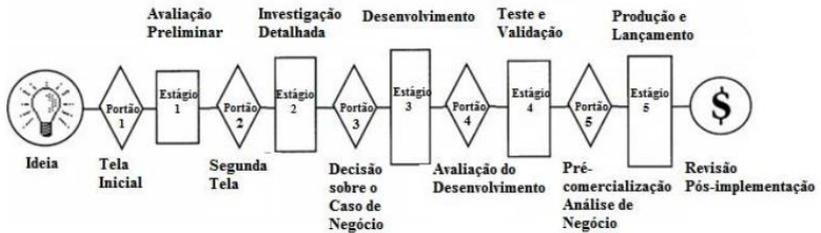
Outro ponto importante nos Sistemas de Gestão de Ideias aborda o fato de uma ideia ser encontrada diversas vezes na base de dados com classificações diferentes, porém relacionadas. Estas ideias quando analisadas individualmente não despertam interesse, mas quando agrupadas revelam grande potencial. Nestes casos, a aplicação de análise de agrupamentos pode conduzir a detecção de grupos revelando possíveis tendências ou demandas (MAGNUSSON; NETZ; WÄSTLUND, 2014).

4.2 MODELOS PARA GESTÃO DE IDEIAS

A Gestão de ideias possui modelos que visam servir de suporte para o processo de gerenciamento de ideias. A literatura apresenta a inovação através de duas interpretações, a saber, inovação como processo e inovação como resultado. Cooper (1990) apresenta inovação como processo, e como tal, pode ser gerenciado. Com o passar do tempo, este modelo foi aprimorado mantendo sua essência.

4.2.1 Modelo *Stage Gate*

O modelo proposto por Cooper (1990) é estruturado em fases e tem a ideia como a primeira etapa no processo de criação de um produto. A primeira fase do modelo contempla a avaliação preliminar do projeto. A definição sobre implementação do projeto esta alocada na segunda fase, precedendo ao desenvolvimento alocado na terceira fase. Na quarta fase são realizados os testes e a validação do projeto. Na quinta fase o produto é comercializado. Ao final o modelo apresenta ainda a revisão pós-implementação. A Figura 5 ilustra o modelo proposto.

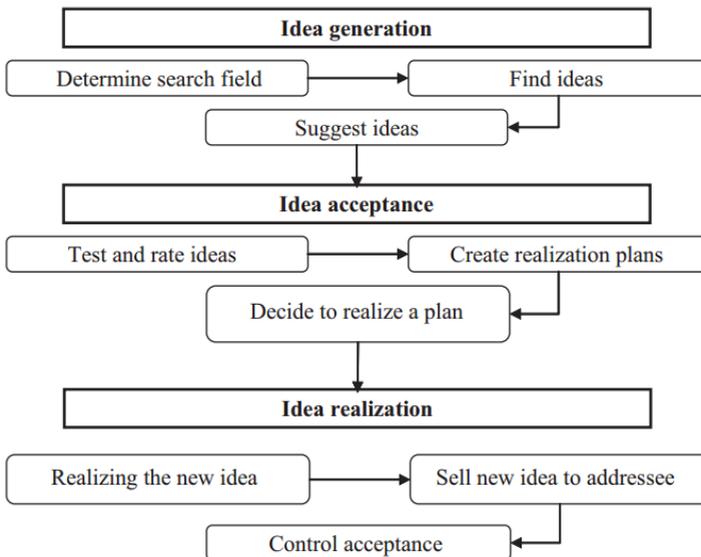
Figura 5 - Modelo *Stage Gate*

Fonte: Cooper (1990).

4.2.2 Estágios da inovação

De acordo com Brem e Voigt (2009), para obter sucesso com a aplicação da gestão de ideias é necessário que este processo seja aplicado de maneira correta. Sua estrutura deve estar adequada aos diferentes tipos de ideias e sua implementação necessita ser compatível com a empresa. Este processo pode ser organizado em estágios agrupados em três seções: *idea generation*, *idea acceptance* e *idea realization*, conforme ilustrado na Figura 6.

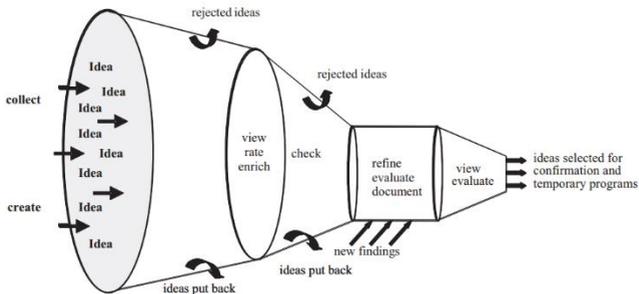
Figura 6 - Estágios da Inovação



Fonte: Brem e Voigt (2009).

Este modelo idealizado por Brem e Voigt (2009), foi inspirado em uma abordagem orientada para o fluxo, o chamado "túnel de ideias" (VOIGT; BREM, 2009) ilustrado na Figura 7 e originada de um conceito antigo denominado "funil de desenvolvimento", considerado o modelo básico elementar para considerações *front-end*.

Figura 7 - Túnel de ideias



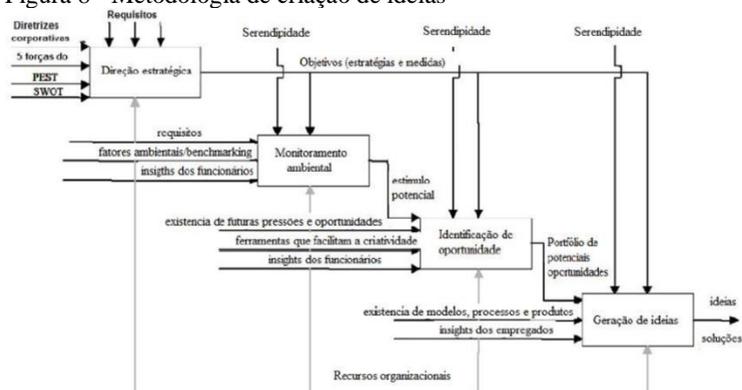
Fonte: Voigt e Brem (2009).

4.2.3 Metodologia de criação de ideias

Flynn *et al.* (2003) expõem uma metodologia fundamentada nos colaboradores da organização, nas habilidades, nos conhecimentos e na comunicação.

Esta metodologia busca prover suporte ao processo de criação de ideias, ao alinhamento das metas e ao reconhecimento da oportunidade de definição da ideia. Assim, alinhando a ideia com a estratégia da organização e a identificação de oportunidade, visando escolher a ideia que será colocada em prática conforme demonstrado na Figura 8.

Figura 8 - Metodologia de criação de ideias



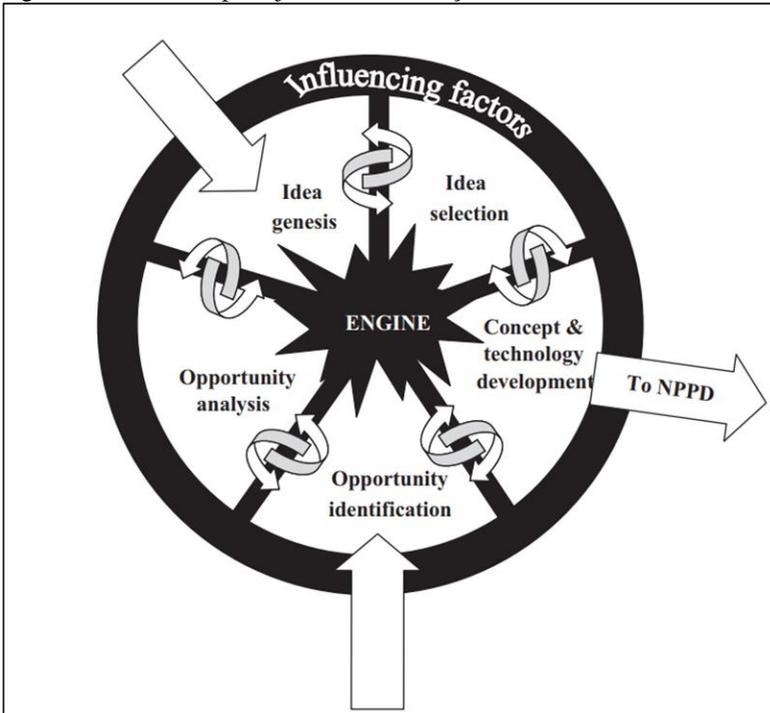
Fonte: Flynn *et al.* (2003).

4.2.4 Modelo de Desenvolvimento de Novo Conceito.

Entre os modelos mais populares está o de Koen *et al.* (2002), conhecido como Modelo de Desenvolvimento de Novo Conceito (originalmente, *New Concept Development Model*) que propõem fornecer uma linguagem comum para atividades de *front-end*. O modelo exibe uma forma circular de fluxo, contendo cinco etapas, sendo, a identificação de oportunidades, a análise de oportunidade, a geração de ideias, a seleção de ideias e a definição de conceito. O modelo ainda considera a influência de fatores externos no processo.

No centro do modelo, está a engrenagem que gera o fluxo cíclico das cinco etapas conforme ilustra a Figura 9.

Figura 9 - Modelo de planejamento da inovação.



Fonte: Koen *et al.* (2002).

4.3 INICIATIVAS COM GESTÃO DE IDEIAS.

Muitas organizações como Microsoft®, IBM®, Whirlpool®, Dell®, Starbucks®, Cisco® e a Canonical®, têm aplicado Sistemas de Gestão de Ideias como uma forma de absorver o conhecimento coletivo de seus funcionários e clientes. Exemplos desta classe de sistema incluem o *Idea Storm* da Dell® iniciado em Fevereiro de 2007 e o *My Starbucks Ideas* da Starbucks® em execução desde Março de 2008 (BAILEY; HORVITZ, 2010, WESTERSKI; DALAMAGAS; IGLESIAS, 2013).

Embora os Sistemas de Gestão de Ideias auxiliem no processo de inovação de produtos e serviços, muitas organizações encontram problemas para estabelecer, através destas plataformas, comunidades capacitadas de manter iniciativas de Gestão de Ideias, pois não conseguem atrair um número adequado de participantes capacitados. Neste caso, estas organizações têm recorrido às redes sociais como

Facebook® (SALDIVAR et al., 2016) e Twitter® (SALDIVAR et al., 2017), visando construir comunidades de inovação.

Analisando o âmbito empresarial nacional, pode-se destacar a pesquisa realizada por Quandt *et al.* (2014) que foi aplicada em 41 empresas situadas na região sul do Brasil, todas contendo mais de 500 funcionários e receita operacional bruta anual superior a R\$ 300 milhões. Conforme a pesquisa, 70% dessas empresas confirmaram o uso de programas ou sistemas que solicitam, estimulam, avaliam, e implementam sugestões oferecidas por seus colaboradores, recompensando-os por ideias que tenham aplicações práticas e produzam resultados para organização. Em geral, as empresas incentivam seus colaboradores a apresentarem ideias relacionadas à inovações em processos, produtos, relacionamento com clientes, novas oportunidades de negócios ou novos mercados, e melhorias no local de trabalho, como: segurança, conforto, eficiência e produtividade.

Pesquisas recentes buscam aprimorar a aplicação da Gestão de Ideias, dentre elas pode-se citar o trabalho de Sérgio, Gonçalves e Souza (2015), que propõem um modelo baseado em ontologia e na análise de agrupamento para auxiliar no processo de tomada de decisão no domínio de gestão de ideias. Neste trabalho observou-se que através da análise de agrupamento pode-se evidenciar padrões, tendências ou demandas de mercado conforme as ideias presentes em um determinado domínio. Objetiva, portanto, o auxílio a especialistas em suas decisões de quais ideias são passíveis de implementação.

Kultima e Karvinen (2016), em seu trabalho analisam o impacto da Gestão de Ideias no domínio da indústria de jogos eletrônicos. Os autores destacam que as práticas de Gestão de Ideias variam em diferentes empresas bem como entre os atores que participam do processo de desenvolvimento exercendo influência em todas as fases do processo de criação e manutenção de projetos de sucesso.

Porém a aplicação da Gestão de Ideias não se limita apenas aos domínios comerciais. Com o auxílio da *Web 2.0* esta área vem contribuindo no domínio da política e da participação cívica aberta ao público. Um exemplo proeminente dessa prática ocorreu recentemente na Finlândia, onde o público participou de uma iniciativa de reforma da lei do tráfego rodoviário. Nesta ocasião, os cidadãos participaram do processo legislativo apresentando suas ideias, comentando e votando em ideias através de uma plataforma *Web* (SALDIVAR et al., 2017).

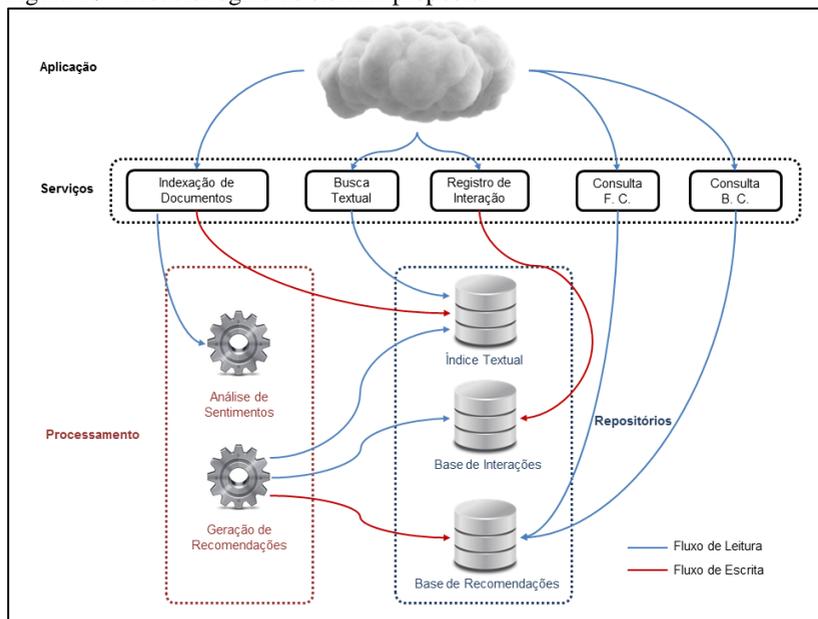
5 MODELO PROPOSTO

O objetivo deste capítulo é apresentar o modelo proposto. Tal apresentação será feita em duas etapas. Na seção 5.1 será explicado o modelo lógico, seus detalhes e suas interações entre os módulos e componentes. A seção 5.2 destina-se ao modelo físico com o intuito de descrever seus componentes tecnológicos e a justificativa para a utilização dos mesmos.

5.1 MODELO LÓGICO

O modelo lógico ilustrado pela Figura 10 é composto por camadas que possibilitam a interconexão de conteúdo, visando gerar a recomendação deste conteúdo aos usuários. Na Figura 10 o ponto inicial é representado pela nuvem que simboliza a *Web*, seus usuários e o conteúdo gerado e consumido por eles em um ambiente.

Figura 10 - Modelo lógico do sistema proposto.



