

**FACULDADE METROPOLITANA DA GRANDE FORTALEZA**

**CURSO DE SISTEMA DE INFORMAÇÃO**

**ANTONIO JACKSON DOS SANTOS PASSOS**

**FILTRAGEM COLABORATIVA DE SISTEMA DE RECOMENDAÇÃO – Uma abordagem unificada de Interface-Homem-Máquina com Machine Learning usando Apache Mahout**

FORTALEZA - CE

2017

ANTONIO JACKSON DOS SANTOS PASSOS

**FILTRAGEM COLABORATIVA DE SISTEMA DE RECOMENDAÇÃO – Uma abordagem unificada de Interface-Homem-Máquina com Machine Learning usando Apache Mahout**

Artigo científico apresentado ao curso de Sistema de Informação da Faculdade Metropolitana da Grande Fortaleza – FAMETRO – como requisito para obtenção do grau de bacharel, sob a orientação do profº Francisco Nauber Bernardo Gois

Fortaleza - CE

2017

ANTONIO JACKSON DOS SANTOS PASSOS

**FILTRAGEM COLABORATIVA DE SISTEMA DE RECOMENDAÇÃO – Uma abordagem unificada de Interface-Homem-Máquina com Machine Learning usando Apache Mahout**

Este artigo científico foi apresentado no dia 00 de mês de 2017 como requisito para obtenção do grau de bacharel em Sistemas de Informação da Faculdade Metropolitana da Grande Fortaleza – FAMETRO – tendo sido aprovada pela banca examinadora composta pelos professores abaixo:

BANCA EXAMINADORA

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Profº Esp.

Coordenador – Faculdade Metropolitana da Grande Fortaleza

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Profº Esp.

Coordenador – Faculdade Metropolitana da Grande Fortaleza

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Profº Esp.

Coordenador – Faculdade Metropolitana da Grande Fortaleza

**AGRADECIMENTOS**

A Deus, por acreditar que nossa existência pressupõe outra infinitamente superior, dando-me o discernimento e sabedoria para alcançar meus objetivos.

Aos familiares e amigos, que com muito carinho e apoio, contribuíram de forma indireta e diretamente para que eu chegasse a esta etapa de minha vida.

Ao Coordenador, Júlio Alcântara Tavares, e todos os professores do Curso de Sistema de Informação da Fametro, em especial ao Doutorando Nauber Gois, pelo convívio, pelo apoio, por aceitar me orientar neste artigo de fundamental importância em minha vida acadêmica, pela compreensão e pela amizade.

O melhor resultado acontece quando todos em um grupo fazem o melhor por si próprios e pelo grupo.

John Nash

**FILTRAGEM COLABORATIVA DE SISTEMA DE RECOMENDAÇÃO – Uma abordagem unificada de Interface-Homem-Máquina com Machine Learning usando Apache Mahout**

**Antonio Jackson dos Santos Passos1**

**RESUMO**

Este artigo refere-se a como ajudar as pessoas a viverem melhor, economizando tempo e elevando o nível de satisfação da experiência do usuário, auxiliando-o na tomada de decisão em seu cotidiano com acesso a sistemas de recomendações, bem como associar esse tipo de sistema com a performance da interação humano-computador e seus principais benefícios no nível de satifsfação.

Para fundamentar a problematização deste documento, foi escolhido a filtragem colaborativa em sistemas de recomendações usando a plataforma apache mahout, um ambiente para criação rápida de aplicações escaláveis e automáticas para machine learn, com implementação na linguagem java, onde o mesmo busca criar recomendações para um usuário alvo, baseado em experiências e declarações correlacionadas de outros usuários de mesma similaridade com o usuário alvo. Ao fim da execução desse tipo de sistema, é possível avaliar o grau de importância no impacto da experiência do usuário, bem como as estratégias de mercado para explorar.

