



Daniel Guilherme da Silveira Filho

**Sistema de recomendação de restaurantes,
baseado em três tipos diferentes de filtragem
de dados, nos bairros do Recife–PE**

**Recife
2021**

Daniel Guilherme da Silveira Filho

**Sistema de recomendação de restaurantes, baseado em
três tipos diferentes de filtragem de dados, nos bairros
do Recife-PE**

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Estatística e Informática

Curso de Bacharelado em Sistemas de Informação

Orientador: Gilberto Amado de Azevedo Cysneiros Filho

Recife

2021

Daniel Guilherme da Silveira Filho

Sistema de recomendação de restaurantes, baseado em três tipos diferentes de filtragem de dados, nos bairros do Recife-PE

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em: 20 de Dezembro de 2021.

BANCA EXAMINADORA

Gilberto Amado de Azevedo Cysneiros Filho (Orientador)
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Elizabeth Regina Tschá
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Cícero Garrozi
Departamento de Estatística e Informática
Universidade Federal Rural de Pernambuco

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

- F481s Filho, Daniel Guilherme da Silveira
Sistema de recomendação de restaurantes, baseado em três tipos diferentes de filtragem de dados, nos bairros do Recife - PE / Daniel Guilherme da Silveira Filho. - 2021.
45 f. : il.
- Orientador: Gilberto Amado de Azevedo Cysneiros Filho Orientador .
Inclui referências, apêndice(s) e anexo(s).
- Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Sistemas da Informação, Recife, 2021.
1. Sistemas de Recomendação. 2. Restaurantes de Recife. 3. Polo Gastronômico. I. Orientador ,
Gilberto Amado de Azevedo Cysneiros Filho , orient. II. Título

*A minha família que me apoiou sempre nessa jornada confusa e incerta,
principalmente meus pais que me ajudaram a sedimentar os atributos necessários
ao caminho e a minha esposa que está sempre comigo.*

AGRADECIMENTOS

Agradeço a todos que fazem possível o curso de Graduação em Sistemas de Informação da Universidade Federal Rural de Pernambuco, que conseguiu superar até mesmo a pandemia da COVID-19, para que sonhos possa se tornar realidade.

“O sucesso nasce do querer, da determinação e persistência em se chegar a um objetivo. Mesmo não atingindo o alvo, quem busca e vence obstáculos, no mínimo fará coisas admiráveis.” (José de Alencar)

RESUMO

Tendo em vista que a cidade de Recife está na terceira colocação de polos gastronômicos do Brasil, e que atualmente a quantidade de informações, presentes na internet, beira o infinito, os turistas, e até população local, procuram àqueles restaurantes mais conhecidos e divulgados. Uma consequência disso é que pequenos estabelecimentos, que não possuem uma forte divulgação de sua marca, acabam por falir nos primeiros cinco anos de existência. Por isso, realizou-se uma pesquisa sobre sistemas de recomendação, a fim de analisar o desenvolvimento de um sistema de recomendação de restaurantes, baseado em três tipos diferentes de técnicas de filtragem de dados, nos bairros da cidade do Recife. Para tanto, é necessário explicar os fundamentos das técnicas de filtragem que serão utilizadas no desenvolvimento do sistema, além de identificar como definir, coletar, tratar e analisar os dados necessários à construção do sistema, analisar o desenvolvimento de um sistema de recomendação baseado no conteúdo, analisar o desenvolvimento de um sistema de recomendação, baseado no usuário, analisar o desenvolvimento de um sistema de recomendação híbrido e por fim identificar qual técnica de recomendação retorna melhores resultados. Realiza-se então, um estudo sobre a base de dados, seleção, coleta e tratamento dos dados necessários para a construção do sistema, além de modelar uma amostra de pessoas e restaurantes a serem recomendados, também foi feito um estudo sobre sistemas de recomendação, definição, surgimento e principais técnicas de filtragem, utilizadas em trabalhos semelhantes, com isso se fez necessário definir quais os atributos e parâmetros a serem coletados, além de uma modelagem do formulário de captura de dados e desenvolvimento das três técnicas de filtragem colaborativa selecionadas, e por fim, realiza-se testes do sistema para comprovar seu funcionamento e análise de dados coletados. Diante disso, verifica-se que apesar da técnica de filtragem baseada em conteúdo se destacar nos resultados, a diferença entre esse tipo de filtragem e as filtragens colaborativas não foi expressiva, o que impõe a constatação de que o sistema de recomendação, baseado por bairros, traz facilidade e praticidade aos usuários, além de divulgar a marca de restaurantes de bairro e capilarizar o turismo na cidade de Recife, porém necessita de mais interações e informações dos usuários cadastrados, para tornar-se mais preciso em suas recomendações.

Palavras-chave: Sistemas de recomendação. Restaurantes de Recife. Polo gastronômico.

ABSTRACT

Considering that the city of Recife is in third place for gastronomic centers in Brazil, and that currently the amount of information present on the internet is bordering on infinity, tourists and even the local population go to the most popular restaurants. A consequence of this is that small accessories, which do not have a strong publicity for their brand, end up going out of business in the first five years of existence. Therefore, a research was carried out on recommendation systems, an analysis purpose for the development of a restaurant recommendation system, based on three different types of data filtering techniques, in the neighborhoods of the city of Recife. Therefore, it is necessary to explain the fundamentals of the filtering techniques that will be used in the development of the system, in addition to identifying how to define, collect, treat and analyze the data provided for the construction of the system, analyze the development of a recommendation-based system in the content, analyze the development of a recommendation system, based on the user, analyze the development of a hybrid recommendation system and finally identify which recommendation technique best results. A study is then carried out on the database, selection, collection and processing of data provided for the construction of the system, in addition to modeling a sample of people and results to be recommended, a study was also carried out on recommendation systems, definition, emergence and main filtering techniques, used in similar works, thus it was necessary to define which attributes and parameters to be collected, in addition to a modeling of the data capture form and development of the three selected collaborative filtering techniques, and finally, system tests are carried out to prove its functioning and analysis of collected data. Therefore, it appears that although the content-based filtering technique stands out in the results, the difference between this type of filtering and collaborative filtering was not significant, which imposes the observation that the recommendation system, based on neighborhoods, brings ease and convenience to users, in addition to promoting the brand of neighborhood restaurants and expanding tourism in the city of Recife, but it requires more interaction and information from registered users, to become more precise in their recommendations.

Keywords: Recommendation systems. Recife restaurants. Gastronomic pole.

LISTA DE ILUSTRAÇÕES

Figura 1 – tipos de técnicas de filtragem	14
Figura 2 - divisão bairro de Afogados	20
Figura 3 - bairro de Boa Viagem	20
Figura 4 - formulário campo e-mail	25
Figura 5 - formulário campo bairro mora.....	25
Figura 6 - formulário campo bairro costuma comer	26
Figura 7 - formulário campo seleção especialidade	27
Figura 8 - formulário classificação dos restaurantes	27
Figura 9 - formulário recomendações	28
Figura 10 - modelagem banco	40

LISTA DE TABELAS

Tabela 1 - tabela usuário x restaurante	31
Tabela 2 - resultados parciais	33
Tabela 3 - resultados finais	33

SUMÁRIO

1	INTRODUÇÃO	11
1.1	Apresentação	11
1.2	Motivações	11
1.3	Objetivos	12
1.4	Metodologia.....	13
1.5	Estrutura do Trabalho	13
2	SISTEMAS DE RECOMENDAÇÃO	14
2.1	Filtragem Baseada no conteúdo.....	15
2.2	Filtragem colaborativa.....	15
2.3	Filtragem Híbrida.....	16
3	COLETA E ANÁLISE DE DADOS	16
3.1	Definição dos dados de restaurante	17
3.1.1	Bairros	17
3.1.2	Especialidades	17
3.1.3	Restaurantes	18
3.2	Fonte de dados	18
3.3	Processo de coleta	19
3.4	Tratamento de dados.....	21
3.5	Definição de amostra	22
3.6	Formulário	23
3.6.1	Desenvolvimento do Formulário	23
3.6.2	Campos do Formulário.....	24
4	Recomendação Baseada em Conteúdo.....	29
4.1	Desenvolvimento.....	29
5	Recomendação colaborativa baseada no usuário	30
5.1	Desenvolvimento.....	30
6	Recomendação Híbrida	32
6.1	Desenvolvimento.....	32
7	Análise de Resultados	33

8	Considerações Finais.....	34
8.1	Limitações.....	35
8.2	Trabalhos Futuros.....	35
	Hospedagem do sistema de recomendação	37
	Interações com o servidor.....	38
	Banco de dados	39
	Apache Mahout	40

1 INTRODUÇÃO

Na idade média, poucas pessoas eram letradas ou tinham acesso às obras escritas, pois nessa época, a igreja católica concentrava as informações e formações das pessoas e o acesso à informação era altamente restrito [25]. Já no ano de 2011, por exemplo, pesquisadores, de uma universidade na Califórnia, estimaram que o mundo produziu cerca de 296 exabytes de informações entre os anos de 1986 e 2007, com isso conclui-se que o problema nos dias de hoje é o excesso de informações disponíveis e não a falta de informações ou restrições para acessá-la, como na idade média [26].