Fonte: Elaborado pelo autor.

Abaixo da nuvem está a camada de serviços que promove a interação no modelo proposto com a *Web* e seus usuários. Esta camada é composta pelos seguintes serviços:

- Indexação de documentos: Este serviço envia o documento que então é analisado para identificar a sua polaridade, ou seja, se o texto é positivo ou negativo. Após receber o resultado da análise o documento é enviado para ser persistido no repositório de índice textual.
- Busca textual: Fornece ao usuário uma busca no repositório de índice textual populado pelo serviço de indexação.
- Registro de interações: As interações realizadas pelos usuários são persistidas no repositório de interações. Entre as interações citam-se, o acesso a determinado documento e a avaliação do documento em alguma escala indicando seu interesse no mesmo.
- Consulta Filtragem Colaborativa: Possibilita o acesso às recomendações realizadas pelo viés da Filtragem Colaborativa.
- Consulta Baseada em Conteúdo: Promove o acesso às recomendações geradas pelo prisma da Filtragem Baseada em Conteúdo.

O modelo lógico apresentado na Figura 10 está agrupado em 3 seções demarcadas por áreas pontilhadas (serviços, processamento e repositórios). O principal foco deste trabalho está voltado para a seção de processamento. As demais seções assim como a nuvem constam para representar os serviços consumidores de recomendações e a necessidade de informações básicas, como por exemplo, as interações dos usuários e um conjunto de documentos.

Na seção de processamento encontram-se os processos de análise de sentimentos e geração de recomendações. O primeiro processo é executado antes que o documento seja indexado no repositório de índice textual. Este processo percorre o texto retirando cada palavra (*token*) que por sua vez será analisado através da comparação com um dicionário (*lexicon*) organizado em dois conjuntos, sendo um com palavras positivas e outro com palavras negativas. Uma vez aplicada a comparação de todos os *tokens* que constroem o texto, este processo realiza um cálculo para identificar se o texto possui mais palavras positivas ou mais palavras negativas. A polaridade representada pela maioria dos *tokens* será aplicada ao texto como uma característica.

Devidamente polarizado este documento é retornado ao serviço de indexação de documentos para ser persistido no repositório de índice textual.

O segundo processo do modelo proposto é denominado geração de recomendações. Como o próprio nome sugere este processo realizará a recomendações Baseadas em Filtragem Colaborativa e as recomendações Baseadas em Conteúdo. Este processo será executado sobre a base de interações e de documentos e avaliará os itens que ainda não foram processados ou sofreram atualização através das interações dos usuários.

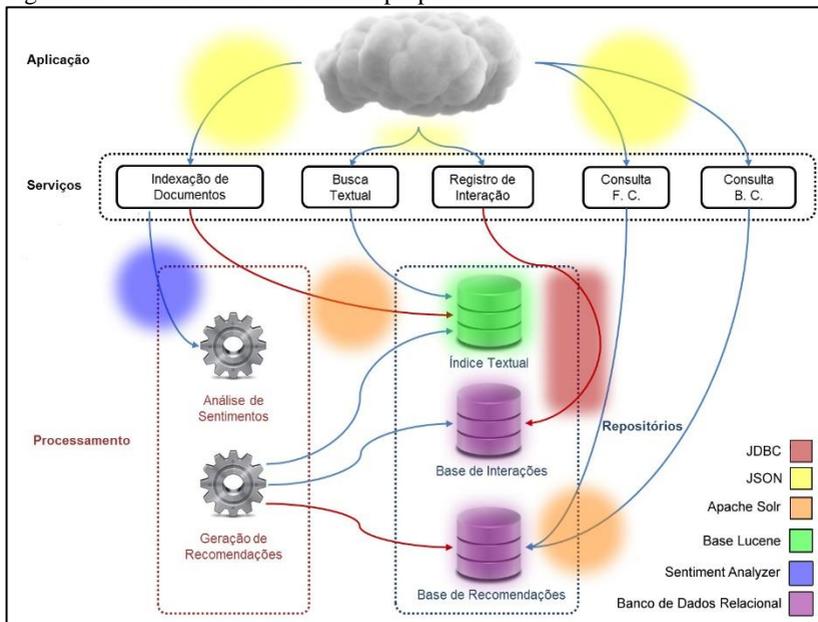
Focando em um determinado documento, o processo de geração de recomendações executará os cálculos de similaridade relacionando os documentos. Ao final as informações são persistidas no repositório denominado base de recomendações. Esta base possuirá os dados já tratados pelo processo de geração de recomendações. Após a realização dos cálculos necessários para efetuar as associações, as informações ficam disponíveis para que a camada de serviços obtenha as recomendações através das consultas de Filtragem Colaborativa e Filtragem Baseada em Conteúdo.

5.2 MODELO FÍSICO

O modelo físico representado na Figura 11 exhibe com maiores detalhes os componentes tecnológicos que compõem o modelo e como as interações entre eles acontecem. Nesta seção tem-se como objetivo apresentar uma visão mais detalhada do modelo proposto.

Os elementos tecnológicos que compõem o modelo estão apresentados na Figura 11 e ressaltados através de colorações conforme legenda. Como ilustrado, a comunicação que ocorre entre o cliente *Web* e a camada de serviços oferecida pela aplicação é realizada através de mensagens JSON (*JavaScript Object Notation*) ou Notação de Objetos JavaScript e está representada pela cor amarela. JSON é um padrão para troca de dados que possui um formato texto e é independente de linguagem. Suas convenções são familiares às linguagens C e muitas outras, como por exemplo, C++, C#, Java, *JavaScript*, Perl, *Python*. Estas propriedades tornam JSON um formato adequado para a troca de dados (JSON.ORG, 2017).

Figura 11 - Modelo físico do sistema proposto.



Fonte: Elaborado pelo autor.

Representada pela cor lilás encontram-se as bases de dados implementadas com o banco de dados relacional PostgreSQL®. A interação entre estas bases de dados com a camada de serviços e com o processo de geração de recomendações foi implementada através da tecnologia *Java Database Connectivity* (JDBC). O JDBC proporciona a manipulação de informações através da execução de comandos escritos na linguagem SQL (ORACLE, 2017). A aplicação do JDBC é ressaltada nas áreas com a cor vermelha.

A base de dados destacada na cor verde foi desenvolvida utilizando a tecnologia Apache Lucene™, uma biblioteca de mecanismo de pesquisa em texto completo de alto desempenho e escrita inteiramente em Java®. Apache Lucene™ É uma tecnologia adequada para praticamente qualquer aplicação que exija pesquisa em texto completo, especialmente em plataformas heterogêneas (LUCENE, 2017).

A indexação de documentos na base de índice textual, assim como a consulta baseada em conteúdo são providas pela tecnologia Apache Solr™ e representada pela cor laranja. O Solr é um servidor de

pesquisa empresarial independente com uma API. Através do Solr documentos são indexados utilizando-se os formatos JSON, XML, CSV ou binário através de requisições HTTP. Ele disponibiliza consultas via HTTP GET e fornece resultados também nos formatos JSON, XML, CSV ou binários. Solr tem se apresentado como uma plataforma confiável, escalável e tolerante a falhas, fornecendo indexação distribuída, replicação e balanceamento de carga, recuperação automatizada, configuração centralizada e vários outros recursos. Solr vem sendo utilizado como base em grandes sites ao redor do mundo (SOLR, 2017).

A cor Azul identifica o processo de Análise de Sentimentos denominado como *Sentiment Analyzer*. Este processo foi implementado utilizando a linguagem Java[®]. Esta mesma linguagem também foi utilizada para o desenvolvimento do processo de geração de recomendações.

Nas seções a seguir serão abordadas com mais detalhamento os processos de Análise de Sentimentos e Recomendação de Conteúdo.

5.2.1 Processo de Análise de Sentimentos

O processo de Análise de Sentimentos nesta seção é tratado como um analisador e para sua execução é necessário que as ideias coletadas estejam estruturadas em padrão XML. Uma vez estruturadas, as ideias são indexadas na base de dados Lucene[™] utilizando a tecnologia Solr[®]. Neste ponto, as ideias são classificadas com base nos comentários relacionados a elas, identificados no XML pela *tag* <ITEM COMMENT>. O conteúdo de cada comentário é subdividido em palavras denominadas *token*. Para cada *token*, o analisador faz uma pesquisa nas bases léxicas (também chamado de *lexicon*), positiva e negativa. Uma vez identificado este *token* em uma destas bases o algoritmo incrementa uma das duas variáveis que identificam a polaridade como positiva ou negativa. Ao final, o analisador define a polaridade do comentário de acordo com a polaridade predominante descoberta através de um somatório realizado sobre polaridade dos *tokens*.

Ciente da polaridade dos comentários, o analisador estabelece agora a polaridade da ideia como um todo seguindo a predominância polar dos comentários. Ou seja, se a maioria dos comentários forem positivos estes definem uma ideia positiva. Porém, se não houver predominância de uma das polaridades (positiva ou negativa) a ideia será registrada pelo analisador como uma ideia neutra.

5.2.1.1 Analisador aprimorado

Em um segundo estágio, realizou-se um aprimoramento no analisador de sentimentos através do tratamento de negação. Este processo é invocado no momento em que os comentários pertencentes a cada ideia são subdivididos em *tokens*. A partir disso, quando um *token* é selecionado para análise e a sua polaridade é identificada, o tratamento de negação é acionado visando determinar se a polaridade já estabelecida deve ou não ser invertida.

Este tratamento trabalha com 2 parâmetros principais, sendo, janela e a direção. O parâmetro janela define o limite de palavras consideradas no processo. O parâmetro direção define se as palavras consideradas estão à direita ou à esquerda da posição do *token* que está sendo analisado.

De modo geral, a partir da definição da polaridade de determinado *token* no texto, o seu contexto é analisado utilizando a quantidade de *tokens* a direita ou a esquerda e a busca em um *lexicon* de modificadores de negação para estabelecer a polaridade definitiva. Ou seja, se determinado *token* do contexto estiver no léxicon de modificadores negativos a polaridade é invertida, positiva para negativa ou negativa para positiva.

Como exemplo, configurando a janela com tamanho de 5 e a direção como esquerda, o tratamento de negação utilizará as 5 palavras do comentário que antecedem o *token* que pertence ou ao *lexicon* positivo ou ao *lexicon* negativo. Após esta configuração de parâmetros, as 5 palavras que antecedem o *token* serão pesquisadas em um *lexicon* de modificadores negativos denominada *stoplist_negative*, contendo palavras como (*aren't, can't, cannot, couldn't, didn't, entre outras*). Caso algumas dessas palavras estejam contidas na janela, o analisador realizará a inversão da polaridade atribuída originalmente ao *token*.

5.2.2 Processo de geração das recomendações

O processo de geração de recomendações visa sugerir ao usuário documentos que sejam semelhantes a um documento específico. Este processo seleciona os documentos na base de índice textual e para cada documento são localizados os similares através da aplicação da similaridade entre vetores, transformando cada documento em um vetor. A similaridade entre os documentos (vetores) é encontrada através do cálculo do cosseno descrito na seção 2.1.1.2 deste trabalho.

Após encontrar a similaridade entre os documentos, esta informação é persistida na base de dados relacional para ser utilizada como recomendação. Através do serviço de Consulta Baseada em Conteúdo é disponibilizado ao usuário os documentos juntamente com a recomendação dos que são similares a este. O retorno é obtido através de uma busca textual, sendo que a camada de serviço fornece ao usuário um objeto JSON contendo uma lista de documentos referentes aos termos utilizados na busca.

Considerando cada documento retornado em uma busca pode-se sugerir outros documentos relacionados ao usuário. Para tal, utilizando a base de recomendações foi desenvolvido um processo interno que localiza os documentos similares a partir de algum documento de interesse. Esta ação de busca é denominada de MLT (*More Like This*) e é responsável para retornar as recomendações de cada documento.

A Figura 12 ilustra o retorno de uma busca textual por documentos que contenham a palavra *reward*. O retorno é provido em formato JSON e apresenta todos os campos que constam na base de índices. O documento é representado por uma ideia. A base de dados é formada por ideias coletadas de ambientes públicos será detalhada na seção 6.1.2.

Figura 12 - Retorno de uma busca textual.

```

1  {
2    "responseHeader": {
13   "response": {
14     "numFound": 16,
15     "start": 0,
16     "maxScore": 1.261458,
17     "docs": [
18       {
19         "id": "starbucks-269",
20         "title": [ "Push notification on app that reminds reward expiring"],
21         "idea": ["Send a push notification from the app telling that reward will expire in a day.
22         "authors": ["[Lisacmomof3]"],
23         "categories": ["[New Technology]"],
24         "keywords": "[]",
25         "comments": ["Yes please. I moped around for a week with my tail between my legs when I l
26         "positive_vote": 0,
27         "negative_vote": 0,
28         "neutral_vote": 0,
29         "polarity": "positive",
30         "type": "starbucks",
31         "search": ["push notification on app that reminds reward expiring send a push notificatio
32         "_version_": 1564934703340847000,
33         "score": 1.261458
34       }

```

Fonte: Elaborado pelo autor.

A seguir a Figura 13 ilustra o retorno representando os documentos relacionados a partir de determinado documento. Neste exemplo são apresentadas, na seção *response*, os documentos mais similares a um documento em particular identificado por “Starbucks-269”. Como retorno os campos de identificação (*id*), título (*title*), texto em si (*idea*), polaridade (*polarity*), entre outros, são apresentados.

Figura 13 - Retorno de recomendação.

```

1  {
2    "responseHeader": {
3      "numFound": 407, "start": 0, "maxScore": 0.3430384, "docs": [
4        {
5          "match": {"numFound": 1, "start": 0, "maxScore": 6.547128, "docs": [
6            {
7              "id": "starbucks-269",
8              "title": ["Push notification on app that reminds reward expiring"],
9              "idea": ["Send a push notification from the app telling that reward will expire in a day. User opts to accep"],
10             "polarity": "positive",
11             "search": ["push notification on app that reminds reward expiring send a push notification from the app tell"],
12             "score": 6.547128}],
13           },
14         "response": {"numFound": 407, "start": 0, "maxScore": 0.3430384, "docs": [
15           {
16             "id": "starbucks-01",
17             "title": ["push notifications"],
18             "idea": ["I don't check my email often and I use the Starbucks app regularly. I think push notifications for"],
19             "polarity": "negative",
20             "search": ["push notifications i dont check my email often and i use the starbucks app regularly i think pus"],
21             "score": 0.3430384},
22           {
23             "id": "starbucks-248",
24             "title": ["Enter Star Codes on the App e Make it EASY and SECURE! (and more)"],
25             "idea": ["I frequently get my coffee at Starbucks locations but I get it from the grocery store every week a"],
26             "polarity": "neutral",
27             "search": ["enter star codes on the app e make it easy and secure and more i frequently get my coffee at sta"],
28             "score": 0.16135012},
29           }
30         ]
31       }
32     }
33   }

```

Fonte: Elaborado pelo autor.