**SUMÁRIO**

[1. INTRODUÇÃO 7](#_Toc483570316)

[2. SISTEMAS DE RECOMENDAÇÃO E SEUS CONCEITOS 9](#_Toc483570317)

[2.1 Introdução aos Sistemas de Recomendação 9](#_Toc483570318)

[2.2 Definição da Filtragem Colaborativa 10](#_Toc483570319)

[2.3 Classificação da Informação 12](#_Toc483570320)

[2.3.1 Classificação Explícita 12](#_Toc483570321)

[2.3.2 Classificação Implícita 12](#_Toc483570322)

[3. IMPLEMENTAÇÃO DA BIBLIOTECA MAHOUT COM DADOS REAIS 12](#_Toc483570323)

[3.1 Perguntas de pesquisa experimental 12](#_Toc483570324)

[3.2 Variáveis 13](#_Toc483570325)

[3.3 Hipóteses 13](#_Toc483570326)

[3.4 Ameaças a validade do experimento 13](#_Toc483570327)

[4. RESULTADOS 13](#_Toc483570328)

[5. CONCLUSÃO 13](#_Toc483570329)

[9. REFERÊNCIAS 14](#_Toc483570330)

# 1. INTRODUÇÃO

A partir de pesquisas militares no auge da Guerra Fria, na década de 1960, surge a rede mundial dos computadores, ou internet. Desde então, com os avanços tecnológicos, surge o princípio da era da informação, porém pode-se afirmar que já vivemos a chamada, era da recomendação (FERREIRA, 2007); Nesse momento, as exigências de mercado provém de usuários e consumidores que julgarão a qualidade e satisfação da pesquisa feita pelos mesmos em sites de busca na internet, como por exemplo o Google¹, que facilita diversas ações no cotidiano das pessoas que consequentemente encontram uma diversidade muito grande de conteúdo, dando origem a um fenômeno chamado sobrecarga de informação (KIELGAST et.al, 1995).

No mundo pós-internet surge a necessidade de criação de novas ferramentas que auxiliem os usuários e projetistas de sistemas, a nortear a navegação para o destino procurado sem causar frustações ou arrependimentos. Em mecanismos de busca, usuários precisam ter uma ideia mínima do que deseja encontrar (*“push” de informação)*, porém essa nova área de pesquisa despertou interesse de pesquisadores independentes em meados de 1990 surgindo assim os Sistemas de Recomendação (SR´s) que trabalham com a perspectiva de descoberta, onde a informação é servida ao usuário (“*pull” de informação),* possibilitando-o sensações inesperadas e experiência personalizada.

O tempo é cada vez mais precioso para pessoas do século XXI, que buscam otimizar o tempo de suas atividades cotidianas e consequentemente deparam-se com o paradoxo da escolha (SCHWARTZ et.al, 2007), gerando ansiedade e confusão, dificultando a escolha da melhor opção. Perguntas tais como: Como investir meu dinheiro para garantir um futuro melhor para meus filhos? Qual o melhor destino de férias? Que filme posso escolher para assistir com meus amigos? Que livro comprar? Que curso devo escolher para minha carreira? Enfim, a todo instante é preciso efetuar decisões, buscando sempre a melhor escolha, aquela que satisfaça o usuário, daí encontra-se a fundamentação para utilização dos SR´s. Tal descoberta, almeja permitir ao usuário desse sistema, ter contato com uma gama de informações não convencionais porém com uma grande relevância para o indivíduo. Por exemplo, ao fazer uma pesquisa na internet, os resultados obtidos seguem o requisito de popularidade, mas nada garante que essa característica seja relevante para o usuário, que dificilmente irá ter acesso às informações menos populares na rede.

Sistemas de recomendação são amplamente usados por lojas virtuais, sites de filmes e livros e todo tipo de comércio eletrônico na web bem como redes sociais com recomendações de pessoas, amizades, grupos de trabalho, empregos entre outros. O primeiro web site a utilizar o SR em grandes proporções foi o My Yahoo, em julho de 1996 (MANBER et.al, 2000) Atualmente, um grande número de websites detém de vantagens dos SR´s, utilizando estratégias de customização para seus clientes como por exemplo, a Amazon.com, Netflix, iTunes Store, Youtube, Facebook, Twitter e Linkedin, tendo por finalidade realizar predições quanto à preferência dos seus usuários, aumentando suas margens de vendas e criando um diferencial competitivo no mercado em relação aos concorrentes.

Para fazer recomendações, o que é preciso é encontrar pessoas com preferências e gostos parecidos e cruzar as escolhas para fazer recomendações baseando-se nessas informações. Não se resume apenas a mostrar mais do que você já gosta, mas oferecer a melhor experiência com base em quem aquela pessoa é. Ou seja, as recomendações devem ser similares, mas não iguais. (Antonio Molins - Engenheiro espanhol que atuou no sistema de recomendação da NetFlix)

Atualmente existem três técnicas principais para implementação de um sistema de recomendação. A primeira é baseada em filtragem de conteúdo, que utiliza a correlação entre o perfil do usuário, sendo necessário adquirir dados pessoais do mesmo, e os itens recomendados. A segunda, baseada na filtragem colaborativa, analisa a classificação prévia dos utilizadores para estabelecer uma associação entre itens e usuários, que ao contrário da primeira, poderá recomendar itens de difícil acesso ou mesmo indisponíveis. A terceira, também denominada de filtragem híbrida, é o resultado da combinação de duas ou mais técnicas com objetivo de otimizar seus desempenhos. A utilização de SR tem sido uma abordagem eficiente para reduzir o esforço do usuário em encontrar informações de seu interesse.