Segundo a Abrasel (Associação Brasileira de Bares e Restaurantes) e reforçado pela prefeitura, Recife é o primeiro polo gastronômico do Nordeste e o terceiro do país, com mais de 10.000 estabelecimentos em funcionamento, perdendo apenas para São Paulo e Rio de Janeiro [27] [28]. Todos os anos a gastronomia da cidade se destaca, através de eventos como o caso do Bar em Bar, Fliporto Gastronomia, Comida de Boteco, Restaurante Week, entre outros [29].

Com toda essa quantidade de restaurantes, presentes na cidade do Recife, fica difícil saber qual o que melhor atende as preferências do cliente. E para resolver esse tipo de problema, foram criados, na década de 90, os sistemas de recomendação [17].

1.1 Apresentação

Este trabalho se propõe a analisar o desenvolvimento de um sistema de recomendação de restaurantes, baseado em três técnicas de filtragem de dados, nos bairros da cidade do Recife. Tal análise possibilitará uma melhor opção na escolha de restaurantes pela população local e turistas, além de fortalecer a divulgação dos pequenos estabelecimentos, muitas vezes desconhecidos pela população.

1.2 Motivações

Os sistemas de recomendação são, hoje em dia, uma área de bastante estudo e atuação no mercado, grandes empresas investem bastante nessa área, para tornar a experiência do usuário cada vez melhor. Um exemplo disso foi a divulgação da Netflix, onde apresentou a métrica de que 80% dos filmes assistidos em sua

plataforma de streaming, são provenientes de seu sistema de recomendação [17]. Isso reflete diretamente na experiência que o usuário final tem na plataforma e por isso ela é, hoje em dia, uma das maiores do mundo.

Por outro lado, Recife é o primeiro polo gastronômico do Nordeste, segundo a própria prefeitura [27], e obteve um aumento de 175% em sua malha aérea no último ano [31]. Isso implica em um desenvolvimento do turismo da cidade e uma quantidade maior de turistas utilizando os serviços dos restaurantes da região. Porém muitos desses restaurantes estão em localizações afastadas dos principais pontos turísticos da cidade e isso colabora para que, esses restaurantes mais afastados, não recebam esses novos clientes.

Combinando esse mundo de sistemas de recomendação, que melhora cada dia mais a experiência do usuário final, com o primeiro polo gastronômico do Nordeste, pode-se concluir, que um equipamento turístico que possa, como o compositor pernambucano, Petrúcio Amorim na sua música Parte da Minha Vida diz: “juntar a fome com a vontade de comer...”, terá condições de capilarizar o turismo da cidade e trazer simplicidade e praticidade para o usuário final, na hora de escolher entre os inúmeros restaurantes, um que possua produtos e serviços que atendam melhor às suas preferências pessoais.

1.3 Objetivos

O objetivo deste trabalho é analisar o desenvolvimento de um sistema de recomendação de restaurantes, baseado em três tipos diferentes de técnicas de filtragem de dados, nos bairros da cidade do Recife.

E para isso será necessário:

- Explicar os fundamentos das técnicas de filtragem que serão utilizadas no desenvolvimento do sistema de recomendação de restaurantes.
- Identificar como definir, coletar, tratar e analisar os dados necessários à construção do sistema de recomendação.
- Analisar o desenvolvimento de um sistema de recomendação baseado no conteúdo.
- Analisar o desenvolvimento de um sistema de recomendação, baseado no usuário.
- Analisar o desenvolvimento de um sistema de recomendação híbrido.

- Identificar qual técnica de recomendação retorna melhores resultados.

1.4 Metodologia

No desenvolvimento deste trabalho serão realizadas as seguintes atividades:

- Aprendizado acerca de base de dados, seleção, coleta e tratamento;
- Modelagem de amostra de pessoas e restaurantes a serem utilizados;
- Aprendizado sobre sistemas de recomendação, definição, surgimento e técnicas de filtragem;
- Estudo de alguns trabalhos semelhantes;
- Definição de atributos e parâmetros a serem coletados;
- Modelagem do formulário de captura de dados;
- Desenvolvimento das três técnicas de sistemas de recomendação;
- Testes do sistema e análise dos dados coletados;

1.5 Estrutura do Trabalho

O capítulo 1 desenvolve a introdução e os objetivos do trabalho.

No capítulo 2 terá uma abordagem didática sobre a definição de sistemas de recomendação e as técnicas de filtragem de dados que serão utilizadas no desenvolvimento do sistema de recomendação de restaurantes.

O capítulo 3 abordará sobre a coleta e tratamento de todos os dados utilizados para o desenvolvimento do sistema de recomendação de restaurantes.

O capítulo 4 tratará da lógica de desenvolvimento do sistema de recomendação baseado no conteúdo.

O capítulo 5 versará sobre a lógica para o desenvolvimento do sistema de recomendação colaborativo, baseado no usuário.

Já no capítulo 6 será abordado a técnica da filtragem que envolve mais de uma técnica junta, a filtragem híbrida.

E por fim no capítulo 7, será realizado as considerações finais, limitações e sugestões de trabalhos futuros.

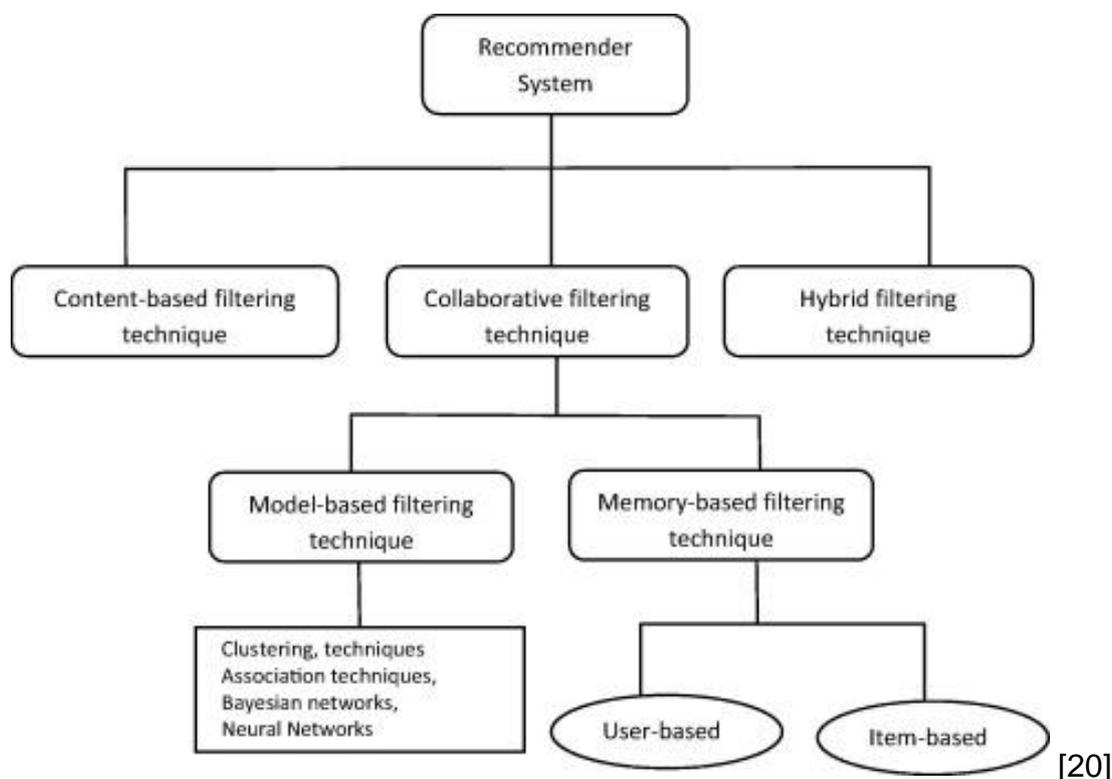
2 SISTEMAS DE RECOMENDAÇÃO

Os sistemas de recomendação iniciam sua jornada com o crescimento do uso do e-mail, por volta da década de 90, quando em Palo Alto, no centro de pesquisas da Xerox, foi criado o Tapestry, cuja finalidade era filtrar e-mails indesejados e retornar, apenas e-mails que correlacionassem com o perfil do usuário utilizador [17]. E ainda assim, nessa época a quantidade de informações ainda era escassa, comparada com o que temos hoje em dia.

Steve Jobs baseava as ações de marketing de seus produtos na premissa de que, as pessoas não sabem o que querem até alguém mostrar a elas [18]. Essa afirmação possui tanta veracidade, que o reflexo disso é a informação divulgada pela Netflix, citando que aproximadamente 80% dos filmes assistidos, em sua plataforma de streaming, derivam do sistema de recomendação que eles desenvolveram [19].

No geral, os sistemas de recomendação são técnicas que serão utilizadas para filtrar, dentre vários itens, àqueles que são mais relevantes à determinado usuário, conforme exposto na figura 1:

Figura 1 – tipos de técnicas de filtragem



Neste Capítulo será explicado o funcionamento dos três tipos de técnicas de filtragem utilizadas no sistema de recomendação analisado neste trabalho.