Como pode ser observado na figura acima o documento principal identificado por “starbucks-269” possui polaridade positiva e está relacionado a outros documentos que também possuem suas polaridades. Ao se apresentar os documentos de origem e as recomendações de documentos relacionados com as respectivas polaridades espera-se prover informações adicionais que auxiliem no processo de tomada de decisão.

6 APLICAÇÃO DO MODELO E ANÁLISE DOS RESULTADOS

Neste capítulo apresentam-se o cenário de aplicação e os resultados mais relevantes obtidos pelo modelo. Serão exibidos detalhadamente os processos de análise de sentimentos e a geração das recomendações baseados em conteúdo apresentando exemplos destes processos para auxiliar a compreensão.

Os passos que envolvem a implementação do modelo proposto através de um protótipo são: desenvolvimento do classificador de sentimentos, aplicação do classificador de sentimentos sobre o conjunto de dados de teste, indexação dos arquivos que compõem o conjunto de dados de teste, e finalmente a geração das recomendações. Sendo assim, a aplicação destas técnicas em um modelo busca oferecer um ferramental que promova suporte à recomendação de conteúdo.

Buscando promover a inserção do modelo de maneira interdisciplinar adotou-se como cenário de aplicação a área de Gestão de Ideias, uma vez que as ideias são a matéria-prima para o processo de criação de novos produtos (COOPER; EDGETT; KLEINSCHMIDT, 2002) e a filtragem de ideias é um ponto crítico do processo de inovação, sendo determinante para a falha ou sucesso da organização (ADAMS, BESSANT, PHELPS, 2006). Porém, selecionar ideias inovadoras é um processo complexo, pois estas ideias podem ser encontradas dentro e fora da organização (MCGRATH, 2004). Sendo assim, a Gestão de Ideias é parte fundamental no processo de inovação (BARBIERI, ÁLVARES, CAJAZEIRA, 2009). Portanto, torna-se de grande valia para as organizações a existência de sistemas de recomendação de ideias para a inovação (LEE et al., 2013).

A base de dados para a aplicação do modelo foi construída utilizando o conjunto de ideias e seus comentários coletados nos portais da Starbucks®. A escolha assegurada em outros trabalhos com alguma semelhança que utilizaram uma base de dados semelhante, como por exemplo, Lee et al. (2013), Sérgio (2016) e Westerski, Dalamagas e Iglesias (2013).

6.1 PROJETO DOS EXPERIMENTOS

Os experimentos abordados neste capítulo serão compostos por duas etapas. A etapa inicial será dedicada à avaliação da Análise de Sentimentos. A segunda etapa, será voltada a fusão da Análise de

Sentimentos com Sistema de Recomendação para suporte à Gestão de Ideias.

6.1.1 Análise de Sentimentos

A primeira etapa do experimento aborda o domínio da Análise de Sentimentos. Inicialmente o experimento foi submetido a uma fase de testes buscando a avaliação do mesmo. Para realizar a avaliação, utilizou-se 1443 textos (*reviews* de usuários) relacionados a um conjunto de dados de câmeras digitais e 2000 textos (*reviews* de usuários) relacionados a um conjunto de dados de DVDs (filmes). Estes textos foram extraídos da Amazon® e fazem parte de um conjunto de dados denominado *Multi-Domain Sentiment Dataset*¹ (versão 2.0). Os textos relacionados às câmeras digitais estão em “unprocessed.tar.gz\sorted_data\camera_>photo” e os textos direcionados a DVDs em “unprocessed.tar.gz\sorted_data\dvd”.

Os textos foram classificados (polarizados) como positivos e negativos com base no *rating* identificado através de uma escala que vai de uma a cinco estrelas. Onde, classificações contendo até duas estrelas foram caracterizadas como negativas e a partir de quatro estrelas classificadas como positivas. Os textos foram então agrupados em dois arquivos: *positive.review* e *negative.review*. No domínio dos DVDs, ambos arquivos possuem 1000 textos. No domínio das câmeras digitais são 726 textos positivos e 717 textos negativos. Esses textos foram extraídos e processados a partir de um documento textual em formato XML como ilustrado na figura 14 a seguir:

¹ Saiba mais sobre o dataset Multi-Domain Sentiment Dataset em <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

Figura 14 - *Review* positivo de uma câmera digital extraído da Amazon®.

```

<review>
  <unique_id>B00009R6TA:great_backpack:w_mutual</unique_id>
  <unique_id>4623</unique_id>
  <asin>B00009R6TA</asin>
  <product_name> Canon Deluxe Photo Backpack 200EG for Canon EOS SLR Cameras: Camera & Photo</product_name>
  <product_type>camera & photo</product_type>
  <helpful>1 of 1</helpful>
  <rating>5.0</rating>
  <title>great backpack</title>
  <date>November 12, 2006</date>
  <reviewer>W. Mutual</reviewer>
  <reviewer_location>California</reviewer_location>
  <review_text>
    This is a great back pack! If there is anything I would change it would be the logo on the front.
    It's brite and shinny, and I'd rather not advertise that I have a camera in there!
    I colored it in with a felt pen and it works great..... Overall, I love the backpack!!
  </review_text>
</review>

```

Fonte: *Multi-Domain Sentiment Dataset*.

Uma vez avaliado o experimento foi aplicado sobre uma base de dados descrita na seção 6.1.2.

Para a realização da análise de sentimentos foi efetuado o processo de polarização dos adjetivos e advérbios extraídos dos textos contidos em cada conjunto de dados. O processo de polarização é efetuado através do uso de um léxico de sentimentos. Para viabilizar esta etapa, foi utilizado o léxico apresentado por Hu e Liu (2004) disponível em “<https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>”. Este léxico é composto por 6800 palavras distintas e classificadas de acordo com sua polaridade (negativo/positivo) conforme exemplo ilustrado na Figura 15. Diversos trabalhos têm se utilizado deste léxico, entre eles, Zhang e Liu (2011), Dong et al. (2013), Ceci et al. (2015). Entre as razões para a utilização de léxicos na implementação de analisadores de sentimento reside a sua flexibilidade nos mais diversos domínios e a não dependência de conjuntos de dados de treinamento.

Figura 15 - Léxico com suas colunas de palavras polarizadas.

Positive Words	Negative Words
◆ +beautiful	◆ +bad
◆ +good	◆ +blurry
◆ +great	◆ +confusing
◆ +nice	◆ +conned
◆ +surprise	◆ +fragile
◆ +usefull	◆ +poorly
◆ -colorful	◆ +slow
◆ -enhance	◆ +stopped
◆ -fairly	◆ +terrible
◆ -hallmark	◆ -cheap
◆ -heros	◆ -concert
◆ -like	◆ -crushed
◆ -mercy	◆ -dark
◆ -patience	◆ -drained
◆ -patient	◆ -dumping
◆ -portable	◆ -dust
◆ -protect	◆ -flare
◆ -protective	◆ -hard
◆ -refund	◆ -indistinguishable
◆ -refunded	◆ -intense
◆ -respect	◆ -limit
◆ -respectable	◆ -loud
◆ -respectful	◆ -mad
◆ -respectfully	◆ -pricey
◆ -super	◆ -resistance
◆ -work	◆ -ruin
◆ -worked	◆ -scam
◆ -works	◆ -sorry

Fonte: Elaborado pelo autor.

6.1.2 Bases de ideias

O cenário de avaliação do modelo proposto foi composto por um conjunto de dados com 514 (quinhentas e quatorze) ideias disponíveis no portal da empresa Starbucks®. Este conjunto de dados foi escolhido seguindo um modelo de estruturação no formato em que as ideias estão dispostas e também por ser utilizada em outros trabalhos relacionados a Sistemas de Recomendação e Análise de Sentimentos, cita-se como exemplo o trabalho de Lee et al., (2013). As ideias que compõem o conjunto de dados foram coletadas no período de outubro de 2014 a março de 2015 e complementadas em março de 2017.

A Starbucks® é uma empresa multinacional fundada em 1971, sendo a maior rede de cafeterias do mundo, hoje sediada em Seattle, EUA. Em Junho de 2015 a empresa atingiu a marca de 22,519 lojas em mais de 50 países, tornando-se mundialmente conhecida não só pela

venda de café, mas também pelo relacionamento diferenciado com seus clientes (STARBUCKS, 2016).

Buscando facilitar a manipulação das ideias através de meios computacionais encontrou-se a necessidade da estruturação dos elementos. Deste modo, cada ideia coletada manualmente foi estruturada em um documento no formato XML. A Figura 16 a seguir ilustra a estrutura de uma ideia coletada para o cenário de estudo a partir do site da empresa Starbucks®.

Figura 16 - Ideia estruturada no formato XML.

```

<DOCUMENT
  ID = ""
  TITLE = "EXPANDED HEALTHY SMOOTHIES"
  DAY = "19"
  MONTH = "06"
  YEAR = "2008"
  IDEA = "ONE STOP BEVERAGE STOP FOR THE ENTIRE FAMILY...If you carried a minimal selection
of the most common frozen fruits for smoothies (strawberry, banana, pineapple & blueberries).
The current line appears NOT NATURAL. This would widen the scope of patrons. However, in addition
offering of a minimal variety of protein options would even more substantially open the door!!!
This would attract the healthy gym crowd, breakfast alternative crowd... wide selection of
friends meeting... pretty endless...Healthy, low fat, natural sugar, protein... to me it's a no
brainer and an easy addition to your growing product line!" >
  <AUTHORS><ITEM NAME = "Kathy G"/></AUTHORS>
  <CATEGORIES><ITEM DESCRIPTION = ""/></CATEGORIES>
  <KEYWORDS></KEYWORDS>
  <VOTES AFFIRMATIVE = "0" NEGATIVE = "110" NEUTRAL = "0"/>
  <COMMENTS>
    <ITEM COMMENT = "I think we'll be carrying fresh bananas but I don't know much about the new drink yet.
And there's another drink with mango? But I don't think that's fresh fruit.. but these
drinks are more like a smoothie but I don't know how much protein or it's health value.
I just saw in a binder.. but i agree and i voted for this.It'd be so much healthier than
drinking a blend that's full of sugar." NAME = "ohitsai" RATING = ""/>
    <ITEM COMMENT = "Fruit & Veggie Smoothies are now available in Starbucks stores in Northern California,
Oregon, Washington, Alaska, select Idaho stores, and nationwide in select Target, Albertsons,
Safeway, and Kroger Starbucks stores." NAME = "sbx_bean" RATING = ""/>
  </COMMENTS>
</DOCUMENT>

```

Fonte: Elaborado pelo autor.

A estrutura do XML ilustrada acima demonstra as características dos elementos que compõem a base de dados. Assim todos os elementos serão compostos pelas seguintes características representadas pelas *tags* XML: Identificador (ID), Título (TITLE), Mês (MONTH), Dia (DAY) e Ano (YEAR) construindo a data da postagem da ideia. O corpo da ideia é representado pela tag IDEA. O autor da ideia é representado por AUTHOR e a categoria ao qual ela se enquadra identificada pela tag CATEGORIES. VOTES informam a avaliação dos leitores, sendo exibida apenas o saldo dos votos. No exemplo ilustrado acima a ideia possui um saldo negativo de 110 votos. Na Seção COMMENTS do XML são relacionados os comentários dos leitores a respeito da ideia, sendo esta seção que terá o conteúdo avaliado pelo analisador de sentimentos.

6.1.3 Métricas de Avaliação

As métricas de avaliação aplicadas neste trabalho foram baseadas nos trabalhos de Hu e Liu (2004), Kazama e Tsujii (2005), Zhang e Liu (2011), Moreo et al. (2012), Cruz et al. (2013), Moraes et al. (2013) e Li e Tsai (2013). As medidas utilizadas para a avaliação foram: acurácia (*accuracy*), precisão (*precision*), revocação (*recall*) e medida F (*f-measure*). O cálculo destas medidas é realizado com o uso de uma matriz de contingência, ilustrada com mais detalhes na Tabela 3.

Tabela 3 - Matriz de contingência.

		Classificação esperada (Correta)	
		Positivo	Negativo
Classificação Retornada	Positivo	TP (Positivo Verdadeiro)	FP (Falso Positivo)
	Negativo	FN (Falso Negativo)	TN (Negativo Verdadeiro)

Fonte: Elaborado pelo autor.

A tabela acima possui os seguintes elementos:

- TP (Positivo Verdadeiro): Casos positivos classificados corretamente;
- TN (Negativo Verdadeiro): Casos negativos classificados corretamente;
- FP (Falso Positivo): Casos positivos classificados de maneira errada, sendo na realidade negativos;
- FN (Falso Negativo) Casos negativos classificados de maneira errada, sendo na realidade positivos.

A equação abaixo é utilizada para encontrar a acurácia:

$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN}$$

Para obter a precisão aplica-se a seguinte equação:

$$Precisão = \frac{TP}{TP + FP}$$

A equação a seguir é utilizada para encontrar a revocação:

$$\text{Revocação} = \frac{TP}{TP + FN}$$

Para encontrar a medida F, aplica-se a equação abaixo:

$$\text{Medida } F = 2 \cdot \frac{\text{precisão} \cdot \text{revocação}}{\text{precisão} + \text{revocação}}$$

6.1.3.1 Significância estatística e métricas de avaliação

Para aumentar a precisão na comparação entre duas amostras correlacionadas, como por exemplo, a evolução de um modelo, adotou-se o método de McNemar. Este método pode aferir a significância estatística entre mudanças comparando os estados anterior e posterior de uma amostra (MCNEMAR, 1947).

Destaca-se no contexto deste trabalho o entendimento de que dois modelos possuem significância estatística, porém, não ocorrem de maneira semelhante e seus resultados não são produzidos ao acaso. Para tal, deve-se definir a hipótese nula que declara que dois modelos (os modelos no contexto do trabalho representam classificadores) executam de maneira equivalente. Ao rejeitar a hipótese nula promove-se a confiança que as observações/afereções dos modelos não executam meramente ao acaso, ou seja, pode-se rejeitar a hipótese nula que dois modelos executam de maneira equivalente.

Para realizar o cálculo de McNemar com base na comparação entre dois modelos (f_1 e f_2), torna-se necessário a coleta das variáveis definidas abaixo (as variáveis c_{00} e c_{11} não fazem parte da equação):

- c_{00} : Número de classificações erradas realizadas por f_1 e f_2 ;
- c_{01} : Número de classificações erradas realizadas por f_1 , mas classificadas corretamente por f_2 ;
- c_{10} : Número de classificações erradas realizadas por f_2 , mas classificadas corretamente por f_1 ;
- c_{11} : Número de classificações corretas realizadas por f_1 e f_2 .

Baseado nesta definição obtém-se o teste de McNemar através da equação abaixo, onde $c_{01} + c_{10} \geq 20$.

$$x^2 = \frac{(|c_{01} - c_{10}| - 1)^2}{c_{01} + c_{10}}$$

Para interpretar o teste de McNemar é utilizada a tabela *Chi-Square* padrão. Neste estudo é utilizada apenas a primeira linha da tabela de contingência para obter um grau de liberdade, pois se trata do uso de uma classificação binária (positivo ou negativo). Em um grau de confiança de 95% onde ($p < 0,05$), é considerado o valor superior a 3,841, ou seja, resultados acima deste valor possuem significância estatística.

