

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

VINÍCIUS GABRIEL ANGELOZZI VERONA DE RESENDE  
Orientador: Prof. Dr. Guilherme Tavares de Assis

**PROPOSTA E CRIAÇÃO DE UM SERVIÇO DE RECOMENDAÇÃO DE  
EVENTOS EM REDES SOCIAIS**

Ouro Preto, MG  
2024

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

VINÍCIUS GABRIEL ANGELOZZI VERONA DE RESENDE

**PROPOSTA E CRIAÇÃO DE UM SERVIÇO DE RECOMENDAÇÃO DE EVENTOS EM  
REDES SOCIAIS**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

**Orientador:** Prof. Dr. Guilherme Tavares de Assis

Ouro Preto, MG  
2024

## SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

R433p Resende, Vinicius Gabriel Angelozzi Verona de.  
Proposta e criação de um serviço de recomendação de eventos em  
redes sociais. [manuscrito] / Vinicius Gabriel Angelozzi Verona de  
Resende. - 2024.  
47 f.: il.: color., gráf., tab..

Orientador: Prof. Dr. Guilherme Tavares Assis.  
Monografia (Bacharelado). Universidade Federal de Ouro Preto.  
Instituto de Ciências Exatas e Biológicas. Graduação em Ciência da  
Computação .

1. Sistema de recomendação. 2. Rede neural. 3. Rede social. 4. Rede  
social baseada em evento. I. Assis, Guilherme Tavares. II. Universidade  
Federal de Ouro Preto. III. Título.

CDU 004

Bibliotecário(a) Responsável: Luciana De Oliveira - SIAPE: 1.937.800



## FOLHA DE APROVAÇÃO

**Vinicius Gabriel Angelozzi Verona de Resende**

### **Proposta e criação de um serviço de recomendação de eventos em redes sociais**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 09 de Agosto de 2024.

#### Membros da banca

Guilherme Tavares de Assis (Orientador) - Doutor - Universidade Federal de Ouro Preto  
Reinaldo Silva Fortes (Examinador) - Doutor - Universidade Federal de Ouro Preto  
Anderson Almeida Ferreira (Examinador) - Doutor - Universidade Federal de Ouro Preto

Guilherme Tavares de Assis, Orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 09/08/2024.



Documento assinado eletronicamente por **Guilherme Tavares de Assis, PROFESSOR DE MAGISTERIO SUPERIOR**, em 09/08/2024, às 15:43, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0754197** e o código CRC **D6038527**.

*Dedico este trabalho aos meus pais, Hélio Fernando Verona de Resende e Selma Augusta Verona de Resende, cujo apoio e encorajamento foram fundamentais para a realização desta monografia. Ao meu irmão, Felipe Henrique Angelozzi Verona de Resende, pela ideia tida como base para a elaboração deste trabalho.*

*Com muito amor, dedico este trabalho também ao meu sobrinho, Felipe Ciola Verona. Que você cresça em um mundo de muitas oportunidades e realizações, e que no futuro possa se orgulhar e se inspirar em minhas tentativas de alcançar novos horizontes, independentemente dos desafios.*

*A todos vocês, dedico este trabalho, com muito amor. Obrigado por estarem ao meu lado e me proporcionarem esta oportunidade.*

# Agradecimentos

Gostaria de agradecer a todos as pessoas ou instituições que contribuíram para a realização do trabalho.

Ao meu orientador, Prof. Dr. Guilherme Tavares de Assis, pela ajuda, paciência e incentivo ao longo da realização da monografia. Seu conhecimento e dedicação foram essenciais para o desenvolvimento do trabalho.

Agradeço ao corpo docente do Departamento de Ciência da Computação da Universidade Federal de Ouro Preto, pelos ensinamentos e orientações durante o meu curso e pela permissão de defesa antecipada do trabalho.

Agradeço ao Prof. Dr. Hongzhi Yin, pela disponibilização de dados necessários para realização deste trabalho.

Aos meus amigos, em especial Carlos Eduardo Romaniello, Carlos Freitas, Daniel Jardim, Daniel Moneteiro, Filipe Ramos, Fábio Henrique Fernandes, Gabriel Mace, Gabriel Niquini, Igor Santiago, João Paulo Prata, Lucas Araújo, Marcus Vinicius Souza e Vinícius Cássio, por estarem ao meu lado e me apoiarem durante toda a jornada da minha graduação.

Agradeço também ao Departamento de Ciência da Computação e Departamento de Física, em especial ao Prof. Dr. Alan Barros de Oliveira e ao Laboratório de Dinâmica Molecular e Estrutura Eletrônica (NANO), pela disponibilização da rede de computadores utilizados durante o trabalho.

Por fim, agradeço a todos aqueles que contribuíram para minha jornada e para a realização desta monografia. Sem vocês, este trabalho não teria sido possível.

“It will not be we who reach Alpha Centauri and the other nearby stars. It will be a species very like us, but with more of our strengths and fewer of our weaknesses, a species returned to circumstances more like those for which it was originally evolved, more confident, farseeing, capable, and prudent.”(Sagan, 1994)

# Resumo

Com a popularização do acesso à *Internet*, surgiram diversas plataformas cujo foco está na socialização dos seus membros. Estas plataformas são comumente chamadas de redes sociais. Dentre as redes sociais existentes, há um grupo denominado “Redes Sociais Baseadas em Eventos” (EBSN), cujo objetivo está no auxílio da organização e distribuição de eventos *online* e *offline*. Com a presença de milhões de usuários e milhares de eventos semanais nas EBSN, existe a necessidade da criação de serviços que realizam recomendação de eventos de forma personalizada a cada um dos usuários, visto que a busca manual pode ser inviável e desmotivadora. Sendo assim, este trabalho propõe o desenvolvimento de um serviço de recomendação de eventos que considera *features* amplamente usados na literatura, como a localização do usuário e evento, assim como *features* não tão comuns, sendo estas a emoção, o sentimento e a personalidade do usuário. Para tal, de uma forma geral, este trabalho produziu: uma nova base de dados sintética para a classe de problemas e suporte do serviço de recomendação, a partir de uma base coletada; um *script* para coleta de dados públicos em plataformas reais e, por fim, o serviço de recomendação de eventos, desde a extração de *features* à coleta de *feedbacks* explícitos. Para averiguar a qualidade dos modelos utilizados na arquitetura de funcionamento do recomendador proposto, alguns experimentos comparativos foram realizados utilizando a base de dados mencionada. A partir dos experimentos realizados, foi possível observar que uma das técnicas de aprendizado propostas inferiu modelos, para o serviço de recomendação, que apresentaram resultados satisfatórios utilizando a base criada; no entanto, para outros, mostrou-se necessárias melhorias. Este trabalho contribui com a possível abertura de novos temas de pesquisa, além do desenvolvimento de um serviço de recomendação e uma nova base de dados sintética. Ademais, quanto ao serviço proposto de recomendação propriamente dito, conclui-se que ele utiliza de um módulo com boa engenharia de recursos (*features*) e de um modelo de classificação de interesse eficaz quanto às métricas coletadas; no entanto, há oportunidades para melhorias, como implementação de um sistema de ranqueamento de itens e aumento de personalização pela utilização de diferentes algoritmos de aprendizado por usuário.

**Palavras-chave:** Sistemas de Recomendação. Redes Sociais Baseadas em Eventos. EBSN. Redes Sociais. Rede Neural. XGBoost. Aprendizado de Máquina. ELM. Pruning ELM.

# Abstract

As access to the *Internet* gets popular, over time, various platforms have emerged with a focus on the socialization of its members. These platforms are commonly referred to as social networks. Among the existing social networks, there is a group called “Event-Based Social Networks” (EBSN), which aims to help in the organization and distribution of *online* and *offline* events. With millions of users and thousands of weekly events on EBSNs, there is a need to create services that provide personalized event recommendations for each user since manual searches can be unfeasible and demotivating. This work, therefore, proposes the development of an event recommendation service that takes into account features widely used in the literature, such as the user’s location and the event, as well as features that are not so common, such as the user’s emotions, feelings, and personality traits. To this end, this work has produced a new synthetic database for the class of problems and support for the recommendation service, based on a collected database; a script for collecting public data on real platforms, and, finally, the event recommendation service, from extracting features to collecting explicit feedback. Some comparative experiments were carried out using the aforementioned database to check the quality of the models used in the proposed recommender’s operating architecture. From the experiments, it was possible to see that one of the models inferred by one of the learning algorithm proposed for the recommendation service showed satisfactory results using the database created for a large number of the users tested; however, for others, improvements are necessary. This work contributes to the possible opening up of new research topics, as well as the development of a recommendation service and a new synthetic database. As for the proposed recommendation service itself, it can be concluded that the service uses a module with good features engineering and a satisfactory interest classification model when analyzing the collected metrics; however, there are clear opportunities for improvement, such as ranking recommended items and increasing personalization by using different learning algorithms per user.

**Keywords:** Event-based Social Network, EBSN, Social Networks, Recommendation Systems. ELM. Pruning ELM. XGBoost. Machine Learning.

# Lista de Ilustrações

Figura 2.1 – Esquema de Filtragem Colaborativa . . . . .	5
Figura 2.2 – Esquema de Filtragem Baseada em Conteúdo . . . . .	6
Figura 2.3 – Esquema de Filtragem Híbrida . . . . .	7
Figura 2.4 – Esquema de Casamento de Personalidade . . . . .	9
Figura 2.5 – Esquema de Filtragem de Personalidade . . . . .	9
Figura 2.6 – Esquema de Filtragem Híbrida de Personalidade . . . . .	10
Figura 2.7 – Esquema de análise de sentimento baseada em léxico. . . . .	11
Figura 2.8 – Esquema de análise de sentimento baseada em aprendizado de máquina. . . . .	12
Figura 2.9 – Esquema de análise de sentimento baseada em aprendizado profundo. . . . .	13
Figura 3.1 – Esquema de arquivos da base de dados criada. . . . .	18
Figura 3.2 – Exemplo de circunferência para seleção randômica de localidade (apenas demonstrativo). . . . .	19
Figura 3.3 – Distribuição de MBTI na República Popular da China . . . . .	20
Figura 3.4 – Arquitetura do serviço de recomendação. . . . .	23
Figura 3.5 – Esquema de recomendação e coleta de Feedback. . . . .	29
Figura 3.6 – Esquema da coleção de dados provenientes de outras plataformas. . . . .	30
Figura 3.7 – Distribuição de cidade por eventos. . . . .	30
Figura 4.1 – Gráfico de acurácia por usuário. . . . .	34
Figura 4.2 – Gráfico de precisão por usuário. . . . .	35
Figura 4.3 – Gráfico de recall por usuário. . . . .	35
Figura 4.4 – Gráfico de F1 por usuário. . . . .	36
Figura 4.5 – Gráfico de acurácia por usuário. . . . .	36
Figura 4.6 – Gráfico de precisão por usuário. . . . .	37
Figura 4.7 – Gráfico de Recall por usuário. . . . .	37
Figura 4.8 – Gráfico de F1 por usuário. . . . .	38
Figura 4.9 – Gráfico de acurácia por usuário. . . . .	38
Figura 4.10–Gráfico de precisão por usuário. . . . .	39
Figura 4.11–Gráfico de Recall por usuário. . . . .	39
Figura 4.12–Gráfico de F1 por usuário. . . . .	40
Figura 4.13–Gráfico de acurácia por usuário. . . . .	40
Figura 4.14–Gráfico de precisão média. . . . .	41
Figura 4.15–Gráfico de Recall médio. . . . .	41
Figura 4.16–Gráfico de F1 médio. . . . .	42