Este documento foi estruturado da seguinte forma: na seção 2, começa-se por introduzir os principais conceitos dos sistemas de recomendação, fundamentando a problematização; na seção 3, apresenta-se a implementação do sistema utilizando a biblioteca mahout com dados reais; na seção 4, mostra-se os resultados da aplicação com uma pesquisa experimental qualitativa e por fim na seção 5, aborda-se a conclusão do trabalho.

# 2. SISTEMAS DE RECOMENDAÇÃO E SEUS CONCEITOS

Esta seção aborda os principais conceitos mais implementados em sistemas de recomendação (SR), bem como a função adotada para possibilitar a recomendação ao usuário. Ao fim da mesma é fundamentada a problematização e as técnicas existentes para classificar a informação.

## 2.1 Introdução aos Sistemas de Recomendação

A predição das preferências do usuário ou clientes, gerou uma disputa muito ampla para a maioria dos web sites na rede, significando um diferencial competitivo e agregando valor considerável para ambas as partes do processo. Isso é possível por causa da utilização dos sistemas de recomendação que tem como principal objetivo analisar dados dos utilizadores, extraindo informações preciosas e relevantes, para recomendar itens, produtos, filmes, músicas, pessoas, carreiras profissionais, cursos, profissões, viagens, enfim tudo que é possível recomendar para uma pessoa.

Para realizar predições precisas, esse tipo de sistema utiliza diversas técnicas baseadas em algoritmos da inteligência artificial chamados de *machine learning* ou aprendizado de máquina. Estes podem levar em consideração dois principais aspectos de análise dos dados: o perfil de cada usuário ou o item que deseja recomendar. Dentre estas considerações, destaca-se três principais técnicas de filtragem sendo elas de conteúdo, colaborativa e híbrida das quais adiante neste documento é explanado suas peculiaridades.

Pelo fato desse tipo sistema auxiliar empresas do comércio *online* a estabelecerem uma relação íntima e personalizada para seus clientes, é que se explica a popularidade e crescente interesse de acadêmicos em incentivar pesquisas na área a fim de melhorias dos algoritmos e suas precisões, tanto é que o início do termo, Sistema de Recomendação ou *Recommender Systems,* surge em meados dos anos 1990 como área de pesquisa independente com a criação do sistema *Tapestry*, que foi o primeiro sistema de recomendação comercial, onde classificava como “bom” ou “mal” as anotações feitas pelos usuários de seus próprios e-mails. Este sistema utilizava algoritmos de colaboração e de conteúdo, tendo dessa forma diversas desvantagens como por exemplo, a falta de capacidade de agrupamento de usuários por interesses similares, pois o mesmo não teria um método capaz de tal feito (GOLDBERG et.al, 1992).