O método de McNemar foi aplicado com o objetivo de comparar um dos componentes do modelo proposto (analisador de sentimento) em relação outros modelos denominados classificadores estatísticos, verificando a significância estatística entre as execuções.

Diante do teste de McNemar e das métricas demonstradas aqui, realiza-se a avaliação e a viabilidade do analisador de sentimento desenvolvido.

6.1.3.2 Similaridades entre vetores

Para identificar as ideias similares, neste trabalho foi aplicada a similaridade entre vetores, onde, cada ideia inicialmente estruturada em um arquivo XML será transformada em um vetor. Uma vez realizado este processo, a similaridade entre as ideias (vetores), será obtida através do cálculo do cosseno descrito na seção 2.1.1.2 deste trabalho. A similaridade entre os vetores permite apresentar ideias relacionadas a partir de uma determinada ideia, fornecendo assim, na visão proposta pelo modelo subsídios para auxiliar na tomada de decisão. No contexto do cenário de aplicação deste trabalho, ou seja, a Gestão de Ideias, a tomada de decisão refere-se à implementação/desenvolvimento ou não de determinada ideia.

6.2 AVALIAÇÃO DOS EXPERIMENTOS

Nesta seção será abordada em mais detalhes a avaliação dos processos de análise de sentimentos e recomendações baseadas em conteúdo.

6.2.1 Análise de Sentimento

Visando testar o analisador de sentimentos são conduzidos quatro testes em cada um dos conjuntos de dados utilizados neste trabalho, *reviews* de câmeras e DVDs. O primeiro teste apresenta os resultados do analisador básico (chamado de Base) e o segundo teste apresenta os resultados do analisador básico acrescido do tratamento de negação. Já o terceiro teste compara o analisador com tratamento de negação e métodos estatísticos amplamente utilizado em tarefas de classificação, *Support Vector Machine* (SVM) e *Naïve Bayes* (NB). Por fim, o último teste analisa a significância estatística das duas versões do analisador de sentimento desenvolvido no trabalho em relação aos métodos estatísticos. O analisar de sentimentos foi detalhado na seção 5.2.1.

6.2.1.1 Conjunto de Dados de Câmeras

A primeira análise apresenta os resultados do analisador base considerando as variáveis da tabela de contingência (TP, TN, FP, FN) que possibilitam o cálculo da acuraria (Acu.), precisão (Prec.), revocação (Rec.) e medida F (F). Pode-se notar na Tabela 4 que o analisador base atingiu uma acurácia de 0,75 e um índice agregado (medida F) de 0,79.

Como pode ser observado o analisador básico classifica 332 documentos como positivo que deveriam ser negativos (FP), enquanto classifica somente 34 documentos como negativo que deveriam ser positivos (FN). Isso reduz a precisão uma vez que esta estabelece uma relação entre os positivos verdadeiros (TP) e os falsos positivos (FP). Por outro lado, a revocação é beneficiada uma vez que esta é uma relação entre os positivos verdadeiros (TP) e os falsos negativos (FN). Para um analisador simples baseado em tabelas léxicas o resultado agregado é compatível com o encontrado na literatura, por exemplo, no trabalho de Ceci et al. (2015).

Tabela 4 - Tabela com os testes iniciais utilizando o analisador base.

Método	TP	TN	FP	FN	Acu.	Prec.	Rec.	F
Base	692	385	332	34	0,75	0,68	0,95	0,79

Fonte: Elaborado pelo autor.

A segunda análise apresenta os resultados do analisador base incrementado com o tratamento de negação considerando as variáveis da tabela de contingência (TP, TN, FP, FN). Este analisador atingiu um

índice de 0,79 de acurácia e um índice agregado de 0,82 conforme a Tabela 5. O tratamento de negação, conforme explicado na seção 5.2.1.1, se utiliza de uma janela ao redor de determinado termo polar (termo presente nos léxicos de sentimento) com o intuito de localizar modificadores negativos. Caso seja encontra um modificador na janela o termo polar tem a sua polaridade invertida.

Pode-se verificar que nesta implementação ocorre menos classificações FPs mantendo praticamente igual os FNs. Isso possui impacto positivo na acurácia e na precisão refletindo no índice agregado. Este comportamento é esperado uma vez que o tratamento de negação objetiva reconhecer contextos positivos que sofrem a interferência de modificadores negativos, ou seja, o usuário utiliza termos positivos, mas anteriormente efetua a negação.

Tabela 5 - Tabela com os testes iniciais utilizando o analisador aprimorado.

Método	TP	TN	FP	FN	Acu.	Prec.	Rec.	F
Base + Negação	693	443	274	33	0,79	0,72	0,95	0,82

Fonte: Elaborado pelo autor.

Nesta próxima análise objetiva-se comparar de maneira geral o analisador com tratamento de negação em relação a métodos estatísticos utilizados tradicionalmente em tarefas de classificação, SVM e NB. Conforme apresentado na Tabela 6 o analisador desenvolvido tem uma acurácia menor em relação aos classificadores estatísticos e uma pequena diferença no índice agregado (medida F).

Apesar de o analisador com tratamento de negação classificar de maneira equivocada FPs com maior intensidade, isso é compensado pelas poucas classificações como FNs. Por outro lado, SVM e NB possuem uma tendência maior de classificarem documento como FN. Neste sentido, pode-se afirmar que melhorias no analisador de sentimentos desenvolvido no trabalho permitiriam a redução de classificações FPs impactando na acurácia total.

Tabela 6 - Tabela comparando o aprimorado com métodos estatísticos.

Método	TP	TN	FP	FN	Acu.	Prec.	Rec.	F
Base + Negação	693	443	274	33	0,79	0,72	0,95	0,82
SVM	594	628	89	132	0,85	0,87	0,82	0,84
NB	620	600	117	106	0,85	0,84	0,85	0,85

Fonte: Elaborado pelo autor.

A última análise do conjunto de dados de câmara trata a significância estatística entre os analisadores desenvolvidos e métodos estatísticos tradicionais. Como pode ser observado em todas as situações, o valor calculado de McNemar ficou acima de 3,841, indicando que existe significância estatística entre os métodos testados, ou seja, pode-se rejeitar a hipótese nula de que os métodos executam de maneira equivalente.

Como pode ser observado na Tabela 7, os Métodos f_1 classificam de forma errada mais elementos. Conforme analisado anteriormente isso é reflexo das classificações de documentos como FPs representando mais que o dobro em relação aos métodos estatísticos.

Tabela 7 - Significância estatística entre o método proposto e os demais.

Método f_1	Método f_2	c01	c10	McNemar
Base	NB	289	146	46,35402
Base	SVM	297	152	46,18263
Base + Negação	NB	239	155	17,48477
Base + Negação	SVM	248	162	17,62195

Fonte: Elaborado pelo autor.

6.2.1.2 Conjunto de Dados de Filmes

A primeira análise deste conjunto de dados apresenta os resultados do analisador base considerando as variáveis da tabela de contingência (TP, TN, FP, FN). Na Tabela 8 pode-se verificar que a acurácia atingida foi de 0,76 e a medida F de 0,79.

Como pode ser observado o analisador básico classifica 337 documentos como positivo que deveriam ser negativos (FP) e 133 documentos como negativos que deveriam ser positivos (FN). Como existe uma diferença menor entre FP e FN quando comparado ao conjunto de dados de câmeras os índices relativos de precisão e revocação ficam mais próximos. Verifica-se que analisador base possui uma deficiência na identificação de casos negativos. Na linguagem natural é comum a utilização de palavras com polaridade positiva que possuem modificadores negativos que as antecedem. Nesta análise os resultados também são compatíveis com trabalhos encontrados na literatura.

Tabela 8 - Tabela com os testes iniciais utilizando o analisador base.

Método	TP	TN	FP	FN	Acu.	Prec.	Rec.	F
Base	863	658	337	133	0,76	0,72	0,87	0,79

Fonte: Elaborado pelo autor.

A segunda análise envolve a avaliação do analisador base incrementado com o tratamento de negação também considerando as variáveis da tabela de contingência (TP, TN, FP, FN).

O analisador atingiu um índice de 0,76 de acurácia e um índice agregado de 0,79 conforme consta na Tabela 9. Nota-se que a acurácia é ligeiramente superior ao método base uma vez que menos classificações erradas foram produzidas em FP e FN. Pode-se verificar que neste caso a precisão do analisador é inferior quando comparado ao conjunto de dados de câmera. Isto é justificável uma vez que no caso do conjunto de dados de câmera, os textos (reviews) referem-se a aspectos mais facilmente mensuráveis, por exemplo, “o foco de determinada câmera é ruim” ou “possui um bom zoom, mas o enquadramento não funciona adequadamente”. Por outro lado, textos (reviews) no conjunto de dados de DVD (filmes) referem-se a contextos mais subjetivos, como por exemplo, o gosto por determinado gênero de filme ou a simpatia/antipatia por determinado ator/atriz ou diretor/diretora.

Tabela 9 - Tabela com os testes iniciais utilizando o analisador aprimorado.

Método	TP	TN	FP	FN	Acu.	Prec.	Rec.	F
Base + Negação	851	689	306	145	0,77	0,74	0,85	0,79

Fonte: Elaborado pelo autor.

A análise da Tabela 10 compara o analisador base com tratamento de negação com os classificadores estatísticos SVM e NB. Como pode ser observado o nível de acurácia está próximo dos classificadores estatísticos enquanto que o índice agregado é o mesmo. A precisão é um pouco menor em função de um número maior de erros na classificação em caso negativos, ou seja, os falsos positivos (FPs).

De modo geral, os classificadores possuem um comportamento similar indicando que o analisador proposto, ainda que passível de melhorias, atinge valores próximos dos obtidos por métodos amplamente utilizados em tarefas de classificação.

Tabela 10 - Tabela comparando o analisador aprimorado com métodos estatísticos.

Método	TP	TN	FP	FN	Acu.	Prec.	Rec.	F
Base + Negação	851	689	306	145	0,77	0,74	0,85	0,79
SVM	785	779	216	211	0,79	0,78	0,79	0,79
NB	822	744	251	174	0,79	0,77	0,83	0,79

Fonte: Elaborado pelo autor.

A última análise do conjunto de dados de filme, assim como na análise cenário de câmera, envolve a verificação da significância estatística entre os analisadores desenvolvidos e métodos estatísticos tradicionais. Como pode ser observado na Tabela 11 em todas as situações o valor calculado de McNemar ficou abaixo de 3,841, indicando que não existe significância estatística entre os métodos testados, ou seja, pode-se aceitar a hipótese nula de que os métodos executam de maneira equivalente.

Esta situação é pouco esperada, uma vez que as abordagens são significativamente diferentes. Contudo, a falta de significância pode ser explicada pela proximidade dos erros e acertos de cada um dos métodos f_1 e f_2 . Por exemplo, na comparação do analisador proposto (base + negação) contra o SVM percebe-se uma proximidade entre as variáveis $c01$ e $c10$, em que $c01$ (308) indica a quantidade de erros de f_1 e a quantidade de acertos de f_2 . Já a variável $c10$ (284) indica a quantidade de erros de f_2 e a quantidade de acertos de f_1 . Esta maior proximidade também pode ser observada nos valores de FPs e FNs apresentados na Tabela 10.

Tabela 11 - Significância estatística entre o método proposto e os demais.

Método f_1	Método f_2	c01	c10	McNemar
Base	NB	307	262	3,40246
Base	SVM	312	269	3,036145
Base + Negação	NB	300	274	1,08885
Base + Negação	SVM	308	284	0,893581

Fonte: Elaborado pelo autor.

6.2.1.3 Considerações sobre as Análises dos Conjuntos de dados

Os resultados apresentados na avaliação do analisador de sentimentos desenvolvido indicam que a sua utilização é adequada no contexto deste trabalho. Isto se justifica uma vez que não existe um conjunto de dados de gestão de ideias com a indicação da polaridade de cada comentário e da ideia como um todo.

Métodos estatísticos como o SVM e NB pressupõem uma fase de treinamento para que um modelo seja gerado e utilizado quando uma nova instância é apresentada para ser classificada. Pressupõem também a necessidade, em algum momento, de uma readequação do modelo gerado para incorporar novos padrões. Ou seja, será necessária uma nova fase de treinamento no futuro.

Considerando a base de dados utilizada no processo de recomendação voltada ao cenário de gestão de ideias, esta não possui um conjunto de dados para treinamento. Ainda que existisse um conjunto de treinamento, ideias são altamente dinâmicas e impactadas pelo aprimoramento do conhecimento das pessoas que submetem ou comentam determinada ideia.

Partindo deste pressuposto, faz-se necessária a utilização de analisadores *ad-hoc*, ou seja, o desenvolvimento de classificadores utilizando estruturas de conhecimento (*lexicons*) que permitam a polarização (classificação em positivo ou negativo) de um texto sem prévio treinamento. A deficiência desta abordagem reside na necessidade de melhorias dos *lexicons* ou implementação de novas regras e comportamentos em função da evolução da linguagem natural, por exemplo, a análise de *emojis* e *emoticons*. Contudo, isso pode ser gradativamente desenvolvido e incorporado ao analisador/classificador.

6.2.2 Recomendações Baseadas em Conteúdo

Com o objetivo de avaliar a integração da análise de sentimentos com os sistemas de recomendação são apresentados e discutidos exemplos no domínio da Gestão de Ideias.

A geração de recomendações baseadas em conteúdo sugere ao usuário ideias semelhantes a partir de uma ideia de interesse. Pensando no fluxo de execução, inicialmente o usuário efetua uma busca utilizando palavras-chave, similar ao realizado através de mecanismos de busca. Como resultado, são apresentadas as ideias mais similares à consulta realizada.

Em seguida, o usuário pode obter um conjunto de ideias que são recomendadas pelo modelo levando em consideração a similaridade entre elas. Ou seja, a partir de uma ideia escolhida pelo usuário serão recomendadas a ele ideias semelhantes à ideia escolhida. Para definir a similaridade entre as ideias é aplicado o cálculo do cosseno apresentado na seção 2.1.1.2 deste trabalho.

O Quadro 1 exibe os resultados de uma consulta buscando pela palavra *Beverages*. Pode-se notar que no retorno são apresentadas as ideias compostas pelo título e o texto das mesmas, assim como, a polaridade calculada pelo analisador de sentimento e mantida na base de índices. No contexto desta avaliação a polaridade é o elemento essencial para auxiliar na análise das ideias por determinada organização. As polaridades da ideia em análise e de suas ideias relacionadas, pode fornecer insumos adicionais durante o processo de análise.

Quadro 1 - Exemplo de busca por ideias que contenham a palavra *Beverages*.