# Lista de Tabelas

Tabela 2.1 – Tabela de comparação de critérios . . . . .	16
Tabela 3.1 – Tabela de símbolos usados na arquitetura do sistema de recomendação . . .	24
Tabela 3.2 – Amostra do resultado da extração de traços de personalidades. . . . .	26
Tabela 4.1 – Tabela de valores utilizados para ajuste fino do <i>ELM</i> e <i>Pruning-ELM</i> . . . .	33
Tabela 4.2 – Tabela de valores utilizados para ajuste fino do <i>XGBoost</i> . . . . .	33

# Lista de Abreviaturas e Siglas

RS	Sistemas de Recomendação
CF	Filtragem Colaborativa
CF/UB	Filtragem Colaborativa Baseada em Conteúdo
CF/IB	Filtragem Colaborativa Baseada em Item
CF/MB	Filtragem Colaborativa Baseada em Modelo
CB	Filtragem Baseada em Conteúdo
HF	Filtragem Híbrida
NLP	Processamento de Linguagem Natural
EBSN	Rede Social Baseada em Conteúdo
PM	Casamento de Personalidades
PF	Filtragem de Personalidades
SVM	Máquina de Vetores de Suporte
MP	<i>Most-Popular</i>
BPR	<i>Bayesian Personalized Ranking</i>
MCLRE	<i>Multi-Contextual Learning to Rank Events</i>
NDCG	Ganho Cumulativo Descontado Normalizado
ELM	<i>Extreme Learning Machine</i>
OS-ELM	<i>Online Sequential Extreme Learning Machine</i>
P-ELM	<i>Pruning Extreme Learning Machine</i>
TF	Frequência de Termo
IDF	Frequência Inversa do Documento
TF-IDF	Frequência do Termo-Inversa da Frequência do Documento
AUC-ROC	Área Sob a Curva ROC
MBTI	<i>Myers-Briggs Type Indicator</i>

SMOTE	Técnica de Sobreamostragem Minoritária Sintética
BERT	<i>Bidirectional Encoder Representations from Transformers</i>
DB	Banco de Dados
IP	Identificação de Personalidade
IES	Identificação de Emoção e Sentimento
ARE	Arquitetura de Recomendação de Eventos

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Justificativa	2
1.2	Objetivos Geral e Específicos	2
1.3	Método de trabalho	3
1.4	Organização do Trabalho	3
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>4</b>
2.1	Fundamentação Teórica	4
2.1.1	Sistemas de Recomendação	4
2.1.2	Classificação de Personalidades	7
2.1.3	Uso da Emoção e Sentimento	9
2.2	Trabalhos Relacionados	12
<b>3</b>	<b>Desenvolvimento</b>	<b>17</b>
3.1	Banco de Dados: Arquitetura e Organização	17
3.1.1	Criação da base <i>Events</i>	17
3.1.2	Criação da base <i>Users</i>	18
3.1.3	Distribuição de Sentimentos e Emoções	20
3.2	Recomendador de Eventos	21
3.2.1	Módulo de Seleção de <i>Features</i>	23
3.2.1.1	<i>Features</i> Espaciais	23
3.2.1.2	<i>Features</i> Temporais	25
3.2.1.3	<i>Feature</i> Semântica	25
3.2.1.4	<i>Feature</i> Social	25
3.2.1.5	<i>Features</i> Psicológicas	26
3.2.2	Módulo de Treinamento e Classificação	27
3.2.3	Recomendação e <i>Feedback</i>	28
3.3	Modulo de Integração de Dados	29
<b>4</b>	<b>Experimentos e Resultados</b>	<b>31</b>
4.1	Avaliação dos modelos de recomendação	31
4.2	Ajustes dos modelos baseado no ELM	32
4.3	Ajustes do modelo XGBoost	32
4.4	Resultados	33
4.4.1	Resultados ELM	34
4.4.2	Resultados Pruning-ELM	34
4.4.3	Resultados XGBoost	35
4.4.4	Resultados Médios	35
<b>5</b>	<b>Considerações Finais</b>	<b>43</b>

5.1	Conclusão . . . . .	43
5.2	Trabalhos Futuros . . . . .	44
	<b>Referências . . . . .</b>	<b>45</b>

# 1 Introdução

Redes sociais baseadas em eventos (EBSN) são tipos específicos de redes sociais que propõem a interação *online* entre usuários, como as normais, mas possuem uma parte primordial não necessariamente presente nas outras, a proposta de interações *offline*. Estas redes auxiliam na criação, organização e distribuição de eventos. Por definição, evento é um acontecimento que reúne pessoas com um mesmo intuito e objetivo; em outras palavras, são festas, conferências, festivais e quaisquer outras atividades formais e informais. O crescimento da utilização destas plataformas na *internet* é exponencial e facilmente identificado, como mencionado por [Liu et al. \(2012\)](#). Algumas, como a Meetup, possuem, no ano de 2023, mais de 50 milhões de usuários e mais de 100.000 eventos semanais<sup>1</sup>.

Assim como apontado por [Wu et al. \(2023\)](#), apesar da extensa utilização destas plataformas gerar grandes benefícios para usuários e empresas, a presença de desafios é proporcional, como a busca por comunidades online às quais os usuários queiram aderir dentre inúmeras disponíveis. Com isso, diversas soluções são desenvolvidas com fim de minimizar estes problemas, sendo uma destas o desenvolvimento de serviços de recomendação. [Shah, Gaudani e Balani \(2016\)](#), [Thorat, Goudar e Barve \(2015\)](#) definem sistemas de recomendação como *softwares* que realizam recomendações de itens a usuários de acordo com requisitos específicos. Esta recomendação ocorre por meio da coleta de informações dos itens e dos clientes e da realização de uma análise da relação entre o cliente e os itens. Conforme a filtragem de itens utilizada para criar a lista de recomendação, esta é então classificada entre “Filtragem Colaborativa”, “Filtragem Baseada em Conteúdo” ou “Filtragem Híbrida”, apresentadas na Subseção 2.1.1.

A existência de serviços de recomendação em aplicações com grande número de usuários e itens é fundamental para uma boa experiência do usuário e o sucesso no alcance do objetivo final da plataforma. Por este motivo, empresas como *Google, Meta, Amazon, Netflix, LinkedIn* e *Meetup* empregam estes serviços em seus sistemas. No entanto, a realização deste serviço nem sempre é uma tarefa simples. Utilizando como exemplo a recomendação de eventos e a natureza psicológica do ser humano<sup>2</sup>, diversos são os fatores que podem levar uma pessoa a participar de uma atividade social ou recusá-la, tornando a recomendação personalizada, uma tarefa complexa.

No estudo de [Maeng, Jang e Li \(2016\)](#), que aborda motivações para frequências em festivais, são apresentados alguns fatores comuns a todos os participantes, sendo alguns dos principais: socialização, tempo familiar, fuga do dia-a-dia e novidades. Aprofundando-se na motivação de participação de eventos, [Wu e Pusaksrikit \(2018\)](#) mostram que traços da personalidade de uma pessoa podem contribuir para sua presença e, além disso, podem ajudar a identificar os possíveis

<sup>1</sup> Dados retirados da plataforma [www.meetup.com](http://www.meetup.com)

<sup>2</sup> “Natureza psicológica humana” corresponde aos aspectos inerentes à mente humana, ao comportamento, à cognição e às emoções.

motivos para sua participação; um dos citados foi que adultos com traços de neuroticismo tendem a participar mais frequentemente de alguns eventos como forma de relaxamento e tentativa de diminuir ou se livrar de depressão.

A seguir, a Seção 1.1 aborda os motivos que levam à realização deste trabalho, assim como sua importância para o mercado. Seguindo, a Seção 1.2 detalha os objetivos geral e específicos deste trabalho. A Seção 1.3 apresenta o método de trabalho. Por último, a Seção 1.4 detalha os capítulos subsequentes do trabalho.

## 1.1 Justificativa

Como mencionado, a presença de milhões de usuários e milhares de eventos diários implica na existência de problemas para os envolvidos, isto é, o organizador, o cliente e a plataforma intermediadora. Um dos maiores desafios enfrentados pelos usuários é encontrar eventos dos quais estes queiram participar em tempo hábil, ou seja, encontrar eventos antes que estes ocorram. Este problema é enfrentado também pela EBSN responsável pela distribuição do evento, visto que esta propõe o uso de seus serviços sob a condição de auxiliar o organizador no aumento do lucro e de auxiliar o usuário na busca de eventos.

Ainda sendo uma área com a presença de milhões de usuários e com diversas pesquisas realizadas, *datasets* focados neste problema não são tão acessíveis. Isso se deve, provavelmente, a problemas de privacidade e segurança de empresas ao liberarem dados para o desenvolvimento. Dessa forma, a disponibilidade de um *dataset* sintética mostra-se de grande contribuição.

Outro desafio presente por parte de usuários está em como realizar a busca de eventos. Geralmente, caso nenhum elemento na página principal de uma plataforma capte o interesse do usuário, este deve realizar uma busca via frases e palavras-chave, podendo não ser simples, visto que nem todos conseguem expressar as preferências concisamente. Sendo assim, surge a necessidade da aplicação de técnicas de recomendação de eventos para usuários de forma personalizada e automática, visto que diferentes pessoas possuem gostos diferentes.

Assim sendo, diante da complexidade e do impacto social e financeiro do problema, a relevância deste trabalho está na concepção de um novo serviço de recomendação, que considera aspectos da natureza psicológica humana, somando-se a outros aspectos considerados, atualmente, no mercado.

## 1.2 Objetivos Geral e Específicos

Este trabalho tem, por objetivo geral, a proposta, o desenvolvimento e a validação experimental de um serviço de recomendação de eventos que utilize de aspectos intrínsecos à psicologia, como diferentes traços de personalidades, com o intuito de recomendar eventos corretos aos

usuários. Vale ressaltar que, neste trabalho, considera-se um evento como sendo correto, quando o usuário ao qual este foi recomendado possui interesse no evento.

Os objetivos específicos alcançados neste trabalho, são:

- criação de uma base de dados para suporte da aplicação e do serviço;
- previsão de interesse de usuários por eventos por meio da proposta de modelos classificadores para tal fim;
- proposta de módulo de integração de dados de outras plataformas.