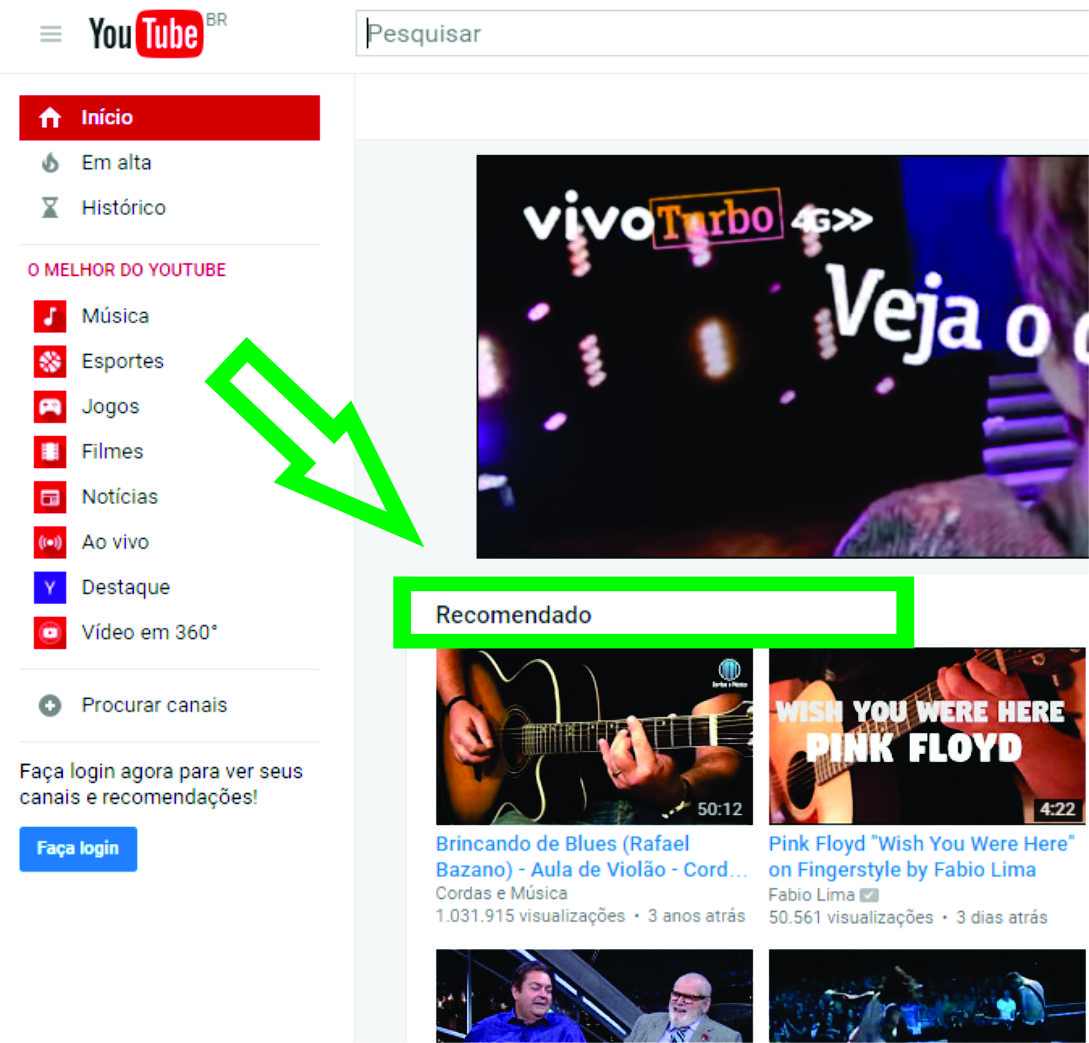
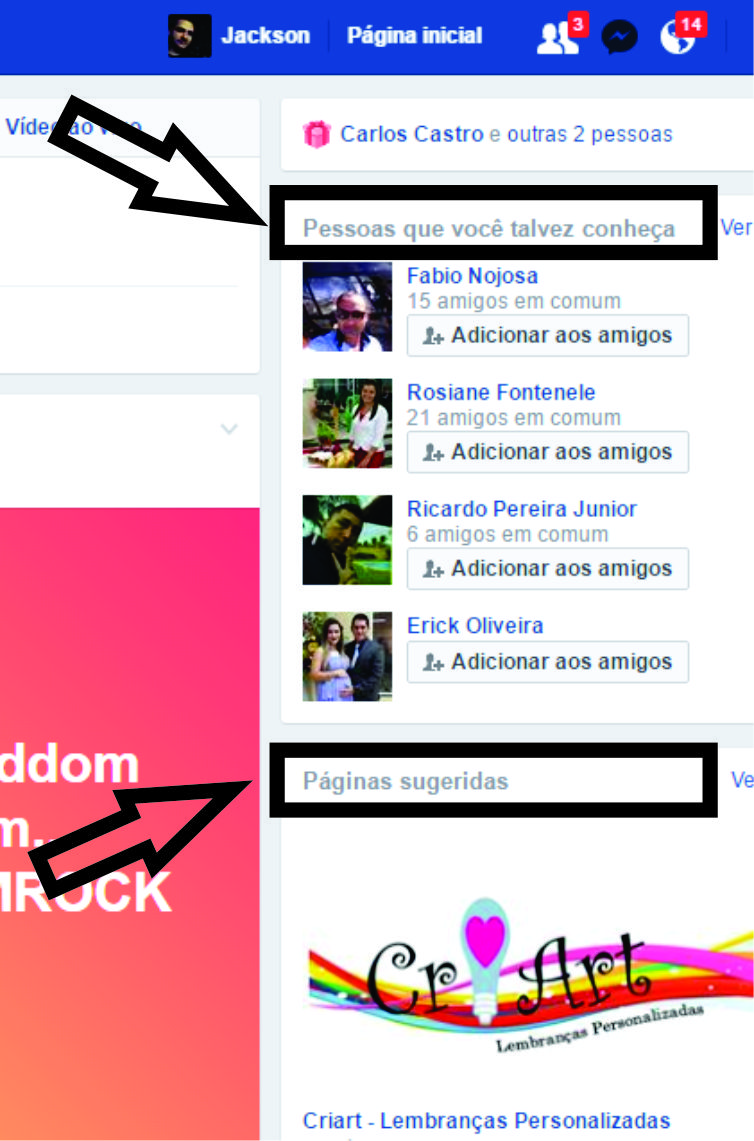
Com o crescimento perceptível de entretenimento e redes sociais na rede mundial de computadores, é mais fácil identificar no dia-a-dia o sistema de recomendação na prática como é o caso das imagens a seguir, que servem como reforço para o entendimento da aplicação dos SR´s. Na figura 1, é representado como o *Youtube.com* recomenda vídeos para o usuário com base em categorias que o mesmo acessa ou que pessoas similares a ele também acessam. De forma semelhante trabalha o *Facebook.com*, apenas com a diferença que ao invés de recomendar vídeos, este o faz com pessoas, sugestões de novas amizades, ou até mesmo páginas web como mostra a figura 2.