Título	Polaridade
Texto	
<i>Bring back Mocha Coconut and Coconut Crème Frappuccino Blended Beverages</i>	Positiva
<i>Please bring back the Mocha Coconut Frappuccino for this summer!! This was my favorite and I haven't seen it in a very long time!!</i>	
<i>Offer a Few Regional Food and Beverages in Some Locations</i>	Positiva
<i>I live in a city surrounded by many cultures. I feel very blessed to have the exposure of many types people, languages and the best thing of all...the food and beverages! Yum! Cultural acceptance and accommodation is an important aspect to the success of any restaurant and store, it shows respect and acknowledgment of it's customers. My idea for Starbucks is to offer a few regional foods and beverages in some locations.....</i>	
<i>Tipping Baristas via Phone App</i>	Positiva
<i>I would have to think purchasing beverages via gift cards or using smart phone apps increase, likewise tips for baristas have conversely decreased.... I would like to have the ability to select a tip amount on my iPhone app that can be added to my purchase before it is scanned. I rarely carry change or cash these ...</i>	
<i>Larger boxes of V-pods</i>	Positiva
<i>Currently the boxes of pods you can buy for the verismo are relatively small, no one who uses their machine regularly can avoid picking up pods at least once a week, which is not a chore, especially if you go to sbux anyway, BUT, you sell boxes of 54 K-cups at bulk retailers like BJ's, why not sell verismo pods in the same large quantity boxes? If a family of four, for example....</i>	
<i>Ability to use Rewards to pay with Mobile ordering</i>	Positiva
<i>I used mobile ordering the other day and I have a reward I need to use but I wasn't able to use it to pay for my drink.</i>	

Fonte: Elaborado pelo autor.

Busca-se neste exemplo, conhecer mais ideias que tenham semelhança com a primeira ideia com contexto positivo. Sendo assim, encontra-se a ideia intitulada como “*Bring back Mocha Coconut and Coconut Crème Frappuccino Blended Beverages*”. O autor da ideia solicita que seja comercializada novamente uma bebida que este considera a melhor bebida. Neste caso especificamente, 31 membros da comunidade comentaram esta ideia, todos sem exceção concordavam com o autor, e se demonstravam esperançosos que este pedido fosse atendido pela Starbucks®.

Dentre estes comentários cita-se como exemplo 5 casos:

- *“Oh my god! Yes. This is my favorite. I was deeply saddened when you took it away.... Wahhh”*
- *“PLEASE PLEASE bring back the Coconut Creme Frappuccino! This is my all time Favorite and I'm very disappointed it hasn't been offered the last couple summers!!!!”*
- *“PLEASE PLEASE PLEASE Bring back the Mocha Coconut Frappuccino. I love it soooooooooo much. Best drink ever!! Pleeeeeeease!”*
- *“I agree!!! This is one of my favorites and I just called and they said they weren't bringing it back...soooo sad! Coming from a super loyal regular, BRING IT BACK!!!”*
- *“Yes!!! Please Bring it back, and this time let's have all year round since you made us suffer this summer.”*

Visto que a ideia obteve aprovação unanime da comunidade, busca-se agora saber se existem semelhantes a esta. Assim, Através da funcionalidade *More Like This (MLR)* implementada, solicita-se ao protótipo que disponibilize mais ideias que sejam semelhantes à ideia “*Bring back Mocha Coconut and Coconut Crème Frappuccino Blended Beverages*”.

Atendendo a esta solicitação o modelo recomenda as ideias por nível decrescente de similaridade conforme demonstradas no Quadro 2 abaixo:

Quadro 2 - Exemplo de recomendação de ideias com base em uma ideia de interesse.

Título	Polaridade
Texto	

<i>Bring back Mocha Coconut and Coconut Crème Frappuccino Blended Beverages</i>	Positiva
<i>Please bring back the Mocha Coconut Frappuccino for this summer!! This was my favorite and I haven't seen it in a very long time!!</i>	
Ideias Recomendadas	
Título	Polaridade
Texto	
<i>Mocha Coconut Frappuccino</i>	Positiva
<i>Bring back the mocha coconut frappuccino!!!!!! Please!</i>	
<i>MOCHA COCONUT</i>	Positiva
<i>Please bring back the Mocha Coconut in summer 2014, AND consider making it a year round drink!</i>	
<i>MOCHA COCONUT FRAPPUCCINO</i>	Positiva
<i>BRING BACK THE MOCHA COCONUT FRAPPUCCINOS PLEASE!!!!</i>	
<i>Bring Back the Mocha Coconut Frappuccino</i>	Positiva
<i>Summer lovin... the sweet scent of coconut in the air!! Summer isn't the same without it. Where is it?? This should be a menu staple. Please bring this hit back. Don't leave it in the One Hit Wonder pile. Christine Sipin</i>	
<i>Coconut syrup</i>	Positiva
<i>Please bring back coconut early and keep it year round it's my favorite!!! I've been dying for a coconut mocha frap</i>	
<i>Where is the Mocha Coconut Frappuccino?</i>	Positiva
<i>Literally, I'm crying over here. Mocha Coconut Frappuccino has been my favorite drink that Starbucks has ever had! I do not understand why it was discontinued. There's NO coconut syrup in the stores at all actually. This is so disappointing, especially because that's a pretty popular flavor for iced coffee. So upset about this... and I just started my Rewards card because I knew I would rack up the stars with coconut iced coffee and fraps. So so sad.</i>	
<i>Popular Frappuccino return</i>	Positiva
<i>Bring back the ever popular Mocha Coconut Frappuccino - it was the MAIN reason 5 of my family members would go to Starbucks together!!! Nothing can match this Frappe!!!!!!</i>	
<i>MOCHA COCONUT FRAPPUCCINO</i>	Negativa
<i>PLEASE BRING BACK THE MOCHA COCONUT FRAPPUCCINO...ITS SEEMS PRETTY CLEAR ITS WHAT EVERYONE WANTS...DON'T KNOW WHY YOU CAN'T LISTEN!</i>	
<i>mocha coconut</i>	Neutro
<i>Bring the mocha coconut back for good.</i>	

Fonte: Elaborado pelo autor.

Seguindo a busca por ideia, neste segundo exemplo, é realizada uma busca por ideias relacionadas ao termo *Holiday*.

Quadro 3 - Exemplo de busca por ideias que contenham a palavra *Holiday*.

Título	Polaridade
---------------	-------------------

Texto	
<i>Reusable RED holiday cups</i>	Positiva
<i>\$1 reusable holiday cups. It means the holidays are coming, and so any would like to enjoy our own coffee in the red cup during the season and reuse them each time we come in.</i>	
<i>Reusable white plastic Starbucks cups. How about a RED REUSABLE CUP during the xmas holiday?!</i>	Positiva
<i>Reusable white plastic Starbucks cups. How about a RED REUSABLE CUP during the xmas holiday?</i>	
<i>\$1 reusable cup</i>	Positiva
<i>During the holiday season when you change your regular cups to red with a holiday design-- create and sell the reusable cup to match that design!!!</i>	
<i>Holiday cups Reuseable</i>	Positiva
<i>I have loved using your re useable cups. I love drinking out of them. However, since the Holiday season is coming around maybe try to make colored cups that match what the cups you have during Christmas. It's the little things ...</i>	
<i>Reusable red cups</i>	Positiva
<i>Every year during the holiday season, everyone looks forward to the Starbucks red cups. Currently they have reusable cups for \$1 but it is not being utilized and it is not really promoted by Starbucks for customers to purchase those reusable cups. I think if during the holiday season, if Starbucks have a limited edition ...</i>	

Fonte: Elaborado pelo autor.

Assim como no caso anterior, busca-se ideias semelhantes a primeira ideia positiva retornada pela busca denominada “*Reusable RED holiday cups*” conforme apresentada pelo quadro acima. Neste caso, o autor sugere que copos criados para edições comemorativas sejam comercializados. Esta ideia originou 20 comentários todos positivos. Dentre estes comentários cita-se 5 exemplos:

- “*I was going to submit this very same idea! I love the red cup during the holidays*”
- “*I just submitted the same idea before looking!! Love it and would buy a few :)*”
- “*I am collecting holiday cups every year and I would love to have a reusable holiday cup!!!*”
- “*I was just submitting the same idea. I think these would be a huge hit!*”
- “*We appreciate the way you are all thinking!! And surprisingly a few of us are already focusing on Christmas/Holiday for next year!*”

Buscando ideias semelhantes a esta, obtêm-se as recomendações contidas no Quadro 4. Em sua maioria, as recomendações relacionadas são positivas. É possível localizar uma ideia negativa em que no comentário o usuário efetua reclamações. Contudo, apesar da ideia ser relacionada como uma ideia negativa esta é pouco similar. A similaridade neste caso poderia ser utilizada para restringir a apresentação de ideias similares, provendo assim um limiar de corte.

Quadro 4 - Exemplo de recomendação de ideias com base na ideia eleita pelo usuário.

Título	Polaridade
Texto	
<i>Reusable RED holiday cups</i>	Positiva
<i>\$1 reusable holiday cups. It means the holidays are coming, and so any would like to enjoy our own coffee in the red cup during the season and reuse them each time we come in.</i>	
Ideias Recomendadas	
Título	Polaridade
Texto	
<i>\$1 reusable cup</i>	Positiva
<i>During the holiday season when you change your regular cups to red with a holiday design-- create and sell the reusable cup to match that design!!!</i>	
<i>Reusable red cups</i>	Negativa
<i>Every year during the holiday season, everyone looks forward to the Starbucks red cups. Currently they have reusable cups for \$1 but it is not being utilized and it is not really promoted by Starbucks for customers to purchase those reusable cups....</i>	
<i>reusable white plastic StarBucks cups. How about a RED REUSABLE CUP during the xmas holiday?!</i>	Positiva
<i>reusable white plastic StarBucks cups. How about a RED REUSABLE CUP during the xmas holiday?!</i>	
<i>Holiday cups Reuseable</i>	Positiva
<i>I have loved using your re useable cups. I love drinking out of them. However, since the Holiday season is coming around maybe try to make colored cups that match what the cups you have during Christmas. It's the little things but the holiday cups always put me in a good mood. It would be great to have reusable ones to enjoy as well.</i>	
<i>Red plastic reusable cups for Holliday season</i>	Positiva
<i>I love the red cups. But don't want to use paper. We need reusable red Christmas cups!</i>	
<i>Go back to the European coffee house</i>	Positiva
<i>Please go back to your original idea of an European coffee house and get rid of the extraneous items like cds, stuffed animals, countless foods and all that factory holiday junk. I love the original Starbucks better. And Howard Schultz, I love you. -</i>	

<i>An idea from our Annual Meeting of Shareholders on March 19, 2008 in Seattle</i>	
<i>Bring Back the Coffee House Feel</i>	Positiva
<i>Starbucks started when Howard Schultz, now again the current President and CEO bought out a local coffee house in Seattle, WA. That coffee house had weathered wood floors, bags of roasted beans laying about and a general warm atmosphere. Now Starbucks has become a cold...</i>	
<i>More PERKS for Gold Card Members</i>	Negativa
<i>When I first became a Gold Card Member, Starbucks offered more benefits from being a member. Now, the only benefit we get is being able to get a free drink after 12 stars. I spend between \$5-\$6 every time I go to Starbucks, spending roughly \$72 (= \$6 x 12) before I get a reward from being a loyal customer....</i>	
<i>Flat Whites in America</i>	Positiva
<i>Can we add Flat Whites added to the menu in America. Recently returned from England and my favorite drink of choice there was a flat white. Its like a smoother, nicer cappuccino, but not as milky as a latte. I would say imagine heaven in a cup and it'll taste like a flat white. We need them here.</i>	
<i>Not a coffee drinker</i>	Positiva
<i>The gals at work (I teach) are/were coffee drinkers. I hate coffee! I also don't care to have tea in the morning. However, I love smoothies! I visit a little local drive thru coffee house that offers smoothies in the morning, but it is a bit out of the way to work. I have managed ...</i>	

Fonte: Elaborado pelo autor

Através das recomendações ilustradas nos dois exemplos, é possível que as ideias bases analisadas encontram-se inseridas em um contexto positivo. Ao buscar o real estado dessas ideias no site da Starbucks®, identifica-se que a implementação foi concretizada.

O Quadro 5, demonstra as ideias contidas no conjunto de dados descrito na seção 6.1.2 que estão relacionadas entre as ideias implementadas no site da Starbucks®.

Quadro 5 - Relação das ideias implementadas pela Starbucks®.

Ideias Implementadas	
Polaridade	Idea
Positivo	<i>Alternatives to dairy and soy</i>
	<i>Be able to use rewards on mobile-ordering</i>
	<i>Bring back Mocha Coconut and Coconut Crème Frappuccino Blended Beverages</i>
	<i>Bring the Flat White to the North American market!</i>
	<i>EXPANDED HEALTHY SMOOTHIES</i>
	<i>Gratuity when using gift card</i>
	<i>Happy Hours</i>

	<i>More perks discounts for gold card member</i>
	<i>Offer a Few Regional Food and Beverages in Some Locations</i>
	<i>Order through app</i>
	<i>Peach flavor</i>
	<i>Reusable RED holiday cups</i>
	<i>Verismo Pods in Bulk</i>
	<i>You have lost the community coffee house feel</i>
Negativo	Bring back the old lemon pound cake with yummy glaze on top.

Fonte: Elaborado pelo autor.

Pode-se verificar que a maior parte das ideias foram classificadas como positivas. A ideia classificada como negativa é na verdade um falso negativo. Ao analisar os comentários verifica-se que os usuários reclamam da nova versão, concordando com a ideia em si de retornar uma versão antiga que agradava mais. Ao solicitar a recomendação de ideias semelhantes a esta ideia negativa, o protótipo retornou o conteúdo demonstrado no Quadro 6.

Quadro 6 - Detalhamento da ideia negativa implementada.

Título	Polaridade
Texto	
<i>Bring back the old lemon pound cake with yummy glaze on top</i>	Negativo
<i>For the love of all foods. The old lemon pound cake slices were the best.</i>	
Ideias Recomendadas	
Título	Polaridade
Texto	
<i>Bring back old bakery items</i>	Positivo
<i>bring back the lemon pound cake please the new lemon loaf stinks has no taste at all like eating cardboard really do you guys taste this items before bringing them out also whats wrong with bringing the maple scone to the states what are the canadiens special we like maple here in the states just as much as canada</i>	
<i>Please bring back the old pastries!!</i>	Positivo
<i>Bring back all of the old pastries!! The lemon loaf was our whole familys favorite!! The new one doesnt do any justice compared to the original lemon that had the awesome glaze on it!!</i>	
<i>Go back to the European coffee house</i>	Positivo
<i>Please go back to your original idea of an European coffee house and get rid of the extraneous items like cds, stuffed animals, countless foods and all that factory holiday junk. I love the original Starbucks better. And Howard Schultz, I love you. - An idea from our Annual Meeting of Shareholders on March 19, 2008 in Seattle</i>	
<i>Offer a Few Regional Food and Beverages in Some Locations</i>	Positivo

<i>I live in a city surrounded by many cultures. I feel very blessed to have the exposure of many types people, languages and the best thing of all...the food and beverages! Yum! Cultural acceptance and accommodation is an important aspect to the success of any restaurant and store, it shows respect and acknowledgment of it's customers. My idea for Starbucks is to offer a few regional ...</i>	
<i>What happened to the Lemon Loaf?!</i>	Negativo
<i>I live in Indiana and was appalled when I ate my lemon loaf this morning. I am a lemon loaf/pumpkin loaf FAN! I asked the barista why this tasted funny and his answer was that Starbucks just changed the recipe and started sending this new stuff. I really loved the lemon loaf before it got a new recipe...</i>	
<i>Bring Back Cocunut!</i>	Positivo
<i>I used to get mocha cocunut and cocunut lattes hot or over ice! They were delicious. Best flavour syrup!</i>	
<i>Order on app before hand</i>	Positivo
<i>Wouldn't it be great if we had the opportunity to order our drinks before hand on our app while we wait in line, or on our way to Starbucks and make lines move much more faster. Yes! There is a way...</i>	
<i>Add A tip Button on my Starbucks APP</i>	Positivo
<i>I use to tip almost 4 out Of 5 drinks that I would order at Starbucks. Now that one of Starbucks manager got me to use the APP on my phone, I probally tip 1 out of 20 drinks I order now. This is due to the convience of just walking in the store with my Phone only...</i>	
<i>Mocha Coconut</i>	Neutro
<i>Bring the mocha coconut back for good.</i>	
<i>Coconut Syrup Return</i>	Neutro
<i>I have been told for the last 6 months that the coconut syrup would be back around May. Now I find out that it's not coming back at all????!!!!!! PLEASE bring back the COCONUT!!!!</i>	

Fonte: Elaborado pelo autor.