### 1.3 Método de trabalho

Para alcançar o objetivo geral descrito, foi elaborado e desenvolvido um serviço de recomendação de eventos a usuários. Tal serviço seguiu uma arquitetura modular, com 3 módulos principais: extrator de recursos, classificador e recomendador; este foi validado por meio de experimentos utilizando diferentes *features* e diferentes classificadores, ambos propostos neste presente trabalho. Um módulo para obtenção de dados de outras EBSNs foi desenvolvido e disponibilizado publicamente. Por fim, uma nova base de dados foi desenvolvida tendo como referência bases já utilizadas em outros trabalhos relacionados.

### 1.4 Organização do Trabalho

Para facilitar a leitura do trabalho, este foi dividido em 6 capítulos: presente Capítulo 1 introduz, detalha e especifica o tema e objetivos do trabalho; Capítulo 2 descreve informações e conceitos importantes para entendimento do tema e resume alguns trabalhos relacionados; Capítulo 3 apresenta em detalhes a arquitetura do serviço proposto de recomendação, a extração dos dados necessários, a organização dos dados utilizados e, por fim, o módulo de integração de dados vindos de outras plataformas. Seguindo, o Capítulo 4 apresenta os experimentos realizados para validação do desenvolvimento, o Capítulo 4.4 aborda os principais resultados obtidos durante as experimentações e, por fim, o Capítulo 5 apresenta a conclusão obtida pelos experimentos acerca do serviço proposto, pontos positivos e negativos da arquitetura do serviço de recomendação e a perspectiva de trabalhos futuros.

## 2 Revisão Bibliográfica

Este capítulo busca, por meio de uma adequada fundamentação teórica (Seção 2.1) e uma discussão de trabalhos relacionados ao tema proposto (Seção 2.2), realizar uma contextualização da pesquisa para melhor entendimento do assunto.

### 2.1 Fundamentação Teórica

Esta seção apresenta conceitos essenciais usados durante o trabalho e durante a revisão de trabalhos relacionados. Para melhor compreensão destes conceitos, a seção é dividida em três subseções: a Subseção 2.1.1 aborda o conceito, métodos e problemas de sistemas de recomendação; a Subseção 2.1.2 apresenta abordagens de sistemas de recomendação baseadas na extração de traços de personalidade de um usuário, e por fim, a Subseção 2.1.3 define os passos para a análise de sentimentos e emoções e a diferença entre emoção e sentimento.

#### 2.1.1 Sistemas de Recomendação

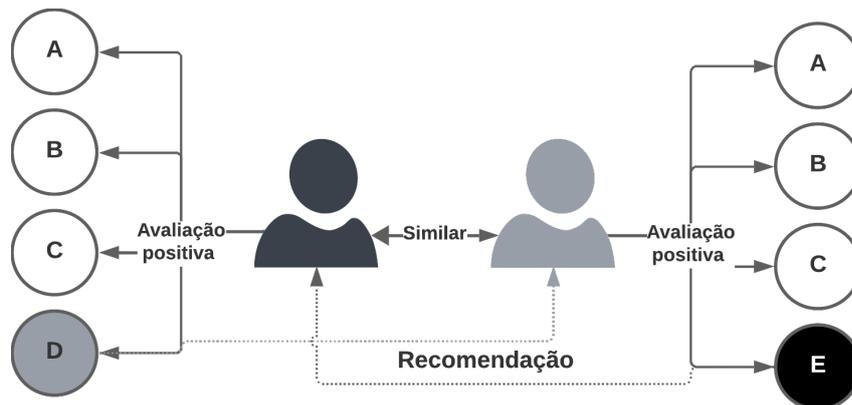
Com o avanço da tecnologia e a oferta de serviços *online*, o número de informações disponíveis tem um crescimento significativo. Sendo assim, com um vasto e rápido acesso a informações, surge a necessidade da criação de sistemas capazes de facilitar a busca de informações relevantes para cada usuário. Os sistemas de recomendação surgem com este intuito, servindo de ferramentas de filtragem de dados (Roy; Dutta, 2022).

Como citado por Ko *et al.* (2022), os modelos de sistemas de recomendação existentes são baseados em três principais técnicas: Filtragem Colaborativa, Filtragem Baseada em Conteúdo e Filtragem Híbrida.

A técnica de *Filtragem Colaborativa (CF)* consiste em criar um modelo de recomendação baseado em banco de preferências dos usuários do sistema. Estas preferências são adquiridas armazenando classificações de itens feitas por usuários do sistema, ou seja, ao final a abordagem busca usuários ou itens parecidos. Existem três principais categorias da CF, a **baseada em usuário (CF/UB)**, a **baseada em item (CF/IB)** e a **baseada em modelo (CF/MB)**. As duas primeiras são também conhecidas como baseadas em memória e consistem comparar similaridades. No caso da CF/UB, a comparação de similaridade é feita entre dois usuários distintos do sistema, analisando a avaliação de um mesmo elemento. Já na CF/IB, a recomendação é feita baseada na similaridade entre um elemento não analisado e um item já avaliado pelo usuário. A última classificação (CF/MB) utiliza do aprendizado de um modelo para predizer uma avaliação de um item baseado em avaliações do usuário (Ko *et al.*, 2022; Bobadilla; Alonso; Hernando, 2020).

A Figura 2.1 detalha o esquema da técnica de CF. Para efetuar a recomendação, dois usuários A (preto) e B (cinza) são comparados e um valor de similaridade definido. Por serem similares, é pressuposto que o gosto entre itens em uma mesma plataforma seja similar. Assim sendo, é feita uma recomendação ao usuário B de itens consumidos por A e não consumidos por B, assim como ao usuário A de itens consumidos por B e não por A.

Figura 2.1 – Esquema de Filtragem Colaborativa



Fonte: Autoria própria

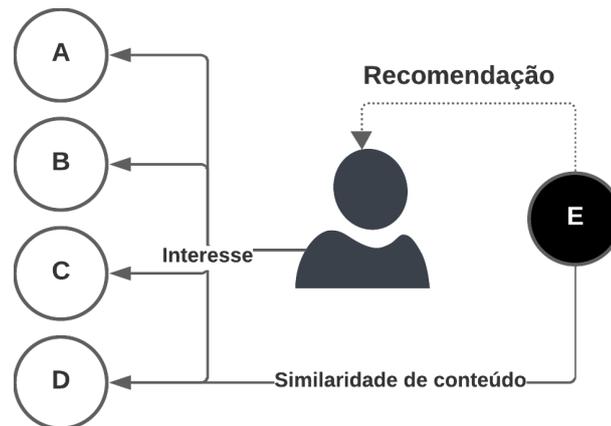
A **Filtragem Baseada em Conteúdo (CB)** é uma técnica com um foco no perfil dos usuários e uma das mais simples entre os modelos de sistemas de recomendação. A CB é caracterizada pela análise do conteúdo de um item, isto é, é um modelo quase independente do usuário. Durante o processo de recomendação, o conteúdo (vídeo, foto, áudio, texto, gênero, entre outros) de um elemento é extraído, analisado e comparado ao conteúdo de um elemento previamente selecionado pelo usuário.

A Figura 2.2 detalha o esquema da técnica de CB. Esta técnica é semelhante à CF; no entanto, para efetuar a recomendação, dois itens D (cinza) e E (preto) são comparados e um valor de similaridade baseado em seus conteúdos é definido. Por serem similares, é pressuposto que o usuário interessado em D esteja também interessado em E, por possuírem conteúdo semelhante. Vale ressaltar que, para a comparação, é necessário que um dos itens pertença a uma lista de interesses do usuário analisado.

Por fim, a técnica de **Filtragem Híbrida (HF)** consiste da mescla de duas ou mais técnicas de recomendação (Tan; He, 2017). Como mencionado por Ko *et al.* (2022), os modelos híbridos são divididos em 7 categorias, sendo estas:

- **Hibridização Ponderada:** combina a avaliação de diferentes algoritmos para formar um único;
- **Comutação de Hibridização:** realiza a seleção de um dos diferentes algoritmos para ser usado como o recomendador;

Figura 2.2 – Esquema de Filtragem Baseada em Conteúdo



**Fonte:** Autoria própria

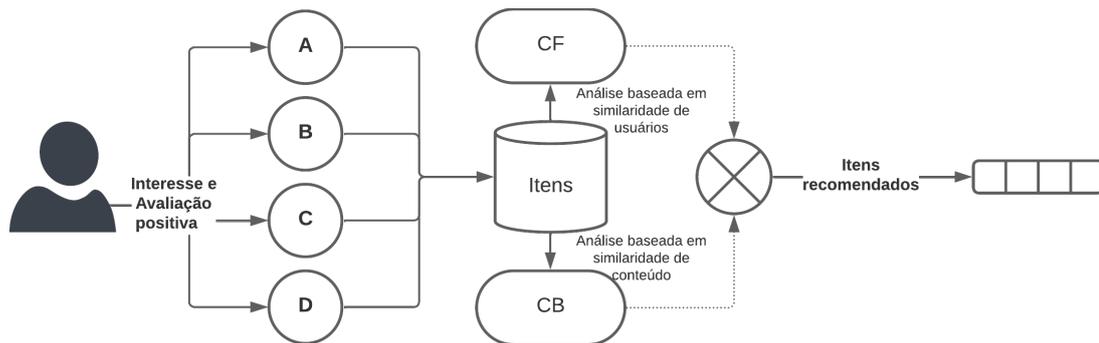
- **Hibridização em Cascata:** utiliza, como entrada, a saída de um algoritmo anterior e realiza refinações nesta saída;
- **Hibridização Mista:** apresenta a recomendação de diferentes algoritmos ao mesmo tempo;
- **Hibridização de Combinação de Recursos:** utiliza recursos de fontes diferentes por diferentes algoritmos ao mesmo tempo;
- **Hibridização de Aumento de Recursos:** utiliza o resultado de um algoritmo como dado de entrada de um próximo;
- **Hibridização Meta-aprendizado:** utiliza o modelo gerado por algoritmos anteriores como dados de entrada.

A Figura 2.3 detalha o esquema da técnica de HF. Como mencionado, esta técnica é a junção das duas anteriores. Para efetuar a recomendação, a lista de interesses e avaliações do usuário (lista A) é comparada com uma lista de itens do banco (lista B) de forma que: na etapa de CB, os itens da lista A são comparados em conteúdo com os da lista B e uma recomendação é gerada; na etapa de CF, os itens da lista A, bem avaliados, são comparados com os bem avaliados por usuários semelhantes na lista B; ao encontrar os não consumidos pelo usuário em análise, uma nova recomendação é gerada. Ao fim, ambas as recomendações geradas nas duas etapas formam a lista de recomendação final.

