

**FIAP – FACULDADE DE INFORMÁTICA E ADMINISTRAÇÃO PAULISTA**

BRUNO BIANCCHI – RM 84351

DOUGLAS ANDRADE SANTOS FILHO – 87004

LUIS HENRIQUE CALDAS ALTERO – RM 88670

PEDRO GUILHERME POLLONI BARRETO - RM 88964

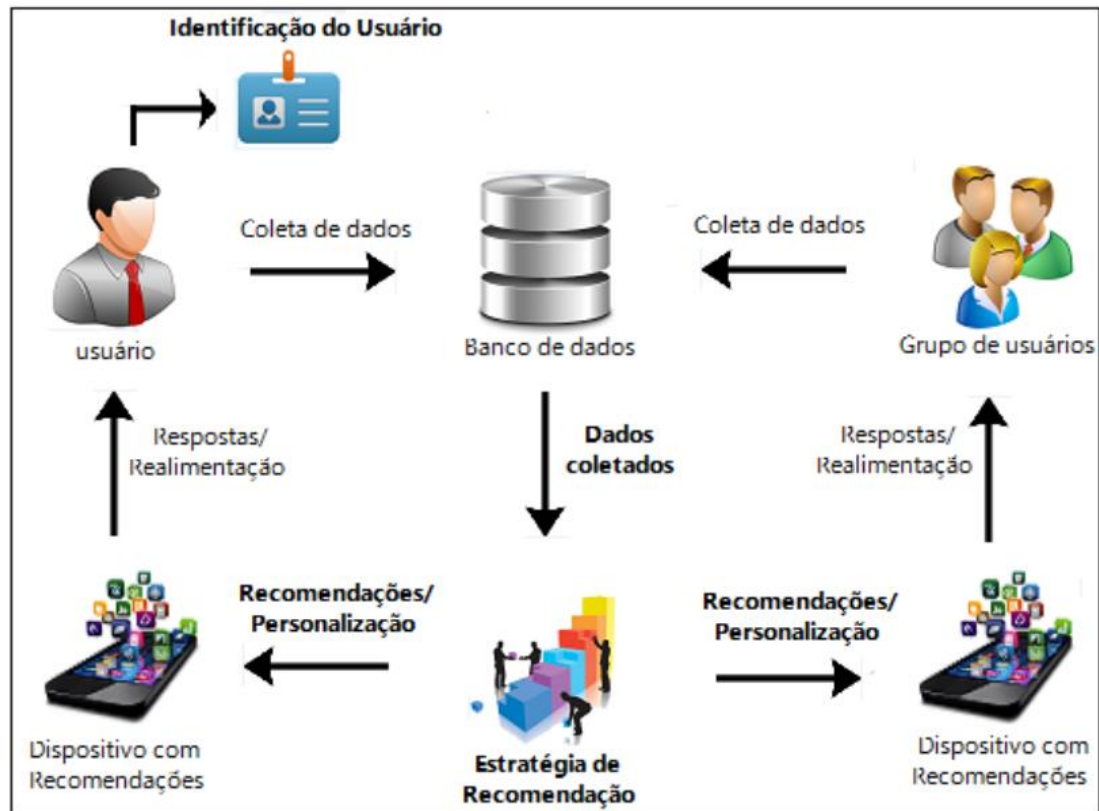
VICTOR LAMPRECHT – RM 86691

**Global Solution – Statistics for Data Science e Machine Learning**

**São Paulo**

**2021**

## Arquitetura Sistema de Recomendação Inteligente



**Figura 1**

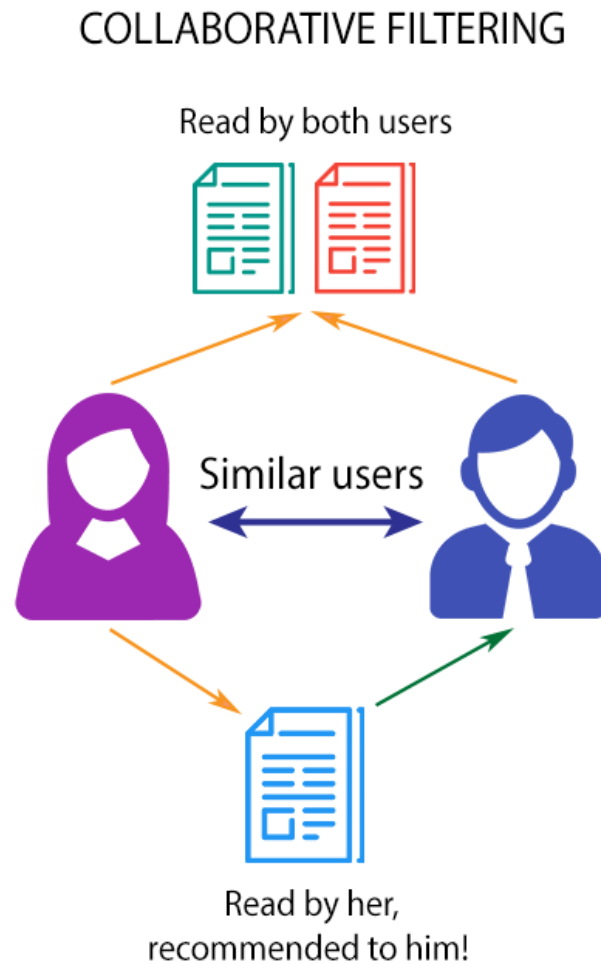
**Fonte:** 2016 TCC, Alexandre Mendonça Lima