2.1 Filtragem Baseada no conteúdo

A técnica de filtragem baseada em conteúdo, como o próprio nome diz, refere-se à similaridade entre os itens a serem recomendados, para um usuário. Para isso utiliza-se atributos (metadados) de um determinado item para calcular o quão parecido esse item é em relação a outros, que o usuário já classificou dentre seus gostos. A fórmula matemática para que isso aconteça envolve uma matriz item x atributos dos itens, onde será calculada a similaridade daquele item classificado como preferido pelo usuário [21].

A principal vantagem desse tipo de técnica é que não se faz necessário ter mais de um usuário utilizando o sistema. Inclusive um único usuário que acabou de ser cadastrado no sistema, desde que ele indique pelo menos uma especialidade de restaurante que gosta, receberá uma recomendação, baseada em conteúdo [22].

Em contrapartida, tal técnica de filtragem, só considera o histórico do usuário, então não consegue recomendar um restaurante novo, de uma especialidade diferente de seu histórico, então a longo prazo essa recomendação perde sua eficácia, tal problema é chamado de superespecialização [21].

2.2 Filtragem colaborativa

O ser humano é um ser social e sempre procura a opinião de um amigo ou conhecido sobre determinado assunto ou produto no dia a dia. E essa opinião do outro, influencia a tomada de decisão e até a preferência dessa pessoa. Baseado nisso, foi criado o conceito de filtragem colaborativa cujo intuito é, a partir de um cabedal de conhecimento e experiências de uma pessoa, assim como opiniões do meio que a cerca, selecionar quais os itens que estão mais propensos a serem escolhidos de acordo com essas preferências, ou seja, a técnica da filtragem colaborativa, utiliza-se de conhecimentos coletivos para realizar uma recomendação [23].

Caso ela seja baseada no usuário, o foco será no usuário, ou seja, depende de informações acerca de um público semelhante ao usuário alvo (aquele que recebe a recomendação). No caso da baseada no item (neste trabalho = restaurante), o foco da análise será sempre os restaurantes e a busca por estabelecimentos semelhantes [21].

Conforme figura 1, existem duas formas de filtragem colaborativa, a baseada em modelo e baseada em memória. Será abordado nesse trabalho apenas a filtragem colaborativa baseada em memória, pois ela será uma das técnicas utilizadas no desenvolvimento do sistema de recomendação em questão.

A principal desvantagem dessa técnica é a falta de dados no início do sistema, pois quanto menos dados o sistema possui, maior é a esparsidade da matriz, impossibilitando uma recomendação precisa. Esse problema é chamado de cold start. Por outro lado, não possui o problema da superespecialização que se encontra na técnica de filtragem baseada em conteúdo, já que novos itens, de categorias diferentes, podem ser recomendados [22].

2.3 Filtragem Híbrida

Os sistemas de recomendação estão no mercado desde a década de 90, e apesar de seu amplo uso atualmente, algumas técnicas ainda apresentam dificuldades em suas recomendações. A filtragem híbrida é utilizada para atenuar as dificuldades encontradas nas técnicas de filtragem, e aproveitar as melhores qualidades de cada uma [23].

A hibridização dessas técnicas depende dos dados que o sistema possui como entrada, e baseado nesses dados é escolhido a melhor técnica para a recomendação.

3 COLETA E ANÁLISE DE DADOS

Nos dias de hoje o acesso aos dados está cada vez mais simples, a quantidade de informações as quais temos acesso beira a infinidade, e o problema a ser resolvido é: como saber filtrar o que é relevante?

Os sistemas de recomendação trabalham com o principal objetivo de filtrar, dentre infinitas informações e trazer aquilo que é relevante ao usuário. Mas para que isso possa ser possível, tais sistemas precisam de dados relevantes a recomendar.

Neste capítulo, será exposto o caminho percorrido na jornada da coleta inicial dos dados necessários ao sistema. Desde a definição dos dados, passando pelo tratamento deles e divisão de amostragem.

3.1 Definição dos dados de restaurante

Antes de iniciar uma busca de fontes de dados, primeiramente, se faz necessário definir quais dados serão relevantes e precisam ser encontrados.

Como o objetivo do sistema é a recomendação de restaurantes por bairros do Recife, a primeira busca a ser feita é: quais são os bairros mais relevantes ao sistema, posterior a isso, os restaurantes a serem recomendados e, como cada usuário tem sua preferência em relação as especialidades dos restaurantes, esse também foi considerado como um dado relevante ao sistema.

3.1.1 Bairros

A informação de bairros do Recife é fundamental para segmentar as recomendações por bairros, por isso buscou-se a relação de todos os bairros do Recife e quais deles seriam relevantes ao sistema, de acordo com o público que receberá as recomendações.

Tais dados estão disponíveis, de forma aberta, no site da prefeitura da cidade do Recife. E segundo o próprio site da prefeitura, a cidade de Recife registra 94 bairros [11], que dentre esses, foram selecionados 29 bairros, para compor essa pesquisa e segmentar as recomendações.

3.1.2 Especialidades

Foi realizada uma entrevista em uma amostra de 25 indivíduos, cuja finalidade era definir, dentro de um escopo de 11 especialidades, quais especialidades seriam de interesse do entrevistado e se ele recomendaria alguma outra fora da lista. O resultado foi a adição de 5 especialidades e a exclusão da especialidade comida mexicana. Com isso, foram definidas 18 especialidades dentre elas: **açaiteria, bar, bozeria, café, churrascaria, comida árabe, comida chinesa,**

comida italiana, comida regional, comida japonesa, doceria, frutos do mar, galeteria, hamburgueria, lanchonete, pizzaria, sorveteria, comida saudável.

3.1.3 Restaurantes

A definição dos dados de bairros e especialidades, são essenciais para refinar as recomendações, servirão como filtros antes de passar pelos algoritmos de recomendação. Todavia, os dados dos restaurantes é sem dúvida, a informação mais importante para que o sistema de recomendação faça sentido, afinal o sistema de recomendação exposto nesse trabalho, recomenda restaurantes.

Assim sendo, foram definidos nome, endereço, e classificação, como dados necessários ao sistema. Onde o nome é utilizado para identificação do restaurante pelo usuário, o endereço coleta o bairro em que o restaurante se encontra e a classificação será utilizada, para dirimir os problemas iniciais das recomendações colaborativas.

3.2 Fonte de dados

A fonte de dados ideal será sempre a que reúne confiabilidade de seus dados com as informações necessárias ao sistema. Pois a falta de confiança nos dados coletados, pode acarretar descrença por parte dos usuários que irão utilizar o sistema.

Para eleger a fonte ideal, listou-se as principais alternativas do mercado que preenchiam esses critérios. E embora o mercado apresente várias opções para coleta de dados de pontos de interesse, como por exemplo: Google Places, TripAdvisor, Kekanto e OpenStreetMap, a API escolhida, além de preencher o requisito de confiabilidade de dados, mostrou uma documentação clara, objetiva e de fácil entendimento, o qual facilitou a implementação do código de comunicação com a referida API e o sistema de recomendações.

Tomando-se esses três pontos como fatores determinantes na escolha da API, confiabilidade dos dados, documentação simples e possuir a maioria dos dados necessários à construção do sistema, a Google Places API foi escolhida como fonte de dados dos restaurantes.

3.3 Processo de coleta

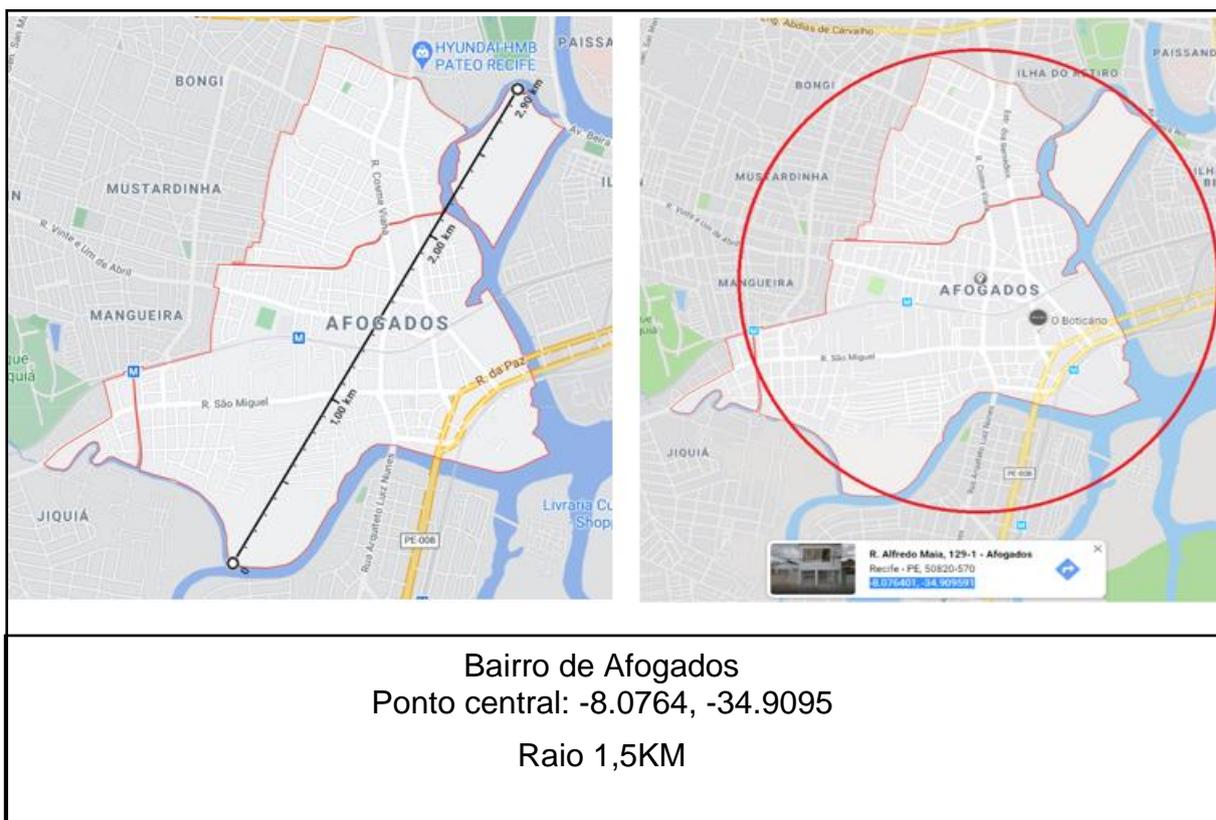
Definidos os dados a serem coletados e a fonte de dados, o próximo passo foi a coleta. A API do Google Places em sua documentação, define os tipos de locais, aos quais libera acesso aos dados [12], dentre esses foram selecionados aqueles que possuíam maior envolvimento com alimentação: bar, café, meal_takeaway e restaurant.