Como mencionado por Roy e Dutta (2022) e Dhelim *et al.* (2021), diversos desafios, apesar dos esforços, dos estudos constantes e da presença no mercado, permanecem existentes em distintas áreas como: plataformas de notícias, filmes, músicas, amizades, artigos, locais de turismo, fotos, dentre outros. Alguns dos principais desafios são:

- **Filtragem ou Partida a frio (Cold-Start):** ocorre quando um sistema não possui dados suficientes de um novo usuário ou item e, portanto, não pode realizar uma recomendação relevante;

Figura 2.3 – Esquema de Filtragem Híbrida



Fonte: Autoria própria

- **Esparcidade e Qualidade de Dados:** ocorre quando há escassez de dados de interação do usuário ou falta de qualidade dos dados, dificultando a recomendação precisa;
- **Superespecialização:** ocorre quando os itens recomendados são sempre muito semelhantes aos outros; em outras palavras, é a falta de diversidade nos itens recomendados;
- **Viés Algorítmico:** ocorre quando o sistema realiza recomendação de conteúdos tendenciosos, perpetuando a existência (ou até mesmo amplificando) de estereótipos ou quaisquer outros vieses existentes;
- **Usuários Atípicos (*Grey Sheep Problem*):** ocorre quando há a existência de usuários com preferências tão únicas que o sistema não consegue classificá-lo em uma categoria usual;
- **Privacidade e Segurança:** ocorre quando o armazenamento de dados e a análise de comportamento dos usuários geram preocupações de privacidade e segurança.

### 2.1.2 Classificação de Personalidades

Um dos métodos utilizados para resolver os problemas de falta de dados e *cold start*, citados anteriormente, é a aplicação de teorias de personalidade estudadas no campo da psicologia. Com essa técnica, como mencionado por *Dhelim et al. (2021)*, surge a computação de personalidades, que visa a aplicação de diversas teorias da psicologia para identificação do tipo de pessoa que é o usuário.

Sistemas de recomendação baseados em personalidade são como os comuns, mas considerando os traços e informações sobre a personalidade do usuário. Geralmente duas etapas são adicionadas à recomendação: a medida de personalidade e a fase de casamento de personalidade. Na primeira, o sistema utiliza de questionários ou algoritmos de identificação automática de personalidade (APR) para avaliar o tipo de pessoa que é o usuário. Na segunda, o sistema avalia a probabilidade de casamento de um elemento e uma pessoa, baseado nos traços da personalidade dela.

Os sistemas de recomendação baseados em personalidade são subdivididos em métodos de filtragem e *deep learning*. Os métodos baseados em filtragem seguem a divisão padrão de RS, como CB, CF e HF, onde a CF possui métodos de vizinhança de personalidade e fatoração de matrizes. Os métodos de recomendação baseados em CF são chamados de *Métodos de Filtragem de Personalidade*, os baseados em conteúdo são conhecidos como *Casamento de Personalidade*, e, por fim, os híbridos são chamados de *Filtragem Híbrida de Personalidade*, encontrando-se descritos a seguir.

- **Casamento de Personalidades (PM):** é bem semelhante à abordagem padrão de filtragem baseada em conteúdo. Via uma análise de atributos ou conteúdo, cada item recebe uma pontuação de personalidade. Pela análise de conteúdo (análise de descrição, categorias), um item recebe a pontuação por um método de reconhecimento automático de personalidade (APR). Já na análise de atributos, a pontuação é dada ao item baseado em traços de um usuário que tenha interagido com aquele item.
- **Filtragem de Personalidades (PF):** Estas técnicas são como as filtrações colaborativas já mencionadas, usando funções de medição de proximidade (como correlação de Pearson<sup>1</sup>) para criação de uma vizinhança de personalidades similares. Outra forma é pela fatoração de matriz, onde o método convencional é apenas adaptado para adotar características da personalidade do usuário.
- **Filtragem Híbrida de Personalidades:** Combinando as duas últimas abordagens, a filtragem híbrida aplica a PF no espaço do usuário e a PM no espaço de itens.

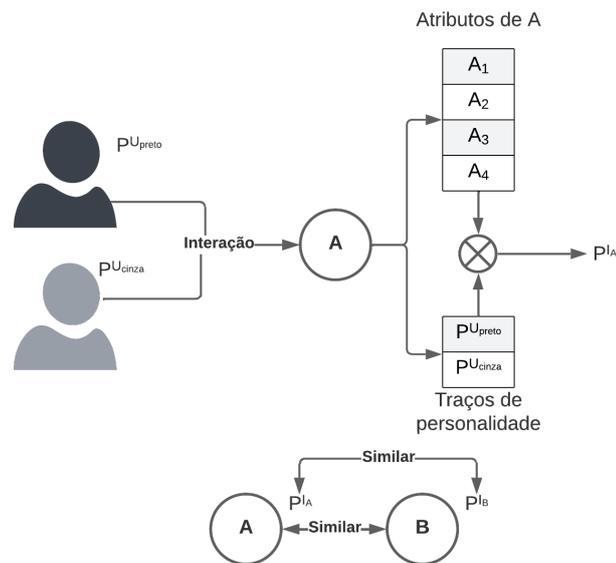
Destas, as técnicas de filtragem híbrida mostram-se bastante promissoras, como mencionado por [Dhelim et al. \(2021\)](#), dado que conseguem aproveitar as vantagens de ambas as outras abordagens.

A Figura 2.4 detalha o esquema da técnica de PM. Inicialmente, traços de personalidade são atribuídos a um item, baseado nos traços de personalidade dos usuários (preto e cinza) que interagiram com este item, com o auxílio de outros atributos do item. Por fim, ocorre uma comparação de similaridade entre dois itens comparando os traços de ambos, assim como na técnica de CB mencionada.

A Figura 2.5 detalha o esquema da técnica de PF. Para efetuar a recomendação, os traços de personalidade de dois usuários  $P^{U_{preto}}$  e  $P^{U_{cinza}}$  são comparados. Por serem similares, é pressuposto que os interesses sejam similares; portanto, itens consumidos por A (preto) e não por B (cinza) são recomendados ao B, assim como os consumidos por B e não por A são recomendados ao A.

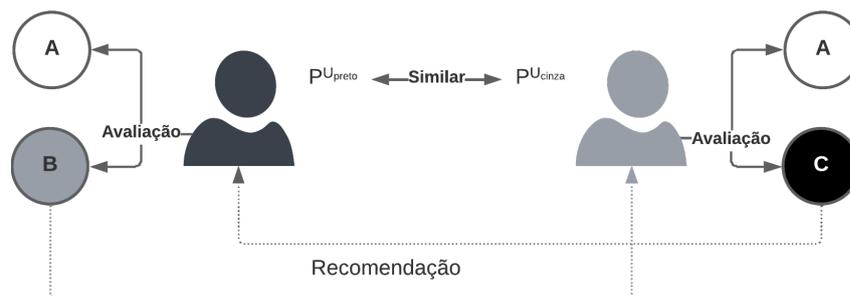
<sup>1</sup> Segundo [Boslaugh \(2012\)](#), a Correlação de Pearson avalia a relação linear entre duas variáveis por meio de um coeficiente que varia de -1 a 1. Quanto mais próximo de 1 e -1, maior a proporcionalidade entre as variáveis. Para o cálculo, a covariância entre as variáveis é dividida pelo produto dos desvios padrão de cada uma das variáveis.

Figura 2.4 – Esquema de Casamento de Personalidade



Fonte: Autoria própria

Figura 2.5 – Esquema de Filtragem de Personalidade



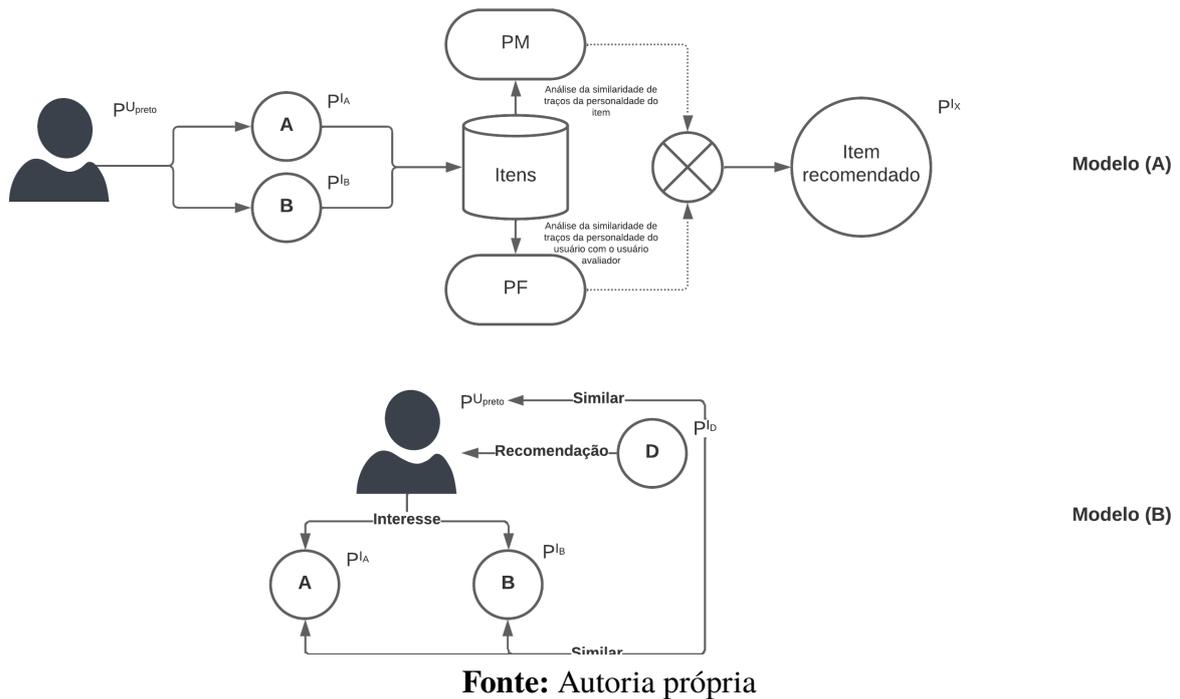
Fonte: Autoria própria

A Figura 2.6 detalha o esquema da técnica de HF. Para exemplificação de técnicas híbridas, dois modelos são apresentados. No modelo A, assim como na Figura 2.3, a técnica é a junção das duas anteriores. Para efetuar a recomendação, a lista de interesses e avaliações do usuário (lista A) é comparada com uma lista de itens do banco (lista B) de forma que: na etapa de PM, os traços de personalidade dos itens da lista A são comparados em conteúdo com a lista B e uma recomendação é gerada; na etapa de PF, os itens da lista A bem avaliados são comparados com os bem avaliados por usuários que possuem personalidade semelhante na lista B; ao fim, ambas recomendações geradas formam a lista de recomendação final. No modelo B, algo semelhante é feito; no entanto, ocorre a verificação da similaridade da personalidade de um item e um usuário ao invés da simples similaridade entre dois usuários.

### 2.1.3 Uso da Emoção e Sentimento

Em qualquer tipo de interação social, o reconhecimento emocional dos envolvidos é um passo fundamental para uma boa comunicação e interação. Com a melhora dos serviços oferecidos pela *internet*, usuários utilizam de meios digitais (principalmente redes sociais) para compartilhar

Figura 2.6 – Esquema de Filtragem Híbrida de Personalidade



experiências, sentimentos e opiniões. Sendo assim, o estudo de técnicas para reconhecimento de sentimentos e emoções surge com o intuito de permitir que máquinas operem como agentes sociais e entendam melhor o comportamento dos usuários, como mencionado por (Shelke *et al.*, 2022).