Figura 1 (esquerda) – Exemplo de recomendação feita pelo serviço do *Youtube.com;*

Figura 2 (direita) – Exemplo de recomendações da página do *Facebook.com*

## 2.2 Definição da Filtragem Colaborativa

Para fundamentação da filtragem colaborativa, é preciso analisar melhor o problema geral dos sistemas de recomendação, considerando 𝐔 como o conjunto de todos os utilizadores de um sistema, e 𝐈 como o conjunto de todos os itens passivos de recomendações, tais como filmes, pessoas, ou músicas. Com o propósito de possibilitar uma recomendação, é considerado 𝑟 como a função medidora da utilidade/relevância de um item para um determinado usuário, representada por: 𝑟:𝑼×𝑰 →𝑹, onde 𝑹 é o conjunto real (MARCELINO, 2014).

Com esses conceitos embasados, pode-se entender melhor a forma utilizada pela técnica de recomendação por filtragem colaborativa que foi projetada para ter a capacidade de recomendar itens indisponíveis ou de difícil acesso aos usuários, pois toma como base a classificação prévia desses itens por outros usuários cujo perfil é similar ao do usuário ativo do qual se quer efetuar alguma recomendação. A tabela 1 apresenta o funcionamento da filtragem colaborativa, com um exemplo fictício de recomendação de um produto para o usuário Nauber; para isso procura-se outros usuários com gostos semelhantes, no caso, Daniel e Júlio já compraram produtos que Nauber também comprou (PROD2). Logo após recomenda-se a Nauber produtos que estes dois usuários (Daniel e Júlio) possuem, porém que Nauber ainda não adquiriu como o PROD1 e PROD5. O resultado final da recomendação toma como base o histórico de avaliações comuns entre os usuários e o valor de predição calculado.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **USUÁRIO** | **PROD1** | **PROD2** | **PROD3** | **PROD4** | **PROD5** | **PROD6** |
| DANIEL |  | **X** |  |  | **X** |  |
| JÚLIO | **X** | **X** |  |  |  |  |
| ANELINE |  |  | X | X | X |  |
| JACKSON |  |  | X |  |  |  |
| CONCEIÇÃO | X |  |  | X |  |  |
| NAUBER | **?** | **X** |  |  | **?** |  |

**Tabela 1 – Recomendação baseada em filtragem colaborativa. Fonte: elaborada pelo autor.**

A técnica de Filtragem colaborativa segue três passos: (1) calcular a similaridade ao usuários alvo; (2) selecionar os vizinhos mais próximos para considerar a predição; (3) ponderar as avaliações dos vizinhos com seus pesos. Esta técnica é conhecida como “*k-nearest-neighbor*” ou “*user-based*”. Existem alguns coeficientes de similaridade como *Coseno, Pearson*, entre outros. A equação a seguir mostra o coeficiente de *Pearson*, pois é amplamente aplicado em literaturas de sistemas de recomendação (CAZELLA et.al, 2005), sendo a função utilizada pela biblioteca mahout, que será implementada mais à frente.

onde é a correlação do usuário ***a***alvo com um usuário ***b***; e são avaliações atribuídas a um item de cada usuário ***a*** e ***b*** respectivamente; e são as médias de todas as avaliações dos usuários a e b respectivamente; É importante destacar que é necessária mais de uma avaliação em comum para que a recomendação seja bem efetivada (CAZELLA et.al, 2015). A média ponderada das avaliações dos vizinhos com coeficiente de similaridade desejado, passado por parâmetro, resulta na predição de um item para o usuário alvo.