A comunicação com a API do Google Places, foi realizada por um programa, desenvolvido para esta coleta, que através de protocolo HTTP, enviava a requisição dos dados e em seu retorno recebia um JSON preenchido com os dados dos restaurantes. Porém a referida API, possui limitações no número de requisições, sendo 20 requisições por página e máximo de 3 páginas, ou seja, na época da coleta, existia uma limitação de 60 pontos de interesse por requisição. [7]

Antes de receber os dados, foi preciso montar a requisição, que seria enviada a API do Google Places, solicitando os dados. Para este fim, foi necessário definir um ponto central, em coordenada geográfica, e um raio de busca a partir daquele ponto, onde seriam mapeados os pontos de interesse selecionados daquela área. Por conta disso, se fez necessário o mapeamento dos bairros um a um, dentre os 29 bairros selecionados.

A definição do ponto central e o raio se deu, conforme exemplifica a figura 2. O ponto central foi descoberto a partir da média do traçado de uma linha, feito de uma ponta a outra do mapa do bairro, de modo que o raio de busca mapeado, contemplasse a maior parte do bairro.

Figura 2 - divisão bairro de Afogados



Em alguns bairros, como por exemplo boa viagem, foi necessário dividir em mais de um ponto central e realizar mais de uma requisição para a API do Google Places, como exemplifica a figura 3.

Figura 3 - bairro de Boa Viagem



Passados alguns dias de coleta, o resultado foi um JSON que possuía 2.912 restaurantes coletados, sendo 783 bares, 306 cafés, 489 meal_takeaway e 1334 restaurantes, com o seguinte formato:

```
{
  "id" : "ChIJ-Yjy0WofqwcRXGBVoDJBbVk",
  "name" : "The Gentleman Loser Pub",
  "localDescription" : "não definido",
  "address" : "R. Cap. Rebelinho, 527 - Pina, Recife - PE, 51011-010, Brasil",
  "geolocation" : {
    "coordinates" : "-8.0953238,-34.8841316",
    "type" : "Point"
  },
  "functionalHour" : ["segunda-feira: Fechado", "terça-feira: Fechado", "quarta-
feira: Fechado", "quinta-feira: 20:00 – 02:30", "sexta-
feira: 22:00 – 04:00", "sábado: 22:00 – 04:00", "domingo: Fechado"],
  "rate" : 4.3,
  "category" : {
    "name" : "1",
    "id" : "bar"
  },
  "urlPhoto" : "https://maps.googleapis.com/maps/api/place/photo?maxwidth=1
000&photoreference=CmRaAAAAWFNwnQ5WEQraw8CXiRGvHSEE-"
}
```

Todos esses dados foram coletados no mês de março de 2019, por isso podem existir divergências, comparando-se com os dados atuais, fruto das condições econômicas enfrentadas pela região, abrindo novos restaurantes e fechando outros.

3.4 Tratamento de dados

Nessa fase, foi realizado um processamento nos dados coletados. O principal objetivo desse tratamento foi excluir dados de restaurantes incompletos, incorretos

e/ou duplicados, visto que tais dados poderiam representar inconsistência nas recomendações, portanto, não poderiam ser utilizados no sistema. Além disso, todos os dados de restaurantes coletados, não dispunham de informações sobre as especialidades dos restaurantes, ou seja, não havia definição se o restaurante selecionado era bar ou pizzaria ou ainda os dois ao mesmo tempo. Tal informação se mostrou extremamente relevante ao sistema, visto que cada usuário demonstra interesse por uma especialidade diferente, além disso, a especialidade, nesse sistema, é um dos filtros para recomendação.

Após finalizado o tratamento dos dados dos restaurantes, o próximo passo é correlacionar as especialidades e com os restaurantes. Essa correlação foi realizada através de busca de palavras-chave e inserção manual no banco de dados e, como não é restrito aos restaurantes a especificidade de um tipo de especialidade, ou seja, cada restaurante pode ter mais de uma especialidade atribuída (1:n), se fez necessário uma análise ainda mais criteriosa dos dados.

3.5 Definição de amostra

Como exposto no capítulo 2, um dos problemas dos sistemas de recomendação de filtragem colaborativa, é a falta de dados (cold start), que necessita de um número mínimo de interações dos usuários para poder recomendar. Por isso, devido à alta quantidade de restaurantes coletados e o tempo disponível para interação dos usuários, se fez necessário uma redução no número de restaurantes, de modo a não prejudicar os algoritmos de recomendação.

Dentre os principais tipos de amostragem, adotou-se a amostragem aleatória estratificada, visto que em sua definição, ela divide a população em um certo número de subpopulações e então extrai uma amostra de cada estrato, sendo assim os dados coletados dos restaurantes, poderiam ser estratificados de acordo com as especialidades de cada um (açaiteria, bar, pizzaria...). [8]

Definido o tipo de amostragem, foi desenvolvido um código para sortear números aleatórios dentro do universo de ids dos restaurantes cadastrados. Após realizada a redução, o número de restaurantes deixou de ser 1709 e passou a ser 171 restaurantes. Sempre levando em consideração a subpopulação de

especialidades, foi coletada uma amostra referente a 10% dos restaurantes de cada especialidade, não havendo assim, exclusão de nenhuma especialidade.

3.6 Formulário

A construção do formulário, se fez necessária com o intuito de obter dados de forma explícita do usuário, a fim de gerar insumos, para os algoritmos de recomendações. Com isso os usuários utilizavam essa ferramenta para interagir com o sistema como um todo, preenchendo dados de preferências, classificando os restaurantes e recebendo recomendações.

Nesta seção será exposto o passo a passo para o planejamento, construção e disponibilização do formulário aos usuários.

3.6.1 Desenvolvimento do Formulário

“Questionários bem feitos produzem informações valiosas, mas os pesquisadores costumeiramente enfrentam uma grande dificuldade: as pessoas hesitam – ou, até mesmo, resistem – em responder às muitas perguntas que lhes são feitas. Isso é compreensível porque responder a um questionário toma tempo, exige atenção e reflexão, e requer tomada de decisão diante de algumas questões.” [10]

Para dirimir os problemas expostos na citação acima, a construção do formulário de coleta de dados do usuário foi pensada, para ser efetuada de maneira simples, direta e intuitiva. Para isso acontecer, foi selecionado um design simples e objetivo, apenas com uma imagem de background padrão e a caixa de resposta em destaque a frente.

Outro fator considerado na construção do formulário, foi o tempo de resposta das perguntas escolhidas. Para que o tempo de resposta fosse mínimo, e isso refletisse em novas classificações e recomendações, foram escolhidas apenas 6 perguntas subjetivas que satisfizessem a pesquisa e não tomassem tempo do usuário, além de minimizar a taxa de respostas erradas na pesquisa.

Cada página de coleta do formulário de pesquisa, continha apenas, um botão de prosseguir, posto sempre do lado inferior direito, e acima uma contagem dos passos a serem seguidos para a finalização da pesquisa. Como a maior parte dos principais softwares, utilizados pela população hoje em dia, possui a mesma configuração, isso gera um sentimento de ambiente conhecido, tornando o manuseio do formulário agradável e de fácil entendimento.

Por estar disponível na internet, foi possível atingir um maior número de usuários, de diferentes lugares do mundo, foi observado inclusive 3 interações provenientes de Los Angeles, nos EUA e 2 outras da cidade de Braga, em Portugal. Outra vantagem da disponibilidade na internet, foi que o formulário ficou disponível 24 horas por dia durante os 18 dias, do dia 1 ao dia 18 de outubro de 2020 e ao fim deste período, todos os dados estavam disponíveis de forma organizada.