Apesar da análise de sentimentos e análise de emoções serem comumente utilizados com um mesmo significado, há diferenças a serem esclarecidas. Como descrito por Taboada (2016), o primeiro visa determinar se um texto (no contexto de análise por meio de textos) é subjetivo, e, caso seja, se possui posição positiva ou negativa. Já o segundo tem por finalidade a detecção dos diferentes tipos de emoções, como fúria, felicidade, surpresa, desprezo, tristeza, entre outros (Nandwani; Verma, 2021). Neste documento, análise de sentimentos e de emoções são usados conforme a descrição feita.

Ambas análises baseadas em texto funcionam com técnicas de processamento de linguagem natural (NLP) e possuem as seguintes etapas (Acheampong; Wenyu; Nunoo-Mensah, 2020; Xu; Chang; Jayne, 2022; Shelke *et al.*, 2022):

- **Pré-processamento do texto:** etapa onde ocorre a limpeza dos dados e do texto (remoção de palavras vazias, por exemplo);
- **Extração de recursos:** etapa onde dados importantes para classificação são extraídos do texto (frequência de cada palavra, por exemplo);
- **Classificação:** etapa onde um valor é dado e há uma classificação das palavras e frases (nessa etapa pode haver uma análise do texto em vários níveis);

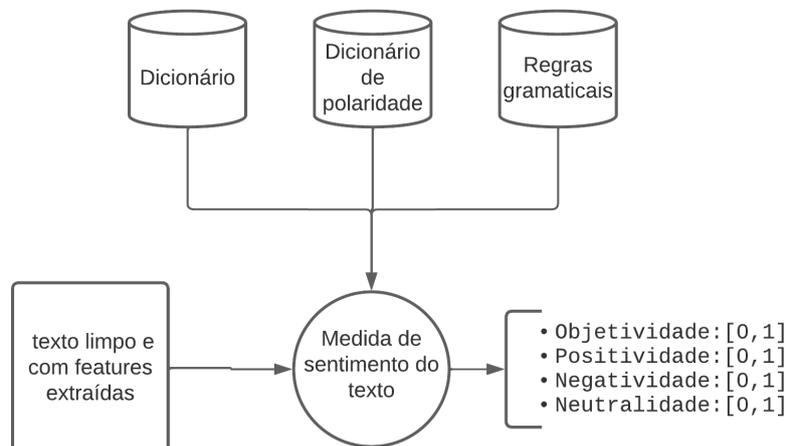
- **Validação:** etapa onde o resultado é validado pelo pesquisador.

Alguns dos métodos mais utilizados para a classificação de sentimentos e emoções são mencionados por [Nandwani e Verma \(2021\)](#), sendo estes:

- **Baseados em léxico:** são métodos que atribuem valores que representam emoções e sentimentos a textos e palavras baseado em um dicionário pré-definido de léxicos;
- **Baseados em aprendizado de máquina:** são métodos que utilizam algoritmos de aprendizado para rotular e classificar textos e palavras baseados em um padrão existente;
- **Baseados em aprendizado profundo:** são métodos que utilizam redes neurais artificiais com diversas camadas ocultas para aprender e extrair características de textos.

A Figura 2.7 detalha um exemplo de como ocorre a análise de sentimento de um texto baseada em léxico. Para a análise, é necessário um dicionário de polaridade, um vocabulário e, opcionalmente, um banco de regras gramaticais da língua em análise. Com um texto limpo (remoção de palavras vazias, e outros caracteres sem valor) e *features* extraídas como, por exemplo, a ocorrência de cada palavra em um texto usada para determinar a polaridade de um texto, basta realizar uma operação como a soma da polaridade de cada palavra. Com isso, obtém-se o valor de cada critério em análise, como objetividade, neutralidade, positividade e negatividade.

Figura 2.7 – Esquema de análise de sentimento baseada em léxico.

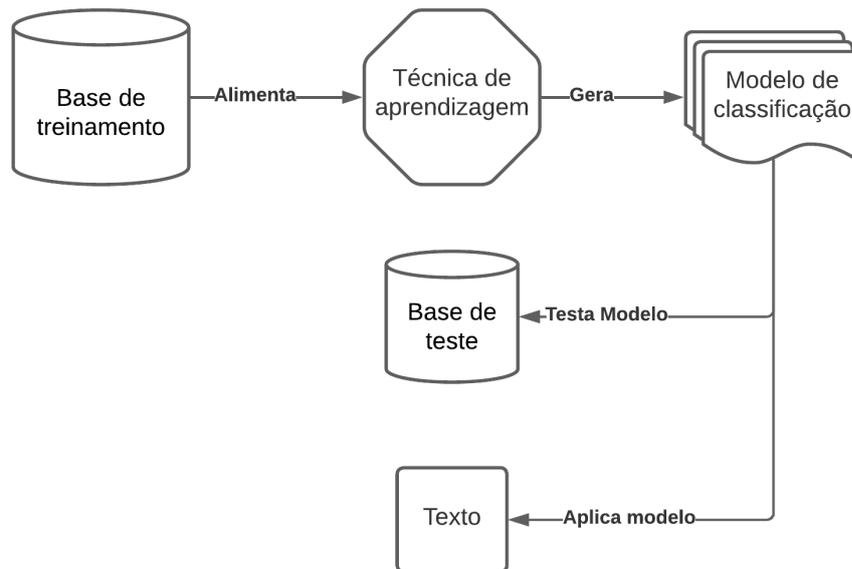


**Fonte:** Autoria própria

A Figura 2.8 detalha um esquema de uso de aprendizado de máquina para análise de sentimento. Para isso, é necessário realizar a coleta de textos para que uma base de treinamento seja montada. Com a base limpa, *features* extraídas e convertidas em representação numérica, usa-se uma técnica de aprendizado, como o algoritmo “Máquina de vetores de suporte (SVM)”<sup>2</sup>, para inferir um modelo de classificação. Este modelo é responsável pela classificação final, aplicado em uma base de teste e no texto em si.

<sup>2</sup> Segundo [Mammone, Turchi e Cristianini \(2009\)](#), SVM é um algoritmo de aprendizado de máquina que busca geralmente classificar itens por meio de um ou mais hiperplanos que separe as instâncias no espaço.

Figura 2.8 – Esquema de análise de sentimento baseada em aprendizado de máquina.



**Fonte:** Autoria própria

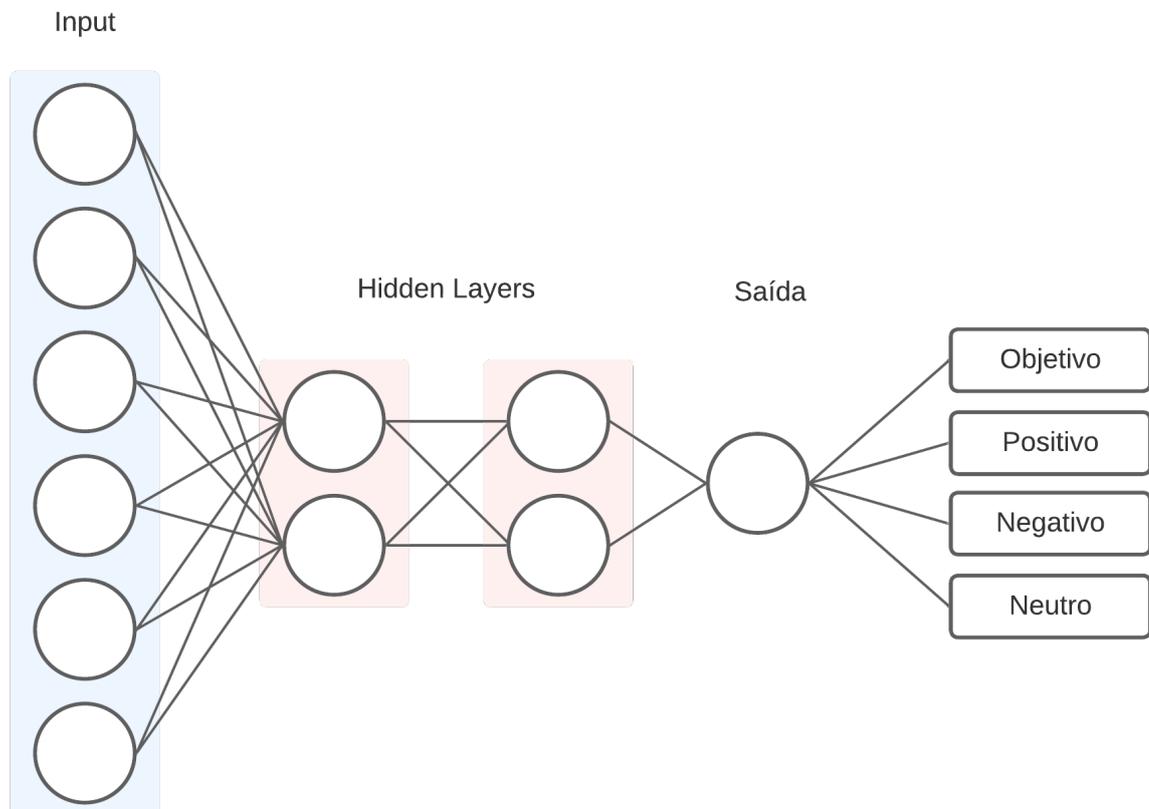
A Figura 2.9 detalha um esquema semelhante ao da Figura 2.8. Modelos de aprendizado profundo são modelos de aprendizado de máquina e, portanto, comportam-se quase como no esquema abordado. Por meio de uma camada de dados de entrada (*input layer*), o texto, já limpo e convertido em representação numérica, é alimentado a camadas chamadas de Camadas Ocultas (*hidden layers*), que internamente são responsáveis por aprender e abstrair *features* dos dados de entrada. Por fim, as camadas ocultas fornecem seus resultados à camada de saída que então classifica o texto nos grupos desejados.

## 2.2 Trabalhos Relacionados

Nesta seção, são apresentados trabalhos semelhantes em algum grau à proposta do trabalho em questão. Para seleção dos trabalhos semelhantes, os seguintes pontos foram considerados: tema principal, ano de publicação e palavras-chave do trabalho relacionado. A pesquisa foi realizada com a busca de palavras-chave em inglês envolvendo eventos, sistemas de recomendação, reconhecimento de emoções, reconhecimento de sentimentos, classificação de personalidade e variações destas palavras. Por fim, foram selecionados trabalhos mais recentes (2015–2023) e de preferência que utilizam base de dados em larga escala e propõem recomendações em tempo real. Ao todo, 4 artigos foram selecionados.