O primeiro passo a se fazer é gerar uma chave primária para o usuário dentro do Banco de Dados, para que seus dados de pesquisa sejam coletados e armazenados de forma segura (minimizando a possibilidade de erros). Após o armazenamento correto dos dados, será utilizado um filtro de recomendação definido no sistema que enviará ao usuário a recomendação feita pelo sistema, sendo essa recomendação podendo ser feita também a um grupo de usuários, que funcionará da mesma forma.

## Algoritmos de Sistema de Recomendação Inteligente

Os dois tipos mais populares de filtros de recomendação com algoritmos são:

**Filtros colaborativos:** Esse tipo de filtro utiliza a similaridade entre os gostos dos usuários para definir uma certa recomendação a cada usuário. Por exemplo, se dois usuários A e B tiverem gostos similares e apenas o usuário A assistiu ao filme X, o algoritmo recomendará o filme X ao usuário B, como visto na imagem abaixo.

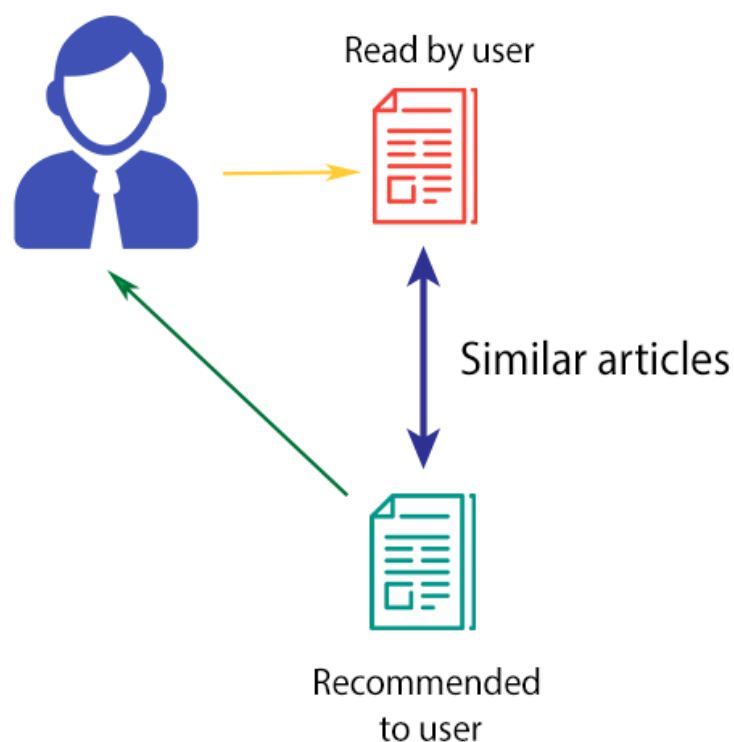


**Figura 2**  
**Fonte:** Internet

Os algoritmos utilizados para esse tipo de filtro são: KNN (o computador irá dizer a qual determinado grupo um dado faz parte), Cosine Similarity (o computador irá dizer qual a similaridade de cossenos) e o Slope one (computa uma avaliação de um determinado item e a similaridade desse item perante outros).

**Filtros baseados em conteúdo:** Esse tipo de filtro utiliza a similaridade entre os itens (suas palavras chaves ou tags) para que a recomendação seja feita, sendo indicado para empresas com produtos com várias características. Por exemplo, se um usuário A assistiu um filme X com as palavras chaves Terror e suspense, o algoritmo irá recomendar filmes similares com as mesmas palavras chaves, como visto na imagem abaixo.

## CONTENT-BASED FILTERING



**Figura 3**  
**Fonte:** Internet

Os algoritmos utilizados para esse tipo de filtro são: K-Means (é um algoritmo de mineração de dados supervisionado, isso significa que ele “aprende” uma regra segundo exemplos de entrada e saída fornecidos anteriormente, criando uma classificação de informações segundo os dados em questão), NLP (é um algoritmo de Word Embeddings que consegue modelar a similaridade das palavras de uma forma semântica).

Existe também a **filtragem híbrida**, que faz a junção das duas anteriores, acarretando numa assertividade mais elevada. Isso ocorre devido aos gostos pessoais de um usuário A serem diferentes de um usuário B dentro de um assunto X que é agradável aos dois, fazendo com que a junção dos dois métodos acima seja necessária para uma recomendação mais precisa.