Além disso, visando diminuir o tempo de resposta do servidor, o formulário possui 6 interações iniciais com o usuário (**e-mail, bairro que mora/está hospedado, bairro que costuma comer, escolha de especialidades, classificação e recomendações**), que serão detalhadas no tópico a seguir, porém realiza apenas 4 interações com o servidor Web (**consulta de e-mail, cadastro de preferências de usuário, cadastro de restaurantes classificados, cadastro de melhor recomendação**). Essa redução na comunicação com o servidor, reduz o tempo de preenchimento dos dados do formulário, refletindo num aumento no número de interações posteriores dos usuários.

E caso o servidor já possua os dados do usuário (usuário já cadastrado), o número de campos a serem preenchidos diminui ainda mais e passam a ser apenas 3 (**e-mail, classificação de restaurantes e recomendações**), por consequência, quanto menos campos, mais classificações.

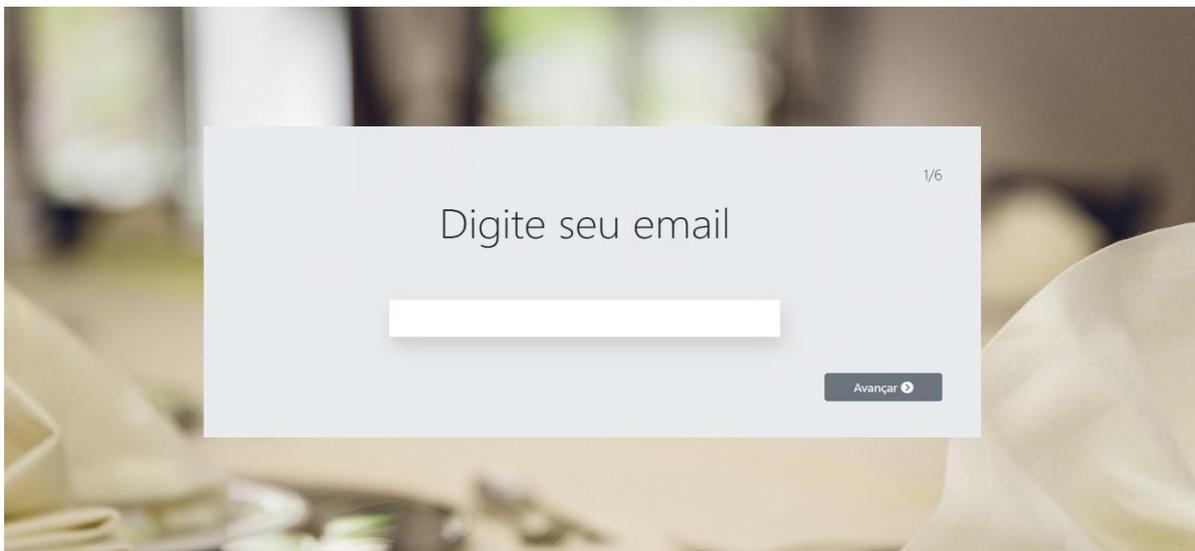
Perguntas diretas, tempo de resposta baixo, interação intuitiva e design simples, foram os principais atributos do formulário que ao fim do período de pesquisa, conseguiu atingir seu objetivo final, coletar um número mínimo de respostas, de modo a permitir que as recomendações fossem geradas.

3.6.2 Campos do Formulário

E-mail – Primeiro campo do formulário, esse campo é responsável pelo armazenamento dos dados do usuário, o e-mail do usuário vai identificá-lo numa

possível volta ao formulário, permitindo mais classificações de restaurantes e aumentando a precisão nas recomendações, baseadas no usuário.

Figura 4 - formulário campo e-mail

A imagem mostra uma interface de usuário com um formulário de campo de e-mail. O formulário é um retângulo cinza claro sobreposto a uma imagem de fundo desfocada de pratos de comida. No topo direito do formulário, há o texto "1/6". No centro, o texto "Digite seu email" está acima de um campo de entrada de texto branco. No canto inferior direito do formulário, há um botão cinza com o texto "Avançar" e uma seta para a direita.

Bairro que mora – Segunda opção de filtro de restaurantes, pois se o usuário já classificou todos os restaurantes do bairro em que ele costuma comer, esse bairro é utilizado para buscar restaurantes mais próximos do usuário, das especialidades que ele escolheu como preferidas.

Figura 5 - formulário campo bairro mora

A imagem mostra uma interface de usuário com um formulário de campo "bairro mora". O formulário é um retângulo cinza claro sobreposto a uma imagem de fundo desfocada de pratos de comida. No topo direito do formulário, há o texto "2/6". No centro, o texto "Qual bairro você mora/está hospedado?" está acima de um campo de entrada de texto branco. No canto inferior direito do formulário, há um botão cinza com o texto "Avançar" e uma seta para a direita.

Bairro que come – Primeira opção de busca de restaurantes, essa variável será crucial para mapear no banco de dados quais os restaurantes que serão relevantes para este usuário.

Figura 6 - formulário campo bairro costuma comer

A imagem mostra um formulário de pesquisa de restaurantes. O formulário é uma caixa branca com um fundo cinza claro, sobreposta a uma imagem de fundo desfocada de pratos e copos. No canto superior direito do formulário, há o texto "3/6". O texto principal do formulário pergunta: "Em que bairro do Recife você costuma comer?". Abaixo da pergunta, há um campo de entrada de texto branco. No canto inferior direito do formulário, há um botão cinza com o texto "Avançar" e um ícone de seta para a direita.

Escolha de especialidades – Especialidades são as categorias gastronômicas, as quais os restaurantes se classificam, como por exemplo Pizzaria, Comida regional, entre outros. Nesse campo o usuário seleciona as especialidades, com as quais o usuário tem mais apreço e usualmente frequenta restaurantes desse tipo. Existem 18 especialidades a serem escolhidas, são elas: açaiteria, bar, boleria, café, churrascaria, comida árabe, comida chinesa, comida italiana, comida regional, comida japonesa, doceria, frutos do mar, galeteria, hamburgueria, lanchonete, pizzaria, sorveteria e comida saudável. Algumas vezes encontra-se restaurantes com mais de uma especialidade definida, como por exemplo um bar que também vende pizza, dessa forma ele será classificado nas especialidades de bar e pizzaria.

Figura 7 - formulário campo seleção especialidade

Classificação – Passado a fase da escolha de especialidades, será exibido ao usuário uma lista de restaurantes, baseados no bairro em que relatou costumeiramente comer e em sua escolha de especialidades.

Nessa lista de restaurantes o usuário irá classificar, de 0 a 5 estrelas, pelo menos um restaurante conhecido, e após isso será direcionado para a próxima página (exibição de recomendações).

Figura 8 - formulário classificação dos restaurantes

Exibição das recomendações – Para finalizar o preenchimento do formulário, nessa última página serão exibidas três recomendações de restaurantes, cada uma coletada através de um algoritmo de recomendação diferente, sendo:

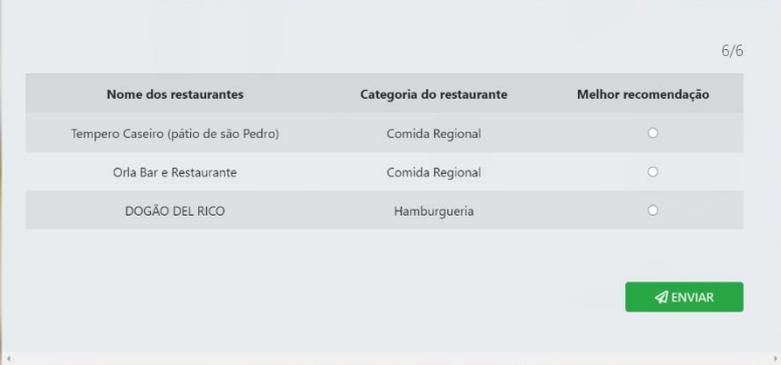
A primeira recomendação será referente ao algoritmo baseado no conteúdo, onde serão coletados, daquelas especialidades escolhidas pelo usuário a com melhor nota, levando-se em consideração a classificação coletada via API do Google Places.

Já na segunda recomendação o algoritmo é baseado no usuário. Aqui o servidor irá procurar perfis de usuário parecidos com o usuário em questão e retornar qual restaurante mais indicado.

A terceira recomendação foi construída de forma híbrida, onde há uma comutação em paralelo dos algoritmos gerando recomendações de diferentes formas.

Após o preenchimento desses dados, o usuário é convidado a retornar para classificar mais restaurantes onde irá popular ainda mais o banco de dados e receber outras recomendações. O intuito desse convite é para que o sistema de recomendação que se baseia no usuário possua mais dados do perfil daquele usuário e possa encontrar com mais facilidade perfis parecidos, aumentando a precisão nas recomendações.