*Wang et al. (2017)* focam no problema de recomendação personalizada de eventos em redes sociais baseadas em eventos (EBSN), sendo o objetivo principal a recomendação dos eventos mais relacionados ao usuário. Como nesta área de sistemas de recomendação o problema de *cold start* é facilmente encontrado, os autores consideram as informações contextuais do evento, assim como influência social dos anfitriões do evento. O modelo proposto é denominado *Social Information Augmented Recommender System (SIARS)* e este unifica os modelos de

Figura 2.9 – Esquema de análise de sentimento baseada em aprendizado profundo.



**Fonte:** Autoria própria

recomendação que são conscientes do anfitrião, tempo, localização, membros e conteúdo para realizar os cálculos das pontuações das recomendações. Ao fim, os  $n$ -melhores eventos são recomendados, ou seja, os  $n$ -eventos com as maiores avaliações pelo sistema. No problema estudado, um evento é formado por uma descrição, um anfitrião, convidados, localização e data.

O *SIARS* explora o histórico de feedback do anfitrião, o histórico de conteúdo do evento e a influência social deste anfitrião. Estes três aspectos possuem o peso modificado e ajustado pelo algoritmo *AdaBPR*<sup>3</sup>. Além disso, os autores propõem a exploração das interações entre os membros de grupos de usuários para cada evento e a extração e análise de tópicos interessantes para usuários, baseados em eventos já participados. O *SIARS* foi avaliado usando métricas de precisão e revocação e utilizando dados coletados do aplicativo Meetup<sup>4</sup> e eventos ocorridos nas cidades de Phoenix, Chicago e San Jose, entre 2012 e 2014; foram também coletadas informações sobre eventos ocorridos em Nova Iorque entre 2014 e 2016. Ao comparar o *SIARS* com outros modelos como *Most Popular (MP)*<sup>5</sup>, *Bayesian Personalized Ranking (BPR)*<sup>6</sup> e *Multi-Contextual*

<sup>3</sup> Segundo Liu *et al.* (2015), Ada-BPR é um algoritmo no qual múltiplos recomendadores de componentes são linearmente combinados. Neste algoritmo, os componentes são construídos usando filtragem colaborativa com pesos dinâmicos nas interações entre usuário e item.

<sup>4</sup> Meetup é uma plataforma que propõe a possibilidade de usuários participarem de grupos e eventos com o intuito de conhecerem novas pessoas e aprenderem novos conteúdos. Mais detalhes disponíveis em: <https://www.meetup.com/>

<sup>5</sup> Segundo Wang *et al.* (2017), é um algoritmo que recomenda os eventos mais populares.

<sup>6</sup> É um algoritmo de fatoração de matrizes para recomendação de itens a partir de feedback implícito. Rendle *et al.*

*Learning to Rank Events (MCLRE)*<sup>7</sup>, o proposto apresentou valores de precisão e revocação melhores e com melhor desempenho.

Yin *et al.* (2018) têm como proposta um novo paradigma de recomendação de eventos que considera amizades e conexões entre pessoas no mundo virtual; não apenas amizades diretas são consideradas, mas todo um círculo social com diferentes níveis de conexão, como amigos de amigos. O modelo proposto é baseado em grafos para o aprendizado da interação entre usuário e eventos. Para contornar o problema de *cold start*, as informações de conteúdo e contexto dos eventos são considerados como em Wang *et al.* (2017), ou seja, localização, tempo e data e descrição. Ao todo, o modelo proposto usa 5 grafos bipartidos, estes são: 1. usuário  $\times$  usuário; 2. usuário  $\times$  evento; 3. evento  $\times$  localização; 4. evento  $\times$  data; 5. evento  $\times$  palavra.

O modelo proposto foi avaliado em eficiência de recomendação *online* com o uso de dois métodos de produção das  $n$ -melhores recomendações e um banco de dados coletado da plataforma chinesa Douban Event<sup>8</sup>, de 2005 a 2012. Em média, um dos métodos, denominado **GEM-TA** obteve as 10 melhores recomendações de pares eventos-parceiros em uma média de 5,45 segundos, com um total de 2.590 novos eventos e 64.113 usuários. Ao filtrar o espaço de busca do algoritmo e armazenando os  $k$ -melhores eventos para cada usuário, os algoritmos atingem eficiência próxima a 1, quando  $k$  é maior que 5% do total de eventos. Quanto ao tempo, o algoritmo GEM-TA possui tempo quase linear em respeito a  $k$ . Yin *et al.* (2018) apresentam resultados satisfatórios e concluem que as técnicas propostas para o problema estudado são altamente escaláveis e eficientes.

Em um trabalho mais recente, Liang (2023) segue uma linha um pouco diferente da tendência de estudos. Ele propõe a recomendação de atividades em grupos (eventos) nas *EBSNs* considerando *fairness* e diversidade. O motivo que levou a este estudo está na demanda de considerar que uma recomendação satisfará as necessidades de todos os membros de um grupo. O objetivo final em sua pesquisa é o aumento de acurácia, diversidade e *fairness* em recomendações de atividades em grupos (eventos). Para verificação da acurácia, foram utilizadas as métricas de revocação e ganho cumulativo descontado normalizado (NDCG); a métrica NDCG considera a relevância e ordem atribuída aos itens de uma lista.

Com o modelo proposto denominado *DFGR*, um grafo de similaridade entre eventos é criado considerando informações contextuais; depois, o problema é transformado em um problema de maximização de cobertura ponderada, classificado como *NP-Hard*. Uma solução próxima à ótima é alcançada mediante um algoritmo guloso. Para experimentação, foram utilizados dados coletados das plataformas Douban e Meetup com eventos nas cidades de Phoenix, Chicago,

---

(2012)

<sup>7</sup> É um algoritmo de aprendizado de ranqueamento; o autor utiliza resultado de diferentes recomendadores como característica de aprendizado. Os recomendadores utilizados consideram informações de tempo, localização, conteúdo e membros presentes. Macedo, Marinho e Santos (2015)

<sup>8</sup> Douban é uma rede social chinesa que permite o registro e criação de conteúdos referentes a temas como música, filmes, livros, eventos e atividades. Mais detalhes em: <https://www.douban.com/>

Pequim e Xangai. Ao concluir, Liang (2023) declara que ao comparar o método com 7 outros métodos de recomendação, o proposto *DFGR* é superior aos demais métodos em recomendação de atividades em grupos nos três indicadores considerados (*fairness*, diversidade e acurácia).

Além da recomendação padrão de eventos comumente estudada, Nkuna e Coetzee (2020) mostram uma preocupação com o histórico de presença do usuário e também com a presença de indivíduos de um mesmo círculo social para aumento da satisfação do usuário. Para realização da recomendação, é aplicada uma técnica de *HF*, onde *CB* é usada para indicação de eventos e *CF* é usada para mapeamento de usuários com interesses similares. O modelo proposto atingiu 82,97% de acurácia, e algumas limitações foram apontadas pelos autores, sendo duas das principais as seguintes:

- a abordagem necessita que o usuário tenha usado o sistema por um tempo para que este consiga realizar boas estimativas de seus interesses; em outras palavras, a abordagem não consegue contornar o problema de *cold start*;
- a análise de sentimentos utilizada pode levar a classificações incorretas, visto que cada avaliação e crítica dos eventos são dadas como independentes, o que não é sempre o caso.

Os autores recomendam a investigação futura de técnicas como *n*-grama para melhor entendimento do contexto de avaliações e lematização de grupo de palavras.

Para melhor entendimento, a Tabela 2.1 compara os critérios de recomendação utilizados nos trabalhos acima mencionados. Os critérios foram classificados em categorias mais genéricas para facilitar a comparação; por exemplo, o conteúdo de um evento envolve palavras-chave, textos, imagem, categoria e tipo do evento.

A partir de uma análise comparativa na Tabela 2.1 é possível ver que localização, tempo e conteúdo são critérios tidos como básicos na recomendação de eventos, visto que são dimensões unânimes nos trabalhos mencionados. Já os demais são dimensões extras e podem auxiliar em uma melhor recomendação e, portanto, possuem diferenças entre os trabalhos. Este trabalho difere dos demais pela junção dos diversos critérios sugeridos nos trabalhos mencionados, introduzindo o uso de traços psicológicos dos usuários e abstendo-se do uso de patrocinadores devido a não relação direta entre a participação de um usuário em um evento e o patrocinador.

Tabela 2.1 – Tabela de comparação de critérios

	Wang <i>et al.</i> (2017)	Yin <i>et al.</i> (2018)	Liang (2023)	Nkuna e Coetzee (2020)	Trabalho atual
<b>Localização</b>	✓	✓	✓	✓	✓
<b>Tempo</b>	✓	✓	✓	✓	✓
<b>Conteúdo</b>	✓	✓	✓	✓	✓
<b>Organizador</b>	✓	✗	✗	✗	✓
<b>Membros</b>	✓	✓	✗	✓	✓
<b>Históricos</b>	✓	✗	✗	✓	✓
<b>Patrocinador</b>	✗	✗	✓	✗	✗
<b>Preferências</b>	✗	✗	✓	✓	✓
<b>Traços Psicológicos</b>	✗	✗	✗	✗	✓

## 3 Desenvolvimento

Este capítulo tem por finalidade a descrição e o detalhamento do desenvolvimento deste trabalho. Para tanto, encontra-se organizado da seguinte forma: Seção 3.1 aborda a criação, a obtenção e a organização dos dados usados para confecção do banco de dados de suporte ao serviço proposto. Seção 3.2 aborda os principais módulos do serviço de recomendação, da extração de *features* à criação da lista de recomendação, apresentados na Figura 3.4. Seção 3.3 aborda a coleta de dados provenientes de outras plataformas para a integração e a melhoria do recomendador.