Analisando o quadro acima, pode-se interpretar que a implementação da ideia, ainda que classificada erroneamente como negativa pelo analisador de sentimento, seria respaldada pelo seu contexto em que as ideias foram classificadas como positivas ou neutras.

A seguir, a seção final é dedicada as considerações finais e sugestões de trabalhos futuros.

7 CONSIDERAÇÕES FINAIS

Neste capítulo apresentam-se as conclusões obtidas durante o processo de desenvolvimento deste trabalho através das etapas de avaliação e análise dos resultados apresentados pelo modelo proposto. Além das conclusões, serão abordados os trabalhos futuros.

7.1 CONSIDERAÇÕES FINAIS

Este trabalho possui o foco em duas áreas de pesquisas, os Sistemas de Recomendação e a Análise de Sentimentos. Através de uma revisão sistemática da literatura, percebe-se um despertar no interesse dos pesquisadores pela união destas duas áreas, através da crescente produção de trabalhos nos últimos anos.

Os Sistemas de Recomendação têm se destacado por auxiliar os usuários na tomada de decisão em cenários complexos que envolvem grande volume de dados. Assim, estes sistemas vêm evoluindo com a união de outras áreas de pesquisa, surgindo a demanda pela integração com a área de Análise de Sentimentos.

Diante disso, este trabalho teve por objetivo geral desenvolver um modelo baseado em Análise de Sentimentos como suporte à Sistemas de Recomendação visando auxiliar o usuário no processo de tomada de decisão. Para atingir o objetivo geral, o trabalho foi conduzido através de objetivos específicos, partindo por uma revisão sistemática da literatura a fim de fundamentar a proposição, o desenvolvimento e a aplicação do modelo. Após esta etapa, percebeu-se a viabilidade da união das duas áreas pesquisadas.

O modelo consistiu na proposição de uma visão lógica e uma visão física. A visão física detalha a tecnologia utilizada e o desenvolvimento como um todo. Contudo, para efeitos de verificação da viabilidade, as fases de análise de sentimento e recomendação baseada em conteúdo foram avaliadas.

No primeiro caso a avaliação ocorreu sobre conjuntos de dados reais colhidos no site da Amazon[®] relacionados a revisões providas por usuários sobre câmeras digitais e filmes, em que cada revisão já possui a sua polarização (classificação positiva ou negativa). O método produzido para tal foi comparado com os métodos estatísticos SVM e NB, amplamente utilizados e referenciados na literatura. Nas avaliações foi possível determinar que a proposição do método (na forma de um analisador de sentimento) possui comportamento similar aos métodos estatísticos com a vantagem de não precisar de uma fase de treinamento.

O desenvolvimento e utilização de um analisador de sentimento baseado em estruturas de conhecimento (*lexicons*), também chamado de *ad-hoc*, se justifica em função da falta de um conjunto de dados no domínio escolhido. Esta falta impossibilita a fase de treinamento e, conseqüentemente, geração de um modelo que possa ser incorporado na fase de execução.

Após a fase de testes do analisador de sentimento, o modelo foi aplicado sobre o cenário composto por um conjunto de dados públicos disponíveis no portal da empresa Starbucks®. Estes dados são compostos por ideias sendo que cada ideia possui um conjunto de comentários. Através da análise da polaridade de cada comentário é possível determinar a polaridade final da ideia. Esta informação é utilizada na fase de recomendação.

Quanto aos modelos de recomendação, optou-se pela aplicação da abordagem baseada em conteúdo. Quanto à abordagem de filtragem colaborativa, o teste se apresentou inviável com o conjunto de dados utilizado pelo modelo. Isto ocorre uma vez que a Starbucks® não disponibiliza os dados que representam as interações dos usuários, por exemplo, quem acessou determinada ideia. Tais informações são fundamentais para recomendações utilizando filtragens colaborativas.

De modo geral, entende-se que o modelo proposto foi capaz de atender o objetivo geral e responder à pergunta de pesquisa deste trabalho. O analisador de sentimentos implementado contendo o tratamento de negação obteve resultados similares aos classificadores estatísticos fornecendo subsídios para as recomendações.

O acoplamento do analisador no método de recomendação de conteúdo permitiu, a partir de uma ideia de interesse, a sugestão de ideias relacionadas considerando a apresentação das polaridades. Pode-se inferir que este conjunto de ideias e as polaridades fornecem subsídios importantes no processo de análise de determinada ideia. Em consultas realizadas no site da empresa Starbucks® foi possível observar que as ideias publicadas no site da organização como ideias concretizadas promovem indícios de que, uma abordagem de recomendação suportada pela análise de sentimentos, pode auxiliar no processo de tomada de decisão.

Sendo assim, a proposição do modelo demonstrou-se apta a prover a união entre a Análise de Sentimentos como suporte aos Sistemas de Recomendação. Os resultados apresentados se mostram consistentes e capazes de auxiliar usuários que utilizam os Sistemas de Recomendação como ferramenta no processo de tomada de decisão.

7.2 SUGESTÕES DE TRABALHOS FUTUROS

Como trabalhos futuros para Análise de Sentimentos vislumbre-se a evolução do algoritmo para identificar e considerar peculiaridades do idioma analisado. Para o caso do idioma utilizado neste trabalho cita-se como exemplo expressões utilizadas por usuários para demonstrar intensidade, como (“YEEEEESSS”, “I AGREEEE”, “I LOOOOOOVE SO MUCH”).

Atualmente, a comunicação textual tem incorporado elementos visuais conhecidos como *emojis* e *emoticons*. Estes elementos visam transmitir uma informação utilizando pouco ou nenhum texto, e são muito utilizados principalmente para demonstrar sentimento ou humor do usuário, por exemplo, :D, ☺, :), :(, ♥.

As gírias também estão muito presentes nas opiniões dos usuários e podem ser de grande valia para definir a polaridade de comentário, dentre muitas encontradas, citam-se: OMG, LOL e FTW.

Outra característica que pode ser agregada ao algoritmo é a consideração de um contexto ao identificar subjetividade como, por exemplo, a expressão “O objeto é muito leve”. Cita-se ainda o tratamento de ironia ou frustração, como nos casos em que o usuário tece elogios em seu comentário conduzindo o leitor a ter uma expectativa que abruptamente é contrastada. No inglês este padrão é chamado de *thwarted-expectations*.

No que tange aos Sistemas de Recomendação, cabe o desenvolvimento da abordagem de filtragem colaborativa associada à análise de sentimentos, utilizando os comentários do usuário para aprimorar seu perfil, visando aumentar a efetividade de recomendações para usuários com perfis semelhantes. Abordagens híbridas ou baseadas em conhecimento também podem ser incrementadas utilizando um analisador de sentimento.

Por fim, o desenvolvimento de uma plataforma de gestão de ideias que registre todas as interações dos usuários e se utilize dos conceitos de recomendação e análise de sentimento pode representar uma importante ferramenta voltada à gestão de inovações nas organizações.

REFERÊNCIAS

ACI. **The Data Explosion in 2014 Minute by Minute – Infographic**. 2014. Disponível em: <<http://aci.info/2014/07/12/the-data-explosion-in-2014-minute-by-minute-infographic/>>. Acesso em: 6 mar. 2016.

ADAMS, R., BESSANT, J., PHELPS, R. Innovation management measurement: a review. **International Journal of Management Reviews**, vol. 8, nº.1, p. 21–47, 2006.

ADOMAVICIUS, Gediminas; TUZHILIN, Alexander. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. **Knowledge And Data Engineering**, S. I., v. 17, n. 6, p.734-749, jun. 2005.

ANDRIOLE, Stephen J., Business impact of Web 2.0 technologies. **Communications Of The Acm**, [s.l.], v. 53, n. 12, p.67-79, 1 dez. 2010. Association for Computing Machinery (ACM).

AGARWAL, Apoorv et al. Sentiment analysis of Twitter data. **Lsm '11 Proceedings Of The Workshop On Languages In Social Media**. Portland, Oregon, p. 30-38. jul. 2011.

BAILEY, Brian P.; HORVITZ, Eric. What's Your Idea? A Case Study of a Grassroots Innovation Pipeline within a Large Software Company. **Chi '10 Proceedings Of The 28th International Conference On Human Factors In Computing Systems**, Atlanta, Georgia, 2010.

BARBIERI, J. C.; ÁLVARES, A. C. T.; CAJAZEIRA, J. E. R. **Gestão de Ideias para inovação contínua**. Porto Alegre: Bookman, 2009.

BALAZS, Jorge A.; VELÁSQUEZ, Juan D., Opinion Mining and Information Fusion: A survey. **Information Fusion**, [s.l.], v. 27, p.95-110, jan. 2016.

BELL, Robert M.; KOREN, Yehuda. Improved Neighborhood-based Collaborative Filtering, **KDD-Cup and Workshop**, ACM press, 2007.

BELLO-ORGAZ, Gema; JUNG, Jason J.; CAMACHO, David. Social big data: Recent achievements and new challenges. **Information Fusion**, [s.l.], v. 28, p.45-59, mar. 2016.

BERNERS-LEE, Tim. **Information Management: A Proposal**. 1989. Disponível em: < <http://www.w3.org/History/1989/proposal.html> >. Acesso em: 27 de fevereiro de 2016.

BREM, Alexander; VOIGT, Kai-ingo. Integration of market pull and technology push in the corporate front end and innovation management—Insights from the German software industry. **Technovation**, [s.l.], v. 29, n. 5, p.351-367, maio 2009.

BOBADILLA, Jesús. et al. Recommender systems survey. **Knowledge-based Systems**, [s.l.], v. 46, p.109-132, jul. 2013. Elsevier BV.

BORRÀS, Joan; MORENO, Antonio; VALLS, Aida. Intelligent tourism recommender systems: A survey. **Expert Systems With Applications**, [s.l.], v. 41, n. 16, p.7370-7389, nov. 2014. Elsevier BV.

BOTHOS, E.; APOSTOLOU, D.; MENTZAS, G. Collective intelligence with web-based information aggregation markets: The role of market facilitation in idea management. *Expert Systems with Applications*, Amsterdam, vol. 39, n°. 1, p. 1333-1345, 2012.

BUNGE, M. *Treatise on Basic Philosophy. Part II: Life Science, Social Science and Technology*. Boston: D. Reidel, 1985.

CANUTO, Sérgio et al. An Efficient and Scalable MetaFeature-based Document Classification Approach based on Massively Parallel Computing. **Proceedings Of The 38th International Acm Sigir Conference On Research And Development In Information Retrieval - Sigir '15**, [s.l.], p.333-342, 2015. Association for Computing Machinery (ACM)

CAPDEVILA, Joan; ARIAS, Marta; ARRATIA, Argimiro. GeoSRS: A hybrid social recommender system for geolocated data. **Information Systems**, [s.l.], v. 57, p.111-128, abr. 2016. Elsevier BV.

CARRER-NETO, Walter et al. Social knowledge-based recommender system. Application to the movies domain. **Expert Systems With Applications**, [s.l.], v. 39, n. 12, p.10990-11000, set. 2012.

CECI, Flávio. **UM MODELO BASEADO EM CASOS E ONTOLOGIA PARA APOIO À TAREFA INTENSIVA EM CONHECIMENTO DE CLASSIFICAÇÃO COM FOCO NA ANÁLISE DE SENTIMENTOS**. 2015. 211 f. Tese (Doutorado) - Curso de Programa de Pós-graduação em Engenharia e Gestão do Conhecimento, Engenharia e Gestão do Conhecimento, Universidade Federal de Santa, Florianópolis, 2015.

CECI, Flávio et al. Adapting Sentiments with Context. **Case-based Reasoning Research And Development**, [s.l.], p.44-59, 2015.

CHEN, Rung-ching et al. A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection. **Expert Systems With Applications**, [s.l.], v. 39, n. 4, p.3995-4006, mar. 2012.

CHEN, Yen-liang; CHENG, Li-chen; CHUANG, Ching-nan. A group recommendation system with consideration of interactions among group members. **Expert Systems With Applications**, [s.l.], v. 34, n. 3, p.2082-2090, abr. 2008.

COLACE, Francesco; CASABURI, Luca. An Approach for Sentiment Classification of Music. **Proceedings Of The 18th International Conference On Enterprise Information Systems (iceis 2016)**, [s.l.], v. 2, p.421-426, 2016.

COLACE, Francesco et al. A collaborative user-centered framework for recommending items in Online Social Networks. **Computers In Human Behavior**, [s.l.], v. 51, p.694-704, out. 2015.

COLOMO-PALACIOS, Ricardo et al. Providing knowledge recommendations: an approach for informal electronic mentoring. **Interactive Learning Environments**, [s.l.], v. 22, n. 2, p.221-240, 21 nov. 2012. Informa UK Limited.

COLOMO-PALACIOS, Ricardo et al. Towards a social and context-aware mobile recommendation system for tourism. **Pervasive And Mobile Computing**, [s.l.], v. 38, p.505-515, jul. 2017. Elsevier BV.

COOPER, R. G.; EDGETT, S. J.; KLEINSCHMIDT, E. J.; Optimizing the Stage-Gate process: What best practice companies are doing – Part I. **Research Technology Management**, v. 45, n. 5, 2002.

COOPER, Robert G. Stage-gate systems: A new tool for managing new products, **Business Horizons**, vol. 33, n. 3, p. 44-54, 1990.

CRUZ, Fermín. L. et al. ‘Long autonomy or long delay?’The importance of domain in opinion mining. **Expert Systems with Applications**, v. 40, n. 8, p. 3174-3184, 2013.

CULOTTA, Aron. Towards detecting influenza epidemics by analyzing Twitter messages. **Proceedings Of The First Workshop On Social Media Analytics - Soma '10**, [s.l.], p.119-128, 2010. Association for Computing Machinery (ACM).