Figura 9 - formulário recomendações



Nome dos restaurantes	Categoria do restaurante	Melhor recomendação
Tempero Caseiro (pátio de são Pedro)	Comida Regional	<input type="radio"/>
Orla Bar e Restaurante	Comida Regional	<input type="radio"/>
DOGÃO DEL RICO	Hamburgueria	<input type="radio"/>

6/6

ENVIAR

4 RECOMENDAÇÃO BASEADA EM CONTEÚDO

Os sistemas de recomendação, foram evoluindo ao passar dos anos e hoje em dia, possuem várias técnicas de filtragem dos dados, para selecionar a recomendação a ser dada. A técnica de filtragem baseada em conteúdo, que será exposta nesse capítulo, foi uma das utilizadas nesse sistema, pois evita o problema de cold start, observado na filtragem colaborativa e será a primeira a ser exposta nesse trabalho.

4.1 Desenvolvimento

Na técnica de filtragem colaborativa, utiliza-se a matriz item x usuário para realizar as predições, porém na técnica de filtragem baseada em conteúdo utiliza-se uma matriz diferente, matriz item x atributos do item [21]. E para realizar tal predição escolheu-se três tipos de atributos diferentes, a serem utilizados: a especialidade dos restaurantes, a localização dos restaurantes, por bairro e, por fim a classificação (rating) dos restaurantes coletada na API do Google Places, que nada mais são do que uma média de notas para cada restaurante, dadas pelos usuários dessa API, que variam de 0 a 5 estrelas.

Após a definição dos atributos dos restaurantes, que seriam utilizados na técnica de filtragem baseada em conteúdo, o próximo passo é coletar tais dados dos usuários para posterior recomendação. Para isso, em um primeiro momento do formulário foi pedido para que o usuário selecionasse o bairro que mora e o bairro em que costuma comer, além das especialidades de restaurantes que teria interesse e posteriormente uma classificação de restaurantes baseada nas especialidades escolhidas.

Após isso, constituiu-se como dados dos usuários: uma lista de restaurantes de cada especialidade que o usuário gosta, além do bairro que o usuário mora e o bairro que ele costuma comer. E de dados dos itens têm-se a classificação dos restaurantes pela API do Google Places, o bairro que o restaurante está localizado e a lista de especialidade de cada restaurante.

De posse desses dados, verifica-se qual a lista de especialidades que o usuário selecionou como preferidas, para que a recomendação do restaurante seja baseada na especialidade escolhida pelo usuário.

Posterior a isso, realiza-se uma busca em banco de dados para coleta de todos os restaurantes, que possuem em sua lista de especialidade semelhantes àquela que o usuário colocou como preferida.

Finalmente, reduz-se essa lista com um filtro daqueles restaurantes, que estejam no bairro que o usuário selecionou onde mora ou costuma comer. E por fim, dessa lista final é selecionado aquele restaurante que possui melhor classificação do Google Places API e ainda não foi recomendado.

Após passar por todos esses passos, o restaurante selecionado desses filtros é exibido na primeira posição da lista com os três restaurantes recomendados.

5 RECOMENDAÇÃO COLABORATIVA BASEADA NO USUÁRIO

Como o filósofo Aristóteles dizia o homem é um ser social, pois é um animal que precisa de outros membros da espécie [24]. A opinião de uma pessoa pode influenciar de várias formas a tomada de decisão de outras. Neste capítulo abordar-se-á o desenvolvimento do sistema de recomendação de restaurantes, utilizando-se a técnica de filtragem colaborativa, baseada no usuário.

5.1 Desenvolvimento

A técnica da filtragem colaborativa, baseada no usuário, utiliza a opinião dos usuários considerados similares ao usuário alvo e recomenda itens que foram considerados relevantes àquele usuário alvo [22].

A tabela 1 representa as notas dadas pelos usuários por cada restaurante listado. Pode-se observar que o usuário 1 e o usuário 3 são similares, pois deram as mesmas notas para os restaurantes 2 e 4. Levando em conta que os dois usuários são similares, podemos prever que a nota que será dada ao restaurante 1 pelo usuário 1 é 4, visto que os gostos do usuário 1 e usuário 3 são bem semelhantes. Logo, podemos recomendar o restaurante 1 ao usuário 1. Seguindo essa lógica o primeiro passo para realizar uma recomendação usando a técnica de filtragem colaborativa, é definir o usuário mais semelhante, após isso realizar as previsões e por fim as recomendações.

Tabela 1 - tabela usuário x restaurante

	restaurante 1	restaurante 2	restaurante 3	restaurante 4	rest... n
user 1		2	3	5	
user 2	5		1		
user 3	4	2		5	
user 4		4	1		
user n					

Para selecionar quais usuários do sistema são mais similares ao usuário alvo, existem alguns cálculos matemáticos conhecidos de correlação. Porém, neste trabalho, utilizou-se o cálculo de correlação de Pearson. Para isso cria-se um modelo no formato .csv dos dados onde a primeira coluna representa o usuário, a segunda coluna representa o restaurante e a última coluna representa as notas atribuídas pelos usuários.

Usuários semelhantes definidos, o próximo passo é definir qual a vizinhança, ou seja, quais desses usuários são mais similares ao usuário alvo. Nesse trabalho foi considerado similar aquele usuário que possui uma diferença de similaridade de até 0,1.

Por fim, o sistema de recomendação irá gerar a recomendação do restaurante para aquele usuário, baseado nos gostos dos usuários mais semelhantes e próximos a ele. Tal recomendação preenche a segunda linha dos três restaurantes recomendados.

Caso o recomendador não consiga recomendar, devido ao problema do cold start, o sistema retornará um restaurante escolhido através do Top N, os n restaurantes mais bem avaliados pela API do Google Places, e tal valor não será contabilizado como uma recomendação baseada no usuário.

6 RECOMENDAÇÃO HÍBRIDA

A recomendação híbrida, procura utilizar algumas formas de filtragem diferentes, mas que se completam. Para que possamos hibridizar um sistema de recomendação, primeiro de tudo, precisamos definir quais as técnicas de filtragem que serão escolhidas e como essas técnicas serão combinadas. [22]

Nesse capítulo, será exposto toda a trajetória a ser percorrida para que a recomendação híbrida pudesse acontecer e quais as técnicas de filtragem utilizadas, atributos escolhidos e como se utilizou tais atributos para a recomendação.

6.1 Desenvolvimento

A filtragem colaborativa, possui o problema do cold start, que não consegue recomendar restaurantes devido à falta de dados. Enquanto a filtragem baseada em conteúdo possui o problema de superespecialização, diminuindo a eficácia das recomendações no longo prazo. Visando utilizar o melhor dos mundos, a forma escolhida para a hibridização do sistema de recomendação é utilizar-se da filtragem colaborativa baseada no usuário, filtragem colaborativa baseada nos restaurantes e por fim filtragem baseada em conteúdo.

Primeiro passo para que isso aconteça é utilizar o sistema de recomendação baseado no usuário para receber uma recomendação. Procura-se usuários semelhantes, define-se a vizinhança e gera-se a recomendação do restaurante.

Com acesso a essa recomendação baseada no usuário, o próximo passo é desenvolver um sistema de recomendação baseado no restaurante, e encontrar um restaurante com características similares ao restaurante recomendado pela filtragem baseada no usuário.

Sabe-se, porém, que a filtragem colaborativa enfrenta problemas com a falta de dados, cold start, e que por esse motivo em alguns momentos, não será possível essa recomendação, e para isso, utiliza-se a recomendação baseada em conteúdo, coletando-se uma lista de restaurantes baseado pela classificação da API do Google Places e retornando o restaurante de maior classificação, de acordo com as preferências do usuário.

Tal recomendação ocupará a terceira linha da lista de restaurantes recomendados na tela final do formulário.

7 ANÁLISE DE RESULTADOS

Neste trabalho foi realizado uma amostragem aleatória estratificada para que o número de restaurantes fosse condizente com a população utilizada na pesquisa de campo. E com isso o número de restaurantes final foi de 171, divididos em 18 especialidades.

O formulário de pesquisa de coleta de dados foi disponibilizado para os usuários no período de 18 dias, iniciando no dia 1 de outubro de 2020 e finalizando no dia 18 de outubro de 2020. Por estar hospedado na web, foi possível o compartilhamento do endereço do formulário, através de redes sociais, canais de comunicação pessoal e indicações de amigos. Ao fim do período de divulgação, observou-se 134 classificações de restaurantes pelos 53 usuários cadastrados. Tais números foram suficientes para superar os problemas iniciais da filtragem colaborativa e recomendar restaurantes.

Como resultado de melhores recomendações tivemos:

Tabela 2 - resultados parciais

Técnica Seleccionada	Melhor Recomendação
Top N	31 vezes
Baseado em conteúdo	43 vezes
Baseado no usuário	33 vezes
Híbrido	27 vezes

A recomendação “Top N” está sendo selecionada como melhor resultado, quando as recomendações baseadas em usuário e híbrida não conseguem recomendar, devido à falta de dados (cold start), por isso ela será desconsiderada.

Descartando-se as classificações da técnica Top N temos:

Tabela 3 - resultados finais

Técnica Seleccionada	Melhor Recomendação	Porcentagem
Baseada em Conteúdo	43 vezes	42%
Baseada no Usuário	33 vezes	32%

Híbrido	27 vezes	26%
---------	----------	-----

Levando-se em consideração tal amostra de dados, observa-se que apesar de ser suficiente para gerar uma recomendação, a pesquisa necessitaria de um maior tempo em exposição, com uma quantidade maior de dados coletados dos usuários. Uma vez que a amostra que a recomendação baseada em conteúdo, foi a mais votada como melhor recomendação, se comparada as demais. Em segundo lugar a filtragem colaborativa baseada em usuário teve um desempenho um pouco melhor que a híbrida, que possuía duas técnicas de filtragem e necessita de uma quantidade de dados maior.