### 3.1 Banco de Dados: Arquitetura e Organização

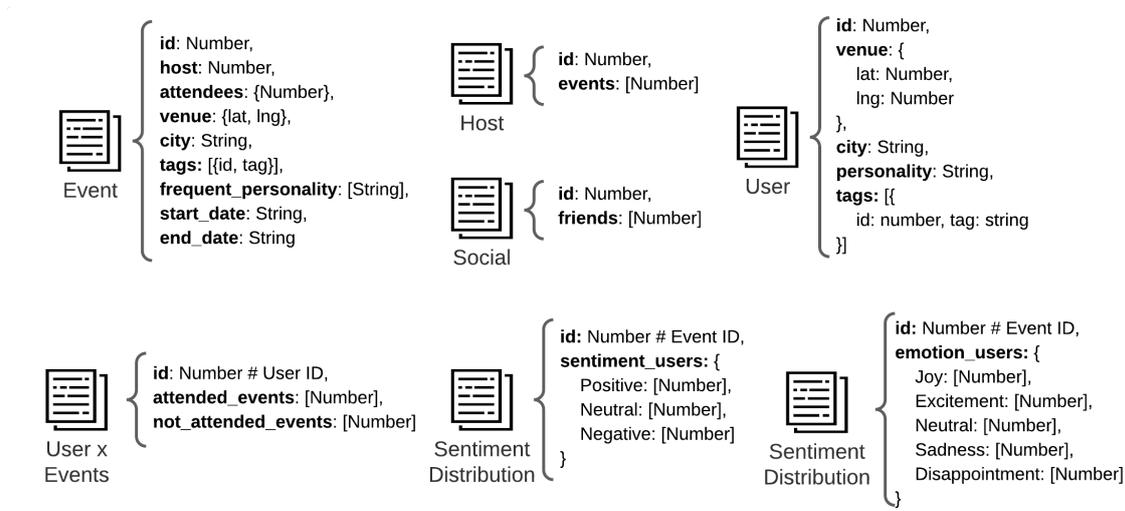
Dada a natureza específica do problema e do ambiente ao qual o problema está inserido, a presença de *datasets* que possuam dados pertinentes para as análises é um desafio. Alguns *datasets* são encontrados; no entanto, possuem poucos dados devido a problemas de privacidade. Sendo assim, um novo *dataset* foi criado<sup>1</sup> tendo como base o utilizado no trabalho (Yin *et al.*, 2018) e nele encontra-se os seguintes dados: eventos, usuários, sentimentos, emoções, conexão entre usuários, e relação de presença de usuários em eventos. A Figura 3.1 representa os dados presentes na base de dados. A identificação dos eventos, dos usuários, assim como as relações usuário  $\times$  usuário e usuários  $\times$  eventos estavam disponíveis no *dataset* original. Quanto ao restante, foram gerados de forma artificial, seja ela de forma aleatória ou baseada em alguma distribuição real, conforme detalhado nas subseções seguintes. Nas subseções seguintes, encontra-se detalhado os dados que seguiram levantamentos ou distribuições reais e aleatórios.

#### 3.1.1 Criação da base *Events*

Como mencionado, no *dataset* fornecido por (Yin *et al.*, 2018), sobre os eventos, apenas sua identificação foi fornecida, sendo necessário preencher parte dos dados de forma artificial, como o endereço do evento e sua categoria, assumindo que estes ocorrem em uma dentre duas cidades, Xangai e Pequim. Para o organizador, foi selecionado aleatoriamente um dentre mil organizadores disponíveis. Os usuários na lista de participantes foram retirados da relação usuário-evento da base original, sendo assim, casos reais de pessoas que frequentaram aquele evento. Quanto à cidade, analisou-se a cidade de origem de cada participante; aquela que possui a maior quantidade de participantes tornou-se o local de acontecimento do evento. Para a localização do evento, utilizou-se uma escolha de coordenadas randômica dentre um limite estabelecido da cidade. Este limite é tido construindo uma circunferência ao redor da cidade cujo raio, escolhido aleatoriamente, possui uma extensão de 10 quilômetros a mais que o raio original da circunferência

<sup>1</sup> Disponível em: <https://1drv.ms/f/s!Ar5b2WXuinYWpi9CWTMHizheRIHy?e=5b2WZV>

Figura 3.1 – Esquema de arquivos da base de dados criada.



**Fonte:** Autoria própria

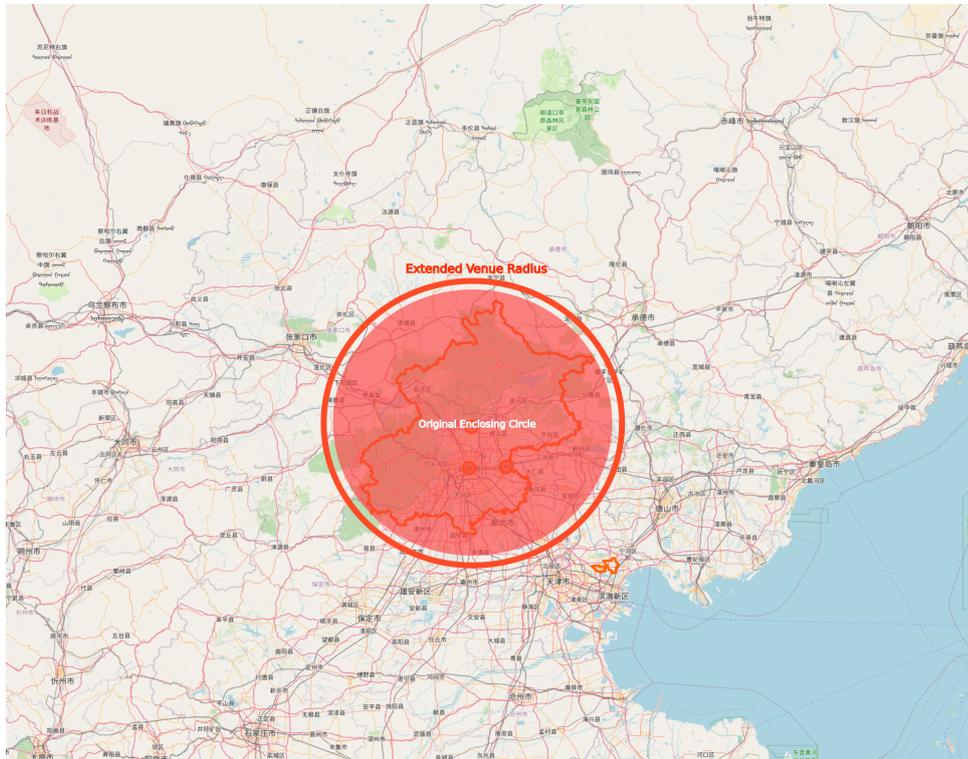
que abrange a cidade, como representado na Figura 3.2. As categorias para cada evento foram escolhidas de forma aleatória. Inicialmente, todas as 80 diferentes categorias são agrupadas em categorias mais abrangentes; dentre estas novas categorias, seleciona-se a mais popular entre todos os participantes. Assim, as categorias inicialmente agrupadas a esta categoria abrangente principal são selecionadas aleatoriamente para representarem o evento. Quanto à lista de personalidades frequentes, foram analisadas as personalidades MBTI de cada participante e as 5 mais frequentes foram mantidas em ordem decrescente. Por fim, ambas datas de início e fim foram escolhidas aleatoriamente dentre um período de 365 dias anteriores e posteriores à data deste trabalho, com um intervalo máximo de 7 dias entre ambas.

### 3.1.2 Criação da base *Users*

Assim como a base mencionada anteriormente, o incremento dos dados em falta da base original foram feitas artificialmente, de forma aleatória ou direcionada com base em dados de pesquisas prévias. Tendo como base que o *dataset* original foi coletado na plataforma *Douban Events*, algumas informações como a distribuição de personalidades MBTI seguiram a distribuição de ocorrências da República Popular da China. No caso da personalidade, utiliza-se a distribuição mostrada na Figura 3.3, retirada do levantamento feito pela *The Mayer-Briggs Company*<sup>2</sup>. Quanto à categoria, assim como no evento, foram escolhidas de forma aleatória uma lista de interesses para o usuário, dentre as 80 categorias existentes. Por fim, quanto à cidade do usuário, foi utilizada uma estratégia que visa simular dois tipos de usuários: usuários com conexões locais e usuários cuja localidade de suas conexões é mais extensa e menos importante. Para simplicidade, chamaremos de usuários tipo A e B, respectivamente, e a distribuição foi feita em duas etapas, descritas a seguir e apresentadas pelos Algoritmos 1 e 2:

<sup>2</sup> Disponível em: <https://hk.themyersbriggs.com/content/MBTIGlobalManualSuppCNS.pdf>

Figura 3.2 – Exemplo de circunferência para seleção randômica de localidade (apenas demonstrativo).



Fonte: Autoria própria

- **Etapa 1:** Seleção inicial

- Caso o usuário tenha mais de 60% de suas conexões em uma cidade A, sua probabilidade de ser do tipo A será equivalente à proporção de usuários da cidade A que fazem parte de sua conexão; caso contrário, terá probabilidade de 40% de ser do tipo A;
- Caso o tipo selecionado seja o tipo A, sua cidade será aquela com a maior frequência de usuários da lista de conexões;
- Caso o tipo selecionado seja o tipo B, sua cidade é aleatória;
- Caso o usuário não tenha ao menos 50% da lista de conexões com cidades definidas, é feita uma seleção randômica do seu tipo (A ou B) e da cidade de origem.

- **Etapa 2:** Propagação para usuários tipo A

- Caso o usuário já não pertence à cidade de maior frequência, a cidade é atualizada.

O Algoritmo 1 representa a etapa inicial de definição da cidade de origem de um usuário. Inicialmente, se o usuário não possui conexões, os atributos tipo e cidade são atribuídos aleatoriamente entre os valores possíveis (tipos 1 ou 2 e cidades Xangai ou Pequim). Para aqueles usuários com conexões, é contado o número de conexões em cada cidade. Se as conexões em Xangai representarem mais de 50% do total e superarem as de Pequim, uma probabilidade  $prob\_type\_1$

é calculada; se  $n_{shanghai}$  for maior que 60% das conexões totais, a probabilidade é tida como a porcentagem de conexões em Xangai com relação às conexões totais; caso contrário, é definida como 0.4. Um valor aleatório entre 0 e 1 é então comparado com com esta probabilidade calculada, para definir o tipo do usuário como 1 e a cidade como Xangai, ou o tipo como 2 com a cidade escolhida aleatoriamente. O mesmo processo é aplicado se as conexões em Pequim predominarem. Caso contrário, ambos os atributos são atribuídos de forma aleatória. O mesmo ocorre para a cidade de Pequim, caso as conexões em Xangai não representarem mais de 50% do total ou não superarem as de Pequim.

Figura 3.3 – Distribuição de MBTI na República Popular da China

Sensing		Intuition			
Thinking	Feeling	Thinking			
<b>ISTJ</b> <i>n</i> = 96 18.4%	<b>ISFJ</b> <i>n</i> = 21 4.0%	<b>INFJ</b> <i>n</i> = 8 1.8%	<b>INTJ</b> <i>n</i> = 23 4.4%	Judging	Introversión
<b>ISTP</b> <i>n</i> = 41 7.9%	<b>ISFP</b> <i>n</i> = 30 5.8%	<b>INFP</b> <i>n</i> = 17 3.3%	<b>INTP</b> <i>n</i> = 25 4.8%		
<b>ESTP</b> <i>n</i> = 45 8.6%	<b>ESFP</b> <i>n</i> = 15 2.9%	<b>ENFP</b> <i>n</i> = 17 3.3%	<b>ENTP</b> <i>n</i> = 21 4.0%	Judging	Extraversión
<b>ESTJ</b> <i>n</i> = 94 18.0%	<b>ESFJ</b> <i>n</i> = 23 4.4%	<b>ENFJ</b> <i>n</i> = 22 4.2%	<b>ENTJ</b> <i>n</i> = 23 4.4%		

Note: *N* = 521.

Fonte: The Myer-Briggs Company, 2018

Já no Algoritmo 2, é feito o ajuste das cidades origem de acordo com o tipo do usuário. Para cada usuário do tipo A, é calculado o número de conexões em cada cidade, Xangai e Pequim, e, baseado nesta quantificação, a cidade mais popular é definida. Se a cidade atual do usuário não for a cidade mais frequente calculada, a cidade do usuário é atualizada para a mais frequente.