CUPANI, A. A Filosofia da tecnologia: um convite. UFSC: Florianópolis, 2011.

D'ADDIO, Rafael Martins; MANZATO, Marcelo Garcia. A Collaborative Filtering Approach Based on User's Reviews. **2014 Brazilian Conference On Intelligent Systems**, [s.l.], p.204-209, out. 2014. Institute of Electrical & Electronics Engineers (IEEE).

DAS, Sanjiv R.; CHEN, Mike Y. Yahoo! for Amazon: Extracting market sentiment from stock message boards. **Proceedings Of The Asia Pacific Finance Association Annual Conference (APFA)**, 2001.

DONG, Ruihai et al. Opinionated Product Recommendation. In: **Case-Based Reasoning Research and Development**. Springer Berlin Heidelberg, p. 44-58. 2013.

DONG, Ruihai; O'MAHONY, Michael P.; SMYTH, Barry. Further Experiments in Opinionated Product Recommendation. **Case-based Reasoning Research And Development**, [s.l.], p.110-124, 2014. Springer Science + Business Media.

ECCLESTON, Derek; GRISERI, Luca. How does Web 2.0 stretch traditional influencing patterns? **International Journal Of Market**

Research: Jmrs: the journal of the Market Research Society. p. 591-616. 2008.

EGGHE, Leo; MICHEL, Christine. Strong similarity measures for ordered sets of documents in information retrieval. **Information Processing and Management: an International Journal**, v. 38, n. 6, p. 823-848, 2002.

ESULI, Andrea; SEBASTIANI, Fabrizio. SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining. **In Proceedings Of The 5th Conference On Language Resources And Evaluation (Irec'06)**, Genoa, Italia, p.417-422, maio 2006.

FERREIRA, F. C.; OLIVEIRA, A. A. Os Sistemas de Recomendação na Web Como Determinantes Prescritivos na Tomada de Decisão. **Revista de Gestão da Tecnologia e Sistemas de Informação**. Brasil, p.353-368. Agosto. 2012.

FLYNN, M. *et al.* Idea management for organizational innovation. **International Journal of Innovation Management**, Washington, v. 7, n. 5, p. 417-442, 2003.

FUKUMOTO, Fumiyo; MOTEGI, Chihiro; MATSUYOSHI, Suguru. Collaborative Filtering based on Sentiment Analysis of Guest Reviews for Hotel Recommendation. **Proceedings Of The International Conference On Knowledge Discovery And Information Retrieval**, [s.l.], p.193-198, 2012.

GARCÍA-CUMBRERAS, Miguel Á.; MONTEJO-RÁEZ, Arturo; DÍAZ-GALIANO, Manuel C. Pessimists and optimists: Improving collaborative filtering through sentiment analysis. **Expert Systems With Applications**, [s.l.], v. 40, n. 17, p.6758-6765, dez. 2013.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (Org.). **Métodos de Pesquisa**. Porto Alegre: Plageder, 2009. 120 p. (Educação a Distância).

GENA, Cristina et al. The Impact of Rating Scales on User's Rating Behavior. **User Modeling, Adaption And Personalization**, [s.l.], p.123-134, 2011.

GEUM, Youngjung; PARK, Yongtae. How to generate creative ideas for innovation: a hybrid approach of WordNet and morphological analysis. **Technological Forecasting And Social Change**, [s.l.], v. 111, p.176-187, out. 2016.

GHAZANFAR, Mustansar Ali; PRÜGEL-BENNETT, Adam. Leveraging clustering approaches to solve the gray-sheep users problem in recommender systems. **Expert Systems With Applications**, [s.l.], v. 41, n. 7, p.3261-3275, jun. 2014.

GHORPADE, Tushar; RAGHA, Lata. Featured based sentiment classification for hotel reviews using NLP and Bayesian classification. **2012 International Conference On Communication, Information & Computing Technology (iccict)**, [s.l.], p.1-5, out. 2012. Institute of Electrical & Electronics Engineers (IEEE).

GIL, Antônio Carlos. **Como Elaborar Projetos de Pesquisa**. 4. ed. São Paulo: Atlas, 2002. 176 p.

GOLDBERG, D.; NICHOLS, D.; OKI, B.; TERRY, D. Using collaborative filtering to weave an information tapestry. **Communications of the Association of Computing Machinery**, v. 35, n. 12, p. 61–70, 1992.

GONZALEZ-CARRASCO, Israel et al. PB-ADVISOR: A private banking multi-investment portfolio advisor. **Information Sciences**, [s.l.], v. 206, p.63-82, nov. 2012. Elsevier BV.

GURINI, Davide Feltoni; GASPARETTI, Fabio. A Sentiment-Based Approach to Twitter User Recommendation. **Conference: Rswb@recsys2013**, Hong Kong, p.1-4, 2013.

GUTIÉRREZ, Yoan; VÁZQUEZ, Sonia; MONTOYO, Andrés. A semantic framework for textual data enrichment. **Expert Systems With Applications**, [s.l.], v. 57, p.248-269, set. 2016.

HALL, Wendy; TIROPANIS, Thanassis. Web evolution and Web Science. **Computer Networks**, [s.l.], v. 56, n. 18, p.3859-3865, dez. 2012.

HASS, B. H., Walsh, G., Kilian, T. (ed.): Web 2.0 - Neue Perspektiven für Marketing und Medien, Berlin Heidelberg, p.73-87, 2008.

HE, Chen; PARRA, Denis; VERBERT, Katrien. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. **Expert Systems With Applications**, [s.l.], v. 56, p.9-27, set. 2016.

HERNANDO, Antonio et al. Trees for explaining recommendations made through collaborative filtering. **Information Sciences**, [s.l.], v. 239, p.1-17, ago. 2013.

HU, Minqing; LIU, Bing. Mining and summarizing customer reviews. **Proceedings Of The 2004 Acm Sigkdd International Conference On Knowledge Discovery And Data Mining - Kdd '04**, [s.l.], p.168-177, 2004. Association for Computing Machinery (ACM).

HUIZINGH, Eelko K.r.e.. Open innovation: State of the art and future perspectives. **Technovation**, [s.l.], v. 31, n. 1, p.2-9, jan. 2011.

IBOPE. **Número de brasileiros com acesso à internet chega a 83,4 milhões de pessoas.** 2012. Disponível em: <<http://www.ibopeinteligencia.com/noticias-e-pesquisas/numero-de-brasileiros-com-acesso-a-internet-chega-a-834-milhoes-de-pessoas/>>. Acesso em: 05 mar. 2016.

IBOPE. **Internauta gasta em média 10 horas e 26 minutos em redes sociais.** 2013. Disponível em: <<http://www.ibopeinteligencia.com/noticias-e-pesquisas/internauta-gasta-em-media-10-horas-e-26-minutos-em-redes-sociais/>>. Acesso em: 05 mar. 2016.

ISSA, Hassan et al. User-sentiment based Evaluation for Market Fitness Trackers: Evaluation of Fitbit One, Jawbone Up and Nike+ Fuelband based on Amazon.com Customer Reviews. **Proceedings Of The 1st International Conference On Information And Communication Technologies For Ageing Well And E-health (ict4ageingwell-2015)**, [s.l.], p.171-179, 2015.

JANNACH, Dietmar. *Recommender Systems: An Introduction*. New York: Cambridge University Press 32 Avenue Of The Americas, New York, Ny 10013-2473, Usa, 2011.

JONES, W. P.; FURNAS, G. W. Pictures of Relevance: A geometric Analysis of Similarity Measures. **Journal of the American Society for Information Science**, Maryland, v.36, n. 6, p.420-442,1987.

JSON.ORG. **Introdução ao JSON: ECMA-404** The JSON Data Interchange Standard. Disponível em: <<http://www.json.org/json-pt.html>>. Acesso em: 20 abr. 2017.

KANG, Daekook; PARK, Yongtae. Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. **Expert Systems With Applications**, [s.l.], v. 41, n. 4, p.1041-1050, mar. 2014.

KARAMPIPERIS, Pythagoras; KOUKOURIKOS, Antonis; STOITSIS, Giannis. Collaborative Filtering Recommendation of Educational Content in Social Environments Utilizing Sentiment Analysis Techniques. **Recommender Systems For Technology Enhanced Learning**, [s.l.], p.3-23, 2014.

KATHARINA, Rasch. An unsupervised recommender system for smart homes. *Journal Of Ambient Intelligence And Smart Environments*, [s.l.], v. 6, n. 1, p.21-37, 2014. IOS Press.

KAZAMA, Jun'ichi; TSUJII, Jun'ichi. Maximum Entropy Models with Inequality Constraints: A Case Study on Text Categorization. **Machine Learning**, [s.l.], v. 60, n. 1-3, p.159-194, set. 2005.

KIM, Dan; YUE, Kwokbun; HALL, Sharon Perkins; GATES, Tracy. Global diffusion of the internet XV: Web 2.0 technologies, principles, and applications: A conceptual framework from technology push and demand pull perspective. **Communications Of The Association For Information Systems**. [s. L.], p. 657-672. 2009.

KIM, Hea-jin; SONG, Min. An Ontology-Based Approach to Sentiment Classification of Mixed Opinions in Online Restaurant Reviews. **Lecture Notes In Computer Science**, [s.l.], p.95-108, 2013. Springer Science + Business Media.

KNIJNENBURG, Bart P. et al. Explaining the user experience of recommender systems. **User Model User-adap Inter**, [s.l.], v. 22, n. 4-5, p.441-504, 10 mar. 2012.

KOEN, Peter A. *et al.* **Fuzzy front end: Effective methods, tools, and techniques**. Wiley, New York, NY, 2002.

KOREN, Yehuda; BELL, Robert; VOLINSKY, Chris. Matrix Factorization Techniques for Recommender Systems. **Ieee Computer Society Press Los Alamitos**, California, v. 42, n. 8, p.30-37, ago. 2009. Institute of Electrical and Electronics Engineers (IEEE).

KOUKOURIKOS, Antonis; STOITSIS, Giannis; KARAMPIPERIS, Pythagoras. Sentiment Analysis: A tool for Rating Attribution to Content in Recommender Systems. **2nd Workshop On Recommender Systems For Technology Enhanced Learning (recsystem 2012)**, Saarbrücken (germany), p.61-70, set. 2012.

KROTZFLEISCH, Harald Von; MERGEL, Ines; MANOUCHEHRI, Shakib; SCHAAR SCHMIDT, Mario. 2008.**Corporate Web 2.0 Applications**.

KULTIMA, Annakaisa; KARVINEN, Juho. Idea practices and attitudes towards innovation in game development. **Proceedings Of The 20th International Academic Mindtrek Conference On - Academicmindtrek '16**, [s.l.], p.331-340, 2016. ACM

LEE, Hanjun et al. The More the Worse? Mining Valuable Ideas with Sentiment Analysis for Idea Recommendation. **Association For Information Systems: AIS Electronic Library (AISel)**, [s.l.], n. 30, p.1-17, Jun. 2013.

LEUNG, Cane Wing-ki et al. A probabilistic rating inference framework for mining user preferences from reviews. **World Wide Web**, [s.l.], v. 14, n. 2, p.187-215, 17 fev. 2011. Springer Science + Business Media.

LEVI, Asher et al. Finding a needle in a haystack of reviews. **Proceedings Of The Sixth Acm Conference On Recommender Systems - Recsys '12**, [s.l.], p.115-122, 2012. Association for Computing Machinery (ACM).

LEWIS, Seth C.; ZAMITH, Rodrigo; HERMIDA, Alfred. Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods. **Journal Of Broadcasting & Electronic Media**, [s.l.], v. 57, n. 1, p.34-52, jan. 2013. Informa UK Limited.

LI, Mingguo; KANKANHALLI, Atreyi; KIM, Seung Hyun. Which ideas are more likely to be implemented in online user innovation communities? An empirical analysis. **Decision Support Systems**, [s.l.], v. 84, p.28-40, abr. 2016.

LI, Sheng-Tun; TSAI, Fu-Ching. A fuzzy conceptualization model for text mining with application in opinion polarity classification. **Knowledge-Based Systems**. v.39 p.23–33, 2013.

LI, Xingsen; LI, Liping; CHEN, Zhengxin. Toward extenics-based innovation model on intelligent knowledge management. **Annals of Data Science**, v. 1, n. 1, p. 127-148, 2014.

LI, Yung-ming; LI, Tsung-ying. Deriving market intelligence from microblogs. **Decision Support Systems**, [s.l.], v. 55, n. 1, p.206-217, abr. 2013.

LIN, Yuming et al. Integrating the optimal classifier set for sentiment analysis. **Soc. Netw. Anal. Min.**, [s.l.], v. 5, n. 1, p.1-13, 11 set. 2015. Springer Science + Business Media.

LINDEN, Greg; SMITH, Brent; YORK, Jeremy. Amazon.com recommendations: item-to-item collaborative filtering. **Ieee Internet Computing**, [s.l.], v. 7, n. 1, p.76-80, jan. 2003.

LIU, Bing. Sentiment Analysis and Opinion Mining. **Synthesis Lectures On Human Language Technologies**, [s.l.], v. 5, n. 1, p.1-167, 23 maio 2012. Morgan & Claypool Publishers LLC.

LIU, Hongyan et al. Combining user preferences and user opinions for accurate recommendation. **Electronic Commerce Research And Applications**, [s.l.], v. 12, n. 1, p.14-23, jan. 2013.

LUCENE, Apache. **Apache Lucene Core**. Disponível em: <<https://lucene.apache.org/core/>>. Acesso em: 20 abr. 2017.

LUNARDI, Alexandre et al. Domain-Tailored Multiclass Classification of User Reviews Based on Binary Splits. **Social Computing And Social Media**, [s.l.], p.298-309, 2016. Springer Nature.

MAGNUSSON, Peter R.; NETZ, Johan; WÄSTLUND, Erik. Exploring holistic intuitive idea screening in the light of formal criteria. **Technovation**, [s.l.], v. 34, n. 5-6, p.315-326, Maio 2014.

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. **Introduction to Information Retrieval**. Cambridge: Cambridge University Press, 2009. 569 p.

MARTÍNEZ-CÁMARA, Eugenio et al. Sentiment analysis in Twitter. **Natural Language Engineering**, [s.l.], v. 20, n. 01, p.1-28, 27 Nov. 2012. Cambridge University Press (CUP).

MARTINEZ-TORRES, Rocio; OLMEDILLA, Maria. Identification of innovation solvers in open innovation communities using swarm intelligence. **Technological Forecasting And Social Change**, [s.l.], v. 109, p.15-24, ago. 2016.

MCGRATH, Michael E. **Next generation product development: how to increase productivity, cut costs, and reduce cycle times**. New York: McGraw-Hill, 2004.

MCNEMAR, Quinn. **Note on the sampling error of the difference between correlated proportions or percentages**. *Psychometrika*, v. 12, n. 2, p. 153-157, 1947.

MELVILLE, Prem; SINDHWANI, Vikas. Recommender Systems. **Encyclopedia Of Machine Learning**, [s.l.], p.829-838, 2011. Springer Science + Business Media.