Os algoritmos utilizados para esse tipo de filtro são os mesmos dos anteriores, já que esse filtro consiste na junção da filtragem baseada em conteúdo e da colaborativa.

Existem várias bibliotecas em Python para a recomendação inteligente e neste texto iremos focar em dois deles, o Surprise e o Crab.

O Surprise é uma biblioteca de sistema de recomendação que faz modelos de unsupervised learning (tipo de algoritmo que aprende padrões a partir de dados não marcados), sendo muito parecido com o Scikit-learn, porém sendo especializado em modelos de recomendação com dados de classificação explícitos.

Já o Crab é uma biblioteca de sistema de recomendação flexível e rápida focada na filtragem colaborativa, contudo suportando a filtragem híbrida também.

### **Quais análises poderiam ser realizadas nos dados da tabela `data_collaborative.csv` e `data_content.csv`?**

Levando em conta as análises simples feitas anteriormente, os dados das tabelas poderiam ser filtrados baseados em uma forma híbrida para garantir uma maior assertividade. Com isso utilizamos o filtro colaborativo para detectar gostos em comum de determinados usuários (como a categoria mostrada no **`data_content.csv`**), unindo-os em grupo, e separando-os dentro dos grupos com a filtragem baseada em conteúdo (separando as categorias de acordo com os títulos dentro das mesmas no **`data_content.csv`**), garantindo maior exatidão ao algoritmo de recomendação.

## Referências

- <https://www.supero.com.br/blog/o-que-sao-sistemas-de-recomendacao/>
- <http://www.each.usp.br/petsi/jornal/?p=2684>
- <https://blog.geekhunter.com.br/algoritmos-de-recomendacao-o-que-sao-e-como-implementa-los/>
- <https://dadosaocubo.com/sistemas-de-recomendacoes-com-surprise/>
- <https://medium.com/@maricatovictor/sistemas-de-recomendação-surprise-5e49e62a304e>
- [https://github.com/NicolasHug/Surprise/blob/master/examples/k\\_nearest\\_neighbors.py](https://github.com/NicolasHug/Surprise/blob/master/examples/k_nearest_neighbors.py)
- <https://github.com/muricoca/crab>
- <https://medium.com/analytics-vidhya/collaborative-based-recommendation-system-using-svd-9adc5b6b3b8>
- <https://medium.com/analytics-vidhya/collaborative-based-recommendation-system-using-svd-9adc5b6b3b8>
- <https://www.google.com/imgres?imgurl=x-raw-image%3A%2F%2F%2F8965d002a0093cd6e83261ba87614d1c9d92b898088c7514ce9d2432bfbbe563&imgrefurl=https%3A%2F%2Fwww.sigaa.ufs.br%2Fsigaa%2FverProducao%3FidProducao%3D1155155%26%26key%3D526226b692e22bc1c9ecd9912f38b8f1&tbnid=1jwJoALQZYm5xM&vet=12ahUKEwifrJfGp4bxAhV2OLkGHSbSA1kQMygBegUIARCjAQ..i&docid=kxxPXqNLikv0QM&w=538&h=400&q=arquitetura%20sistema%20de%20recomenda%C3%A7%C3%A3o&client=opera-gx&ved=2ahUKEwifrJfGp4bxAhV2OLkGHSbSA1kQMygBegUIARCjAQ>
- [https://inferir.com.br/artigos/algoritmo-knn-para-classificacao/#:~:text=O%20algoritmo%20KNN%20\(K%20Nearest,analizando%20seu%20processo%20de%20cálculo.&text=Na%20classificação%20a%20máquina%20irá,de%20um%20contexto%20de%20negócio.](https://inferir.com.br/artigos/algoritmo-knn-para-classificacao/#:~:text=O%20algoritmo%20KNN%20(K%20Nearest,analizando%20seu%20processo%20de%20cálculo.&text=Na%20classificação%20a%20máquina%20irá,de%20um%20contexto%20de%20negócio.)
- <https://www.devmedia.com.br/data-mining-na-pratica-algoritmo-k-means/4584>
- <https://www.ti-enxame.com/pt/python/calcular-semelhanca-de-cosseno-com-duas-cadeias-de-frases/1070875781/>
- <https://medium.com/@analisabavati/filtragem-colaborativa-utilizando-slope-one-173a536607e3#:~:text=Trata-se%20de%20um%20algoritmo,similaridade%20entre%20i%20e%20outros%20itens.&text=O%20algoritmo%20opera%20com%20notas,avaliadas%20pelos%20usuários%20a%20itens.>

<https://sites.northwestern.edu/msia/2018/03/16/movie-recommender-system-based-on-natural-language-processing/>

<https://medium.com/data-hackers/deep-learning-para-sistemas-de-recomendação-parte-3-recomendação-por-similaridade-d788c126d808>