Levando-se em consideração os dados obtidos, expostos na tabela 3, conclui-se que a criação do sistema de recomendação trouxe simplicidade e praticidade à busca de restaurantes pelo usuário, visto que o usuário recebe recomendações de restaurantes, as vezes desconhecidos, dos locais onde mora ou já possui a cultura de realizar suas refeições e sempre levando em consideração as suas preferências pessoais. Porém tais recomendações podem se tornar mais precisas com um maior tempo de uso do sistema e maior aprendizado sobre preferências pessoais.

Além disso, pelo sistema também considerar os bairros do Recife como atributos em suas recomendações, estabelecimentos, antes desconhecidos, aparecem como opções de recomendações e passam a se tornar mais presentes para os usuários do sistema. Isso implica em um aumento na divulgação de sua marca, reduzindo a probabilidade de seu negócio falir prematuramente, além de incentivar o turismo na cidade de forma capilar e não somente em pontos turísticos já bastante conhecidos.

8 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo geral analisar o desenvolvimento de um sistema de recomendação de restaurantes, baseado em três tipos diferentes de técnicas de filtragem de dados, nos bairros da cidade do Recife e teve seu objetivo geral atendido, visto que o sistema de recomendação de restaurantes foi desenvolvido e colocado em produção em ambiente controlado, coletando dados relevantes ao projeto.

Além disso, o trabalho explicou os fundamentos das técnicas de filtragem que foram utilizadas no desenvolvimento do sistema de recomendação de restaurantes. Identificou como definir, coletar, tratar e analisar os dados necessários à construção do sistema de recomendação. Realizou uma análise passo a passo do desenvolvimento de um sistema de recomendação baseado no conteúdo. Analisou o desenvolvimento de um sistema de recomendação, baseado no usuário. Analisou o desenvolvimento de um sistema de recomendação híbrido. E por fim, identificou que a técnica baseada em conteúdo, foi a que mais se adequou à amostra e tempo do trabalho, e trouxe melhores resultados.

8.1 Limitações

Este trabalho foi posto em produção, obtendo interações de usuários reais, porém a amostra de usuários e a amostra de restaurantes a recomendar, foi pequena, considerando-se a quantidade de estabelecimentos reais, que se encaixam no perfil de alimentação e poderiam ser recomendados. Além disso, os dados coletados não se mostraram suficientes para abordar outros aspectos presentes dos restaurantes, como por exemplo acessibilidade, poder aquisitivo do cliente, cardápio do restaurante, entre outros que poderiam enriquecer a qualidade da recomendação.

Além disso, o período o qual foi realizado a pesquisa de campo e a disponibilização do formulário para preenchimento foi antes da pandemia da COVID-19, por isso, é provável que alguns dos restaurantes encontrados no banco de dados, não estejam mais em funcionamento.

8.2 Trabalhos Futuros

Esse trabalho utilizou como base de dados para recomendações, os restaurantes da cidade do Recife, categorizados por bairros. Uma sugestão de trabalho futuro, que complementaria esse trabalho, seria a expansão ou modificação dessa base de dados, utilizando dados de outras regiões do país.

Além disso poderia utilizar-se de dados de locais fechados, como Centro de Compras (Shoppings), onde a densidade do número de restaurantes é maior e as recomendações que se utilizam de sistemas de localização como o GPS não são

precisos. Isso poderia influenciar positivamente no número de clientes de alguns restaurantes menos conhecidos.

Outro fator que pode ser levado em consideração é a implementação de novos contextos, utilizando-se outros tipos de dados como foco, inclusive através de coleta de dados implícitas, a serem considerados pelos algoritmos de recomendação, como o dia da semana, o horário, o tempo e até o perfil financeiro do usuário, sempre com o intuito de melhorar a experiência do usuário no recebimento das recomendações.

APÊNDICE

INFRAESTRUTURA

Neste bloco será definido quais as infraestruturas utilizadas no desenvolvimento do sistema de recomendação.

Hospedagem do sistema de recomendação

A computação em nuvem é um modelo que permite o acesso onipresente e conveniente à rede sob demanda a um pool compartilhado de recursos computacionais configuráveis [5]. Além de indiretamente essa definição, aludir à atributos como a disponibilidade e escalabilidade, o termo *compartilhamento de recursos*, remete também a definição de economia criativa, prática que vem sendo utilizada em diversos países e que possibilita o acesso a bens e serviços sem a necessidade de aquisição do produto [9].

A opção pela computação em nuvem se deu pois, de outra forma, seria necessário a aquisição de um servidor próprio, para a demanda desse trabalho. Isso, além de ser custoso financeiramente, também acarretaria custos de tempo e aprendizado, por isso, optou-se pela utilização de um sistema já configurado e validado pelo mercado.

A Amazon Web Services (AWS) é hoje em dia, a plataforma de nuvem mais adotada e mais abrangente do mundo [3], além disso possui liberdade para escolha do sistema operacional do servidor atuante, com isso, não foi necessário disponibilizar tempo demasiado, em uma curva de aprendizado alta, para se adaptar ao sistema operacional desconhecido.

Ademais, outro e o principal motivo para a escolha da Amazon, foi por permitir a leitura e escrita de um documento de texto dentro do servidor, no meu caso um .CSV como mencionado nesse capítulo na secção de recomendações. Tal atributo não fazia parte do pool de funcionalidades, de outras plataformas de nuvem pesquisadas, como por exemplo o Google App Engine, Google Compute Engine e Heroku.

Interações com o servidor

A linguagem utilizada para o desenvolvimento do sistema de recomendações foi Java e, para simplificar as configurações do servidor e a comunicação com o formulário de pesquisa, foi utilizado o framework Spring, mais precisamente o Spring Boot. Seu foco em simplicidade e produtividade, trazendo consigo um pool de bibliotecas pré-configuradas, otimizou bastante tempo no desenvolvimento do código do servidor [13].

Para que a troca de dados entre servidor e formulário, ocorresse com o mínimo de interações possíveis, a fim de minimizar o tempo de espera do usuário e agilizar a conclusão do formulário, essa comunicação acontecia através de 4 classes com requisições do tipo POST, que serão detalhadas a seguir.

Na primeira interação, o formulário enviava o e-mail do usuário que era conferido no banco de dados se estava cadastrado ou não. Caso o e-mail do usuário já estivesse cadastrado, o servidor retornaria uma lista de restaurantes para o usuário classificar, dar notas, com base em seu bairro, suas preferências de especialidades e nos restaurantes já classificados. Caso contrário, o servidor retorna uma lista vazia e o formulário prossegue com perguntas de cadastro de usuário (**bairro que mora, bairro que costuma comer, preferências de especialidades**), que serão armazenadas em banco pelo servidor, através da classe Primeiro.

Com o usuário já cadastrado, a segunda interação do formulário, coleta dados dos usuários através das classificações de restaurantes, enquanto o servidor armazena essas respostas, a partir da classe Segundo, e retorna as recomendações para apreciação do usuário.