## 2.3 Classificação da Informação

Existem dois modos de extrair, do usuário, as informações necessárias para realizar uma recomendação: explicitamente e implicitamente (PAPAGELIS et.al, 2005). Estes métodos acabam provendo a discussão da questão da privacidade do usuário que pode sentir-se invadido e enfadado dependendo a abordagem utilizada. A seguir é explicado as duas técnicas de mineração das informações.

### **2.3.1 Classificação Explícita**

A obtenção dos dados, realizada de forma explícita, pode ser feita através de questionários ou da classificação por intervalo fixo (exemplo: de 1 a 5). Essa é a forma na qual o usuário optam por expor suas recomendações de um determinado item, de forma natural, por vontade própria. Esse tipo de abordagem é o que melhor se adequa aos sistemas de filtragem colaborativa (PAPAGELIS et.al, 2005), porém exige do usuário uma participação ativa no sistema onde quanto mais ele interagir, mais o sistema vai conhecê-lo, fator que nem sempre é visto de forma positiva pelos mesmos. Para minimizar o impacto negativo ao usuário, são utilizadas estratégias de atribuições de pontos para adquirir produtos grátis, ou participação de sorteios para acesso a conteúdo privilegiado (EKSTRAND, 2010).

### **2.3.2 Classificação Implícita**

Esse tipo de abordagem vem suprir o problema do usuário não querer classificar ou recomendar nada. A estratégia infere os gostos dos usuários por meio de histórico da interação do mesmo com o sistema, como por exemplo, histórico de compras, movimento e cliques de mouses, pesquisas realizadas, entre outros. Porém para essa estratégia há um risco grande de imprecisão, pelo fato dessas ações poderem não ser fiéis ao desejo do utilizador (OARD et.al, 1998), por exemplo, o fato de alguém clicar em uma notícia, não significa dizer que vá ler, ou o fato de fazer uma pesquisa, não quer dizer que represente o interesse real daquele utilizador, ele pode estar realizando pesquisas para outra pessoa.

# 3. IMPLEMENTAÇÃO DA BIBLIOTECA MAHOUT COM DADOS REAIS

## 3.1 Perguntas de pesquisa experimental

## 3.2 Variáveis

## 3.3 Hipóteses

## 3.4 Ameaças a validade do experimento

# 4. RESULTADOS

# 5. CONCLUSÃO

# 9. REFERÊNCIAS

Aletéia Ferreira, “Cybermarketing e a moda: segmentação através dos blogs e orkut”, 2007.

Soeren Kielgast, Bruce A. Hubbard, “Valor agregado à informação – da teoria à prática”, 1995.

Schwartz, Barry. "O paradoxo da escolha: por que mais é menos." São Paulo: A Girafa Editora (2007).

Manber, U.; Patel, A.; Robison, J. (2000). Experience with Personalization on Yahoo!

Communication of the ACM, New York.

D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, “Using Collaborative Filtering to Weave an Information Tapestry”, 1992.

Marcelino. Vasco Ferreira. “Sistema de Recomendação - Filtragem Colaborativa”, 2014.

Reategui, E.; Cazella, S. (2005) Sistemas de Recomendação. Minicurso do IV Encontro Nacional de Inteligência Artificial (ENIA). São Leopoldo, p. 306-348.

Reategui, E.; Cazella, S. “A ciência da opinião: estado da arte em sistemas de recomendação”, (2015).

Papagelis, M., & Plexousakis, D. Qualitative analysis of user-based and item-based prediction algorithms for recommendation agents, 2005.

Ekstrand, M. D. Collaborative Filtering Recommender Systems, 2010.

Douglas W. Oard and Jinmook Kim, “Implicit Feedback for a Recommender Systems”, 1998.

https://cursos.alura.com.br/course/machine-learning-introducao-aos-sistemas-de-recomendacoes