MOHAN, Mayoor; VOSS, Kevin E.; JIMÉNEZ, Fernando R.. Managerial disposition and front-end innovation success. **Journal Of Business Research**, [s.l.], v. 70, p.193-201, jan. 2017.

MONTANER, M., LÓPEZ, B., DE LA ROSA, J. L. A Taxonomy of Recommender Agents on the Internet. **Artificial Intelligence Review**. Netherlands: Kluwer Academic Publishers, p. 285-330,2003.

MONTEIRO, Silvana Drumond; FIDENCIO, Marcos Vinicius. As dobras semióticas do ciberespaço: da web visível à invisível. **Transinformação**, Campinas, v. 1, n. 25, p.35-46, abr. 2013.

MORAES, Rodrigo; VALIATI, João Francisco; GAVIÃO NETO, Wilson P.. Document-level sentiment classification: An empirical comparison between SVM and ANN. **Expert Systems With Applications**, [s.l.], v. 40, n. 2, p.621-633, fev. 2013.

MOREO, Alejandro et al. Lexicon-based comments-oriented news sentiment analyzer system. **Expert Systems with Applications**, v. 39, n. 10, p. 9166-9180, 2012.

MULHOLLAND, Eleanor et al. 360-MAM-Affect: Sentiment Analysis with the Google Prediction API and EmoSenticNet. **Proceedings Of The 7th International Conference On Intelligent Technologies For Interactive Entertainment**, [s.l.], p.217-221, 2015. Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (ICST).

NAKATSUJI, Makoto et al. Linked taxonomies to capture users' subjective assessments of items to facilitate accurate collaborative filtering. **Artificial Intelligence**, [s.l.], v. 207, p.52-68, fev. 2014.

NASUKAWA, Tetsuya; YI, Jeonghee. Sentiment analysis: capturing favorability using natural language processing. **Proceedings Of The International Conference On Knowledge Capture - K-cap '03**, [s.l.], v. 0, n. 0, p.70-77, out. 2003.

ORACLE. **Java SE Technologies - Database: The Java Database Connectivity (JDBC)**. Disponível em: <<http://www.oracle.com/technetwork/java/javase/jdbc/index.html>>. Acesso em: 20 abr. 2017.

OECD. Oslo Manual. **The Measurement Of Scientific And Technological Activities**, [s.l.], p.1-162, 10 nov. 2005. OECD Publishing.

O'REILLY, Tim. **WhatIs Web 2.0: Design Patternsand Business Models for the Next Generationof Software**. 2005. Disponível em:

<<http://oreilly.com/pub/a/web2/archive/what-is-web-20.html>>. Acesso em: 27 de fevereiro de 2016.

PANG, Bo; LEE, Lillian. Opinion Mining and Sentiment Analysis. **Fnt In Information Retrieval**, [s.l.], v. 2, n. 12, p.1-135, 2008. Now Publishers.

PANG, Bo; LEE, Lillian; VAITHYANATHAN, Shivakumar. Thumbs up? **Proceedings Of The Acl-02 Conference On Empirical Methods In Natural Language Processing - Emnlp '02**, [s.l.], p.79-86, 2002. Association for Computational Linguistics (ACL).

PELEJA, Filipa; DIAS, Pedro; MAGALHAES, Joao. A Regularized Recommendation Algorithm with Probabilistic Sentiment-Ratings. **2012 Ieee 12th International Conference On Data Mining Workshops**, [s.l.], p.701-708, dez. 2012. Institute of Electrical & Electronics Engineers (IEEE).

PELEJA, Filipa et al. A recommender system for the TV on the web: integrating unrated reviews and movie ratings. **Multimedia Systems**, [s.l.], v. 19, n. 6, p.543-558, 6 mar. 2013.

POMMERANZ, Alina et al. Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. **User Model User-adap Inter**, [s.l.], v. 22, n. 4-5, p.357-397, 15 mar. 2012.

PPGTIC. **Sobre o PPGTIC**. 2016. Disponível em: <<http://ppgtic.ufsc.br/sobre-o-ppgtic>>. Acesso em: 30 abr. 2016.

PRÖLLOCHS, Nicolas; FEUERRIEGEL, Stefan; NEUMANN, Dirk. Negation scope detection in sentiment analysis: Decision support for news-driven trading. **Decision Support Systems**, [s.l.], v. 88, p.67-75, ago. 2016.

QIU, Likun et al. SELC. **Proceeding Of The 18th Acm Conference On Information And Knowledge Management - Cikm '09**, [s.l.], p.929-936, 2009. Association for Computing Machinery (ACM).

QUANDT, Carlos Olavo et al. PROGRAMAS DE GESTÃO DE IDEIAS E INOVAÇÃO: AS PRÁTICAS DAS GRANDES

EMPRESAS NA REGIÃO SUL DO BRASIL. **Review Of Administration And Innovation - Rai**, [s.l.], v. 11, n. 3, p.176-199, 6 nov. 2014.

RAVI, Kumar; RAVI, Vadlamani. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. **Knowledge-based Systems**, [s.l.], p.1-33, jun. 2015.

RESNICK, Paul; IAKOVOU, Neophytos; SUSHAK, Mitesh; BERGSTROM, Peter; RIEDL, John. GroupLens: An open architecture for collaborative filtering of netnews. **Proceeding of the ACM Conference on Computer Supported Cooperative Work**, p. 175-186, 1994.

RICCI, Francesco; ROKACH, Lior; SHAPIRA, Bracha. **Recommender Systems Handbook**. Berlin: Springer Us, 2011. 702 p.

RODRÍGUEZ-GONZÁLEZ, Alejandro et al. FAST: Fundamental Analysis Support for Financial Statements. Using semantics for trading recommendations. **Information Systems Frontiers**, [s.l.], v. 14, n. 5, p.999-1017, 13 ago. 2011. Springer Nature.

ROSA, Renata L.; RODRIGUEZ, Demsteneso Z.; BRESSAN, Graca. Music recommendation system based on user's sentiments extracted from social networks. **Ieee Transactions On Consumer Electronics**, [s.l.], v. 61, n. 3, p.359-367, ago. 2015. Institute of Electrical and Electronics Engineers (IEEE).

RUI, Huaxia; LIU, Yizao; WHINSTON, Andrew. Whose and what chatter matters? The effect of tweets on movie sales. **Decision Support Systems**, [s.l.], v. 55, n. 4, p.863-870, 1 nov. 2013.

SALDIVAR, Jorge et al. Idea management in social networks: A study of how to tap into the ideas of facebook communities. **Proceedings - 2016 International Conference On Collaboration Technologies And Systems, Cts 2016**, Orlando, Florida, p.3-10, out. 2016.

SALDIVAR, Jorge et al. On the (in)effectiveness of the Share/Tweet button: A study in the context of idea management for civic participation. **Ieee Internet Computing**, [s.l.], p.1-1, 2017. Institute of Electrical and Electronics Engineers (IEEE).

SANDSTRÖM, Christian; BJÖRK, Jennie. Idea management systems for a changing innovation landscape. **International Journal Of Product Development**, [s.l.], v. 11, n. 3/4, p.310-328, jul. 2010.

SCHWEISFURTH, Tim G.; HERSTATT, Cornelius. How internal users contribute to corporate product innovation: the case of embedded users. **R&d Management**, [s.l.], v. 46, n. 1, p.107-126, 28 nov. 2014.

SCHWENKER, Friedhelm; TRENTIN, Edmondo. Pattern classification and clustering: A review of partially supervised learning approaches. **Pattern Recognition Letters**, [s.l.], v. 37, p.4-14, fev. 2014. Elsevier BV.

SEO, Dongback; LEE, Jung. Web_2.0 and five years since: How the combination of technological and organizational initiatives influences an organization's long-term Web_2.0 performance. **Telematics And Informatics**, [s.l.], v. 33, n. 1, p.232-246, fev. 2016.

SÉRGIO, Marina Carradore. **UM MODELO BASEADO EM ONTOLOGIA E ANÁLISE DE AGRUPAMENTO PARA SUPORTE À GESTÃO DE IDEIAS**. 2016. 128 f. Dissertação (Mestrado) - Curso de Pós-Graduação em Engenharia e Gestão do Conhecimento, Centro Tecnológico, Universidade Federal de Santa Catarina, Florianópolis, 2016.

SÉRGIO, Marina Carradore; GONÇALVES, Alexandre Leopoldo; SOUZA, João Artur de. Um Modelo para Auxiliar na Tomada de Decisão no Domínio de Gestão de Ideias. *Future Studies Research Journal: Trends And Strategies*, São Paulo, v. 2, n. 7, p.95-118, jul. 2015.

SERRANO-GUERRERO, Jesus et al. Sentiment analysis: A review and comparative analysis of web services. **Information Sciences**, [s.l.], v. 311, p.18-38, ago. 2015.

SHARDANAND, Upendra; MAES, Pattie. Social information filtering: Algorithms for automating "Word of Mouth". **Proceedings of the SIGCHI Conference on Human Factors in Computing Systems**, p. 210-217, 1995.

SINGH, Vivek Kumar; MUKHERJEE, Mousumi; MEHTA, Ghanshyam Kumar. Combining a Content Filtering Heuristic and Sentiment Analysis for Movie Recommendations. **Communications In Computer And Information Science**, [s.l.], p.659-664, 2011. Springer Science + Business Media.

SINGH, Vivek Kumar et al. A Content-Based eResource Recommender System to Augment eBook-Based Learning. **Lecture Notes In Computer Science**, [s.l.], p.257-268, 2013. Springer Science + Business Media.

SOLR, Apache. **Learn more about Solr**. Disponível em: <<http://lucene.apache.org/solr/>>. Acesso em: 20 abr. 2017.

STARBUCKS. **Starbucks Company Timeline**. 2016. Disponível em: <<https://www.starbucks.com/about-us/company-information>>. Acesso em: 27 Abril 2016.

STEPHEN, Andrew T. The role of digital and social media marketing in consumerbehavior. **CurrentOpinion In Psychology**, [s.l.], v. 10, p.17-21, ago. 2016.

SUN, Jianshan et al. Mining affective text to improve social media item recommendation. **Information Processing & Management**, [s.l.], v. 51, n. 4, p.444-457, jul. 2015.

SURESH, Vaishak; ROOHI, Syeda; EIRINAKI, Magdalini. Aspect-based opinion mining and recommendationsystem for restaurant reviews. **Proceedings Of The 8th Acm Conference On Recommender Systems - Recsys '14**, [s.l.], p.361-362, 2014. Association for Computing Machinery (ACM).

TANG, Duyu et al. A Joint Segmentation and Classification Framework for Sentence Level Sentiment Classification. **Ieee/acm Transactions On Audio, Speech, And Language Processing**, [s.l.], v. 23, n. 11, p.1750-1761, nov. 2015. Institute of Electrical & Electronics Engineers (IEEE).

TAKÁCS, Gábor; PILÁSZ, István; NÉMETH, Bottyán; TIKK, Domonkos. Scalable Collaborative Filtering Approaches for Large

Recommender Systems. *Journal Of Machine Learning Research*, Budapest, v. 10, p.623-656, 2009.

TOFFLER, Alvin. **The third wave**. Bantam Books: New York, 1980.

TONG, Richard M. An Operational System for Detecting and Tracking Opinions in On-line Discussion. **Workshop On Operational Text Classification (OCT)**, [s.l.], Set, 2001.

VALENCIA-GARCÍA, Rafael et al. Informal learning through expertise mining in the social web. *Behaviour & Information Technology*, [s.l.], v. 31, n. 8, p.757-766, ago. 2012. Informa UK Limited.

VAN ZONEN, Ward; VAN DER MEER, Toni G. L. A. Social media research: The application of supervised machine learning in organizational communication research.. **Computers In Human Behavior**, [s.l.], v. 63, p.132-141, out. 2016. Elsevier BV.

VANDENBOSCH, Betty; SAATCIOGLU, Argun; FAY, Sharon. Idea Management: A Systemic View. **Journal Of Management Studies**, [s.l.], v. 43, n. 2, p.259-288, mar. 2006.

VARGAS, M. *Metodologia da pesquisa tecnológica*. Rio de Janeiro, 1985.

VOIGT, Kai-ingo; BREM, Alexander. Integrated Idea Management in Emerging Technology Ventures. **2006 Ieee International Conference On Management Of Innovation And Technology**, Nuremberg/germany, v. 1, p.211-215, jun. 2006.

VOIGT, Kai-ingo; ERNST, Markus. Use of Web 2.0 applications in product development: an empirical study of the potential for knowledge creation and exchange in research and development. **International Journal Of Engineering, Science And Technology**. Sem Cidade, p. 54-68. Nov. 2010.

WANG, Yuanhong; LIU, Yang; YU, Xiaohui. Collaborative Filtering with Aspect-Based Opinion Mining: A Tensor Factorization Approach. **2012 Ieee 12th International Conference On Data Mining**, [s.l.],

p.1152-1157. Dez. 2012. Institute of Electrical & Electronics Engineers (IEEE).

WESTERSKI, Adam; DALAMAGAS, Theodore; IGLESIAS, Carlos A. Classifying and comparing community innovation in Idea Management Systems. **Decision Support Systems**, [s.l.], v. 54, n. 3, p.1316-1326, fev. 2013.

XIE, Yusheng et al. Crowdsourcing recommendations from social sentiment. **Proceedings Of The First International Workshop On Issues Of Sentiment Discovery And Opinion Mining - Wisdom '12**, [s.l.], p.1-8, 2012. Association for Computing Machinery (ACM).

YANG, Dingqi et al. A sentiment-enhanced personalized location recommendation system. **Proceedings Of The 24th Acm Conference On Hypertext And Social Media - Ht '13**, [s.l.], p.119-128, 2013. Association for Computing Machinery (ACM).

YANG, Xiwang et al. A survey of collaborative filtering based social recommender systems. **Computer Communications**, [s.l.], v. 41, p.1-10, mar. 2014.

YOSHIDA, Daichi; MIYAZAWA, Jun-ichi; TAKAHASHI, Shingo. Role of community in user innovation generation and diffusion—Focusing on non-brand communities in the mountain climbing market. **Technological Forecasting And Social Change**, [s.l.], v. 88, p.1-15, out. 2014.

ZHANG, Lei; LIU, Bing. Identifying noun product features that imply opinions. **Hlt '11 Proceedings Of The 49th Annual Meeting Of The Association For Computational Linguistics: Human Language Technologies: short papers**, Portland, Oregon, v. 2, n. 49, p.575-580, jun. 2011.

ZHANG, Weishi et al. Augmenting Chinese Online Video Recommendations by Using Virtual Ratings Predicted by Review Sentiment Classification. **2010 Ieee International Conference On Data Mining Workshops**, [s.l.], p.1143-1150, dez. 2010. Institute of Electrical and Electronics Engineers (IEEE).

ZHANG, Weishi et al. Generating virtual ratings from chinese reviews to augment online recommendations. **Acm Transactions On Intelligent Systems And Technology**, [s.l.], v. 4, n. 1, p.1-17, 1 jan. 2013. Association for Computing Machinery (ACM).