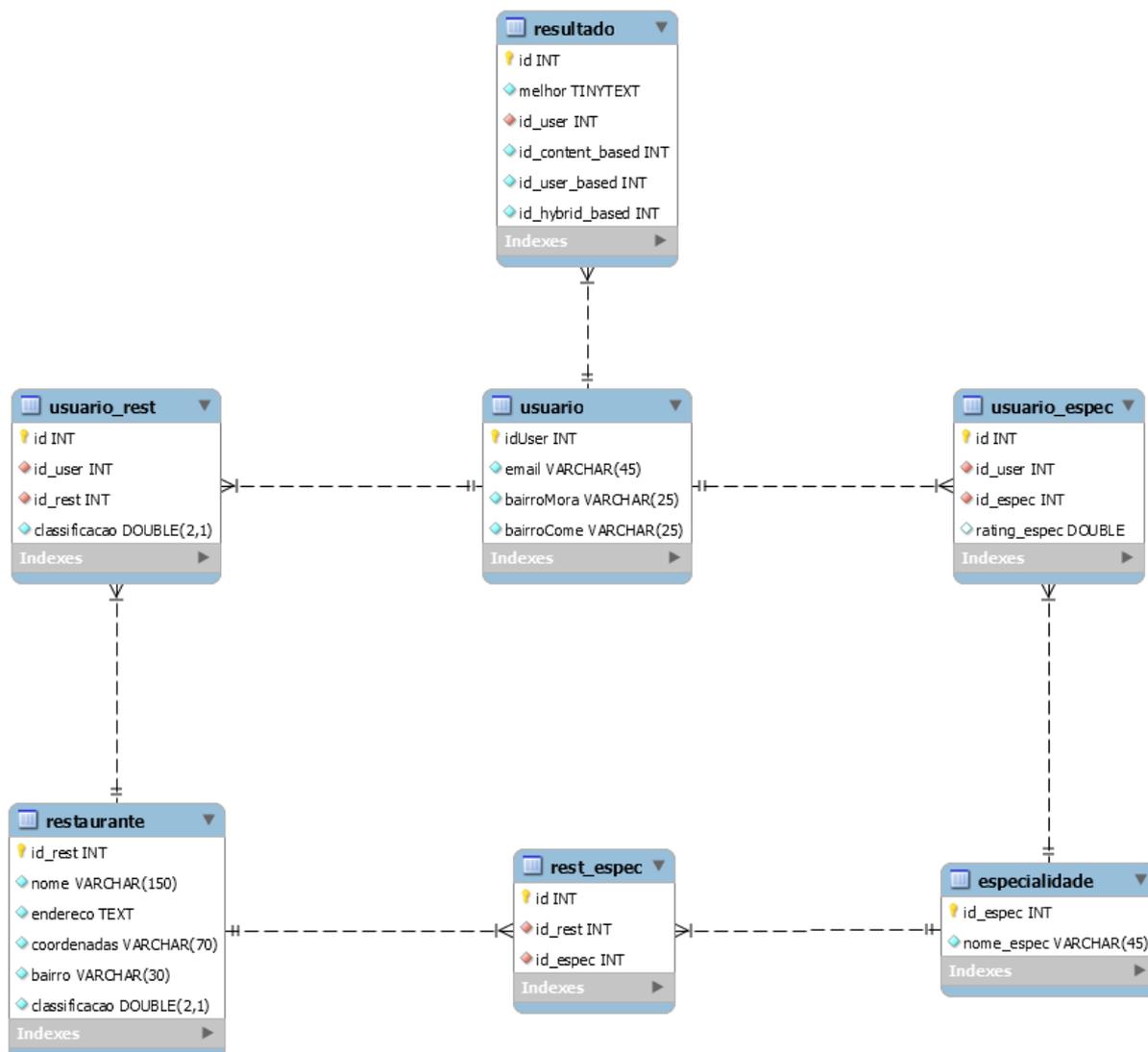
E por fim, na terceira interação, do usuário já cadastrado em banco, ele recebe 3 restaurantes e escolhe qual algoritmo de recomendação obteve maior grau de sucesso em sua recomendação, ou seja, qual desses métodos, mais se aproximou das preferências do usuário. Após enviar, ao servidor, a escolha do usuário, a classe Terceiro armazena os dados do resultado no banco e o processo de coleta de dados se reinicia.

Banco de dados

Na modelagem do banco, foi percebido 3 entidades principais: os restaurantes, as especialidades e os usuários. O relacionamento entre eles é de cardinalidade n:n, logo, se fez necessário a criação de mais 3 tabelas intermediárias. E por fim, criou-se a tabela de resultado, onde serão armazenados os ids dos restaurantes recomendados e qual método foi votado pelo usuário como melhor recomendação.

Nas tabelas *usuario_rest* e *usuário_espec*, observa-se a criação de duas variáveis adicionais, uma armazena os dados referentes a classificação do usuário aos restaurantes, atribuindo-lhes notas de 0 a 5 estrelas. Já do outro lado da imagem na tabela *usuário_espec*, tem-se a variável *rating_espec*, que armazena os valores de probabilidade de classificação do usuário a cada especialidade de restaurante, ou seja, a probabilidade do usuário classificar um bar, uma pizzaria, levando-se em consideração os restaurantes já classificados. Como observado na figura 10, a tabela resultado só se relaciona com a tabela usuário.

Figura 10 - modelagem banco



Apache Mahout

Em 2005, foi criado pela Apache Software Foundation, o Apache Hadoop, com o intuito de processar e armazenar dados em grande escala [16]. Dentro dele, existe um framework chamado MapReduce, que nos permite trabalhar com grandes quantidades de dados de forma paralela, maximizando o desempenho no trato dessa gigantesca enxurrada de dados [16].

Utilizando esse conceito do MapReduce, em 2009, foi criado o Apache Mahout, cujo objetivo é produzir de forma gratuita, implementações de algoritmos de aprendizado de máquina distribuídos ou escalonáveis, e aplicados no campo de análise de dados em grande escala [14][16]. Além disso, Mahout também fornece

bibliotecas para as linguagens Java/Scala [14], e se consolidou sendo usado no desenvolvimento de recomendações baseadas em filtragem colaborativa. [15]

Framework gratuito, consolidado no mercado, trabalha com filtragem colaborativa e possui biblioteca para a linguagem Java. Tais motivos, foram determinantes para a escolha do Mahout no desenvolvimento do sistema de recomendações abordado nesse trabalho.

REFERÊNCIAS

- <https://docs.aws.amazon.com/AmazonS3/latest/gsg/GetStartedWithS3.html> [1]
- <https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html> [2]
- <https://aws.amazon.com/pt/what-is-aws/> [3]
- <https://redmonk.com/sograzy/2020/07/27/language-rankings-6-20/> [4]
- Ruparelia, Nayan B.. Cloud Computing. MIT Press, 6 de maio de 2016. [5]
- <https://developers.google.com/places/web-service/overview> [6]
- <https://developers.google.com/places/web-service/search> [7]
- DONALD COOPER; Pamela S. Schindler. *Métodos de Pesquisa Em Administração*. Bookman; 2003. ISBN 978-85-363-0117-4. p. 164. [8]
- https://pt.wikipedia.org/wiki/Consumo_colaborativo [9]
- Vieira, Sonia. Como Elaborar Questionários. Atlas; 2009 [10]
- http://dados.recife.pe.gov.br/dataset/area-urbana/resource/698d8fe1-d30a-485f-8d5d-307d46163d0c?inner_span=True [11]
- https://developers.google.com/places/supported_types [12]
- <https://spring.io/why-spring> [13]
- "Introducing Apache Mahout". ibm.com. 2011. Retrieved 13 September 2011. [14]
- https://www.researchgate.net/publication/235899480_Collaborative_Filtering_with_Apache_Mahout [15]
- <https://www.devmedia.com.br/descobrimdo-o-apache-mahout/30435> [16]
- GOLDBERG, D. et al. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, v. 35, n. 12, pp. 61-70, 1992. [17]
- Walter Isaacson, Denise Bottmann, Pedro Maia Soares, Berilo Vargas. *Steve Jobs*. Companhia das Letras. [18]
- <https://mobilesyrup.com/2017/08/22/80-percent-netflix-shows-discovered-recommendation/> [19]
- F.O.IsinkayeaY.O.FolajimibB.A.Ojokohc, Recommendation systems: Principles, methods and evaluation, *Egyptian Informatics Journal* Volume 16, Issue 3, November 2015, Pages 261-273. Disponível em <https://www.sciencedirect.com/science/article/pii/S1110866515000341>. Acessado em 07/12/2021 às 22:00 [20]
- Dietmar Jannach, Markus Zander, Alexander Felfernig, Gerhard Friedrich. *Recommender Systems: An Introduction*. Cambridge University press, 2010 [21]
- Francesco Ricci, Lior Rokach, Bracha Saphira. *Recommender Systems Handbook*. Springer, 2015 [22]
- JANNACH, D. et al. Recommender systems — beyond matrix completion. *Commun. ACM*, ACM, New York, NY, USA, v. 59, n. 11, p. 94–102, out. 2016. [23]
- [https://www.psicanaliseclinica.com/o-homem-e-um-ser-social/#:~:text=e%20Hobbes.-,Arist%C3%B3teles%3A%20o%20homem%20%C3%A9%20um%20ser%20social%20porque%20%C3%A9%20um,dos%20outros%20membros%20da%20esp%C3%A9cie&text=N%C3%A3o%20%C3%A9%20%C3%A0%20toa%20que,porque%20%C3%A9%20um%20ser%20carente](https://www.psicanaliseclinica.com/o-homem-e-um-ser-social/#:~:text=e%20Hobbes.-,Arist%C3%B3teles%3A%20o%20homem%20%C3%A9%20um%20ser%20social%20porque%20%C3%A9%20um,dos%20outros%20membros%20da%20esp%C3%A9cie&text=N%C3%A3o%20%C3%A9%20%C3%A0%20toa%20que,porque%20%C3%A9%20um%20ser%20carente.). Acessado em 07/12/2021 às 22:14 [24]
- <https://mundoeducacao.uol.com.br/historiageral/igreja-na-idade-media.htm> [25]
- <https://www.tecmundo.com.br/curiosidade/8567-pesquisadores-estimam-a-quantidade-de-informacao-existente-no-mundo.htm> [26]
- <http://www2.recife.pe.gov.br/servico/cidade-1> [27]

<https://www.terra.com.br/vida-e-estilo/turismo/brasil/nordeste/terceiro-polo-gastronomico-do-pais-recife-agrada-a-todos,73088908cb237310VgnCLD100000bbcceb0aRCRD.html> [28]

<https://viajando.expedia.com.br/recife-capital-gastronomica-do-nordeste-855/> [29]

<https://agenciabrasil.ebc.com.br/economia/noticia/2021-06/sebrae-pequenos-negocios-tem-maior-taxa-de-mortalidade> [30]

<https://aeroin.net/pernambuco-comemora-movimento-aereo-175-maior-que-em-2020-em-recife/> [31]