### 3.1.3 Distribuição de Sentimentos e Emoções

Por fim, para selecionar os sentimentos e emoções, não foram consideradas nenhuma restrição ou padrão em específico, apenas uma escolha aleatória para cada usuário. Dentre os sentimentos, foram utilizados 3: “Positivo”, “Negativo” e “Neutro”; já para as emoções, foram utilizadas 5, sendo elas: “Felicidade”, “Excitação”, “Neutro”, “Tristeza” e “Decepção”.

**Algoritmo 1: Etapa 1**


---

```

1 Entrada: usuário  $u$ , conexões do usuário  $u$  ( $B_u$ )
2 Seleção_Inicial( $u, B_u$ )
3   se  $|B_u| = \emptyset$  então
4      $u\_type \leftarrow \mathbf{random}([1,2])$ 
5      $u.city \leftarrow \mathbf{random}([Shanghai, Beijing])$ 
6    $S_{B_u} \leftarrow \text{conexões em Xangai}, B_{J_{B_u}} \leftarrow \text{conexões em Pequim}$ 
7    $n\_shanghai \leftarrow |S_{B_u}|, n\_beijing \leftarrow |B_{J_{B_u}}|$ 
8   se  $n\_shanghai > n\_beijing$  e  $n\_shanghai > 0.5 \times |B_u|$  então
9     se  $n\_shanghai > 0.6 \times |B_u|$  então
10       $prob\_type\_1 \leftarrow n\_shanghai / |B_u|$ 
11     senão
12       $prob\_type\_1 \leftarrow 0.4$ 
13     se  $\mathbf{random}(0,1) < prob\_type\_1$  então
14        $u\_type \leftarrow 1$ 
15        $u.city \leftarrow Shanghai$ 
16     senão
17        $u\_type \leftarrow 2$ 
18        $u.city \leftarrow \mathbf{random}([Shanghai, Beijing])$ 
19   senão
20     se  $n\_beijing > n\_shanghai$  e  $n\_beijing > 0.5 \times |B_u|$  então
21       se  $n\_beijing > 0.6 \times |B_u|$  então
22          $prob\_type\_1 \leftarrow n\_beijing / |B_u|$ 
23       senão
24          $prob\_type\_1 \leftarrow 0.4$ 
25       se  $\mathbf{random}(0,1) < prob\_type\_1$  então
26          $u\_type \leftarrow 1$ 
27          $u.city \leftarrow 'Beijing'$ 
28       senão
29          $u\_type \leftarrow 2$ 
30          $u.city \leftarrow \mathbf{random}([Shanghai, Beijing])$ 
31     senão
32        $u\_type \leftarrow \mathbf{random}([1,2])$ 
33        $u.city \leftarrow \mathbf{random}([Shanghai, Beijing])$ 

```

---

## 3.2 Recomendador de Eventos

Nesta seção, é detalhado o recomendador proposto para realização das recomendações, que teve como base e inspiração os trabalhos de Li *et al.* (2020) e Macedo, Marinho e Santos (2015). Como mostrado na Figura 3.4, o recomendador é composto por 3 partes principais: o módulo de extração de *features*, abordado na Subseção 3.2.1, o módulo de treinamento e classificação com uso de algoritmos de aprendizado de máquina, detalhado na Subseção 3.2.2, e, por fim, o módulo de recomendação e *feedback*, onde ocorre a criação da lista a ser recomendada e a coleta

**Algoritmo 2: Etapa 2**


---

```

1 Entrada: Conjunto de usuários  $\mathcal{U}$ 
2 Seleção_Inicial( $\mathcal{U}$ )
3   para cada  $usr$  em  $\mathcal{U}$  faça
4     se  $usr\_type == 1$  então
5        $S_{B_{usr}} \leftarrow$  conexões em Xangai,  $BJ_{B_{usr}} \leftarrow$  conexões em Pequim
6        $n\_shanghai \leftarrow |S_{B_{usr}}|$ ,  $n\_beijing \leftarrow |BJ_{B_{usr}}|$ 
7       se  $n\_shanghai > n\_beijing$  então
8          $most\_frequent\_city \leftarrow$  Shanghai
9       senão
10         $most\_frequent\_city \leftarrow$  Beijing
11       se  $n\_shanghai == n\_beijing$  então
12          $most\_frequent\_city \leftarrow$   $usr.city$ 
13       se  $usr.city \neq most\_frequent\_city$  então
14          $usr.city \leftarrow$   $most\_frequent\_city$ 
15   retorne  $\mathcal{U}$ 

```

---

de *feedback* desta lista, detalhado na Subseção 3.2.3. Inicialmente, no “Módulo Extração de *Features*”, utiliza-se o perfil do usuário e a página do evento como *input* para extração das *features*; estas *features* são armazenadas em uma estrutura que é fornecida ao classificador, presente no “Módulo de Treinamento e Classificação”. Em tal módulo, um classificador é treinado utilizando dados de treinamento e baseado no *input* mencionado; além disso, uma lista de recomendação é gerada e fornecida ao usuário, no módulo “Recomendação e Feedback”. Por fim, é feita a coleta de *feedbacks* da lista criada.

Esta seção encontra-se organizada da seguinte forma: Subseção 3.2.1 detalha a extração das *features* utilizadas no recomendador; Subseção 3.2.2 detalha os modelos utilizados para construção do recomendador, assim como o treinamento do modelo; e, por fim, Subseção 3.2.3 detalha a geração da lista de eventos recomendados ao usuário. Durante toda a seção, será usada, caso necessário, a nomenclatura descrita na Tabela 3.1. Será utilizada também a seguinte nomenclatura no restante do trabalho:

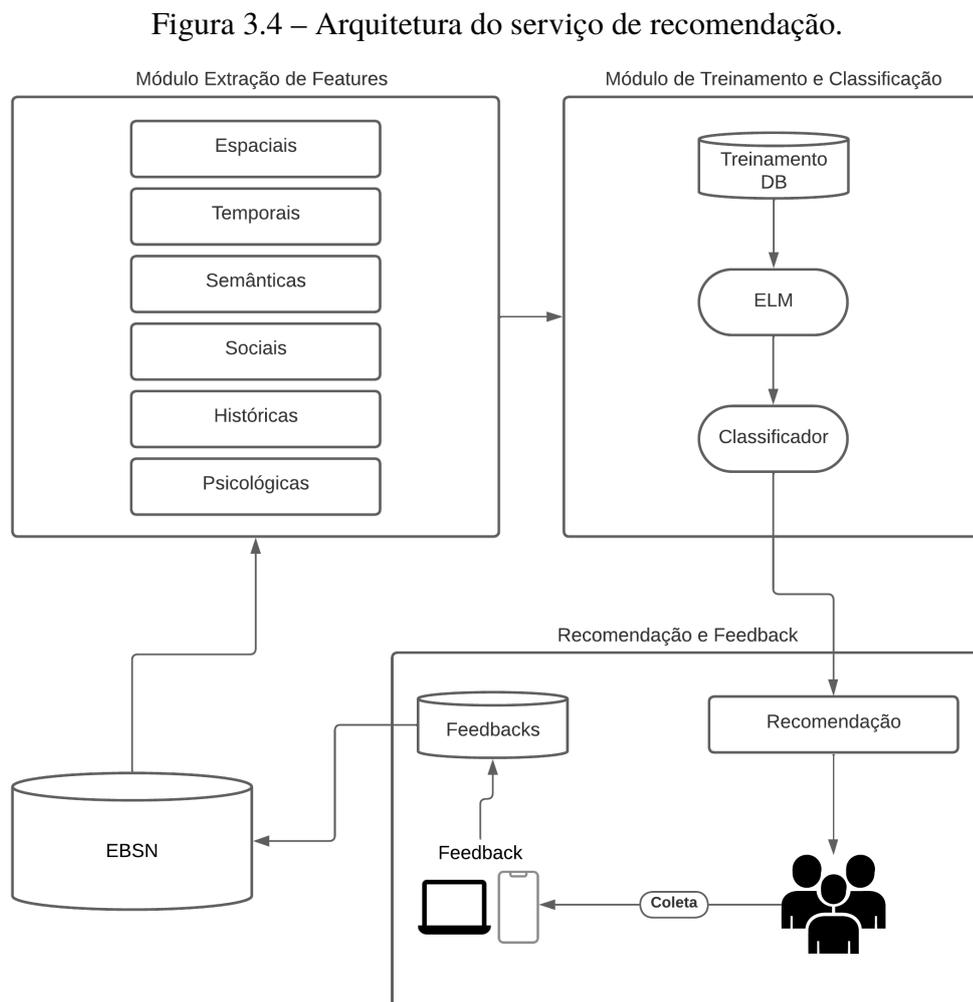
- **Serviço de recomendação:** refere-se a todo o processo representado na Figura 3.4, desde a extração de *features*, até a coleta de *feedbacks* da recomendação de eventos para os usuários;
- **Módulo:** refere-se individualmente a cada uma das três subdivisões da Figura 3.4;
- **Modelo:** refere-se à função utilizada para gerar a lista de recomendação. Em outras palavras, ao classificador inferido por um algoritmo ou técnica de aprendizado que irá definir de um evento deve ou não ser recomendado a um usuário.

### 3.2.1 Módulo de Seleção de *Features*

Como mostrado na Figura 3.4, 6 principais categorias de *features* são utilizadas para a realização da recomendação, sendo estas e encontrando-se detalhadas nas subseções seguintes: espaciais, temporais, semânticas, sociais, históricas e psicológicas. Vale já ressaltar que as *features* psicológicas são mais complexas de serem extraídas; no entanto, como o foco deste trabalho encontra-se no desenvolvimento de um serviço de recomendação e não na extração de personalidade ou emoção e sentimento, assume-se que o dado necessário para as extrações das *features* já foi coletado. Considera-se como dado necessário para tais extrações de *features* psicológicas o identificador MBTI de cada usuário.

#### 3.2.1.1 *Features* Espaciais

As *features* espaciais são aquelas que analisam hábitos comportamentais de um usuário quanto à participação em eventos, mostrando haver uma relação direta entre a participação do usuário e sua distância ao local do evento. Segundo Li *et al.* (2020), as seguintes ocorrências são habituais: (1) indivíduos tendem a participar de eventos dentro de uma certa distância do seu local



Fonte: Autoria própria

Tabela 3.1 – Tabela de símbolos usados na arquitetura do sistema de recomendação

Símbolo	Significado
$E$	Conjunto de todos os eventos.
$U$	Conjunto de todos os usuários.
$L$	Conjunto de todas as categorias no sistema.
$P$	Conjunto de todas as personalidades. $ P  = 16$ .
$A_u$	Conjunto de relação da participação do usuário $u \in U$ no evento $e \in E$ .
$B_u$	Relação de conexão entre usuário $u \in U$ e demais usuários $u' \in U$ .
$l_u$	Conjunto de categorias de interesse do usuário $u \in U$ e $l \in L$ .
$l_e$	Conjunto de categorias do evento $e \in E$ e $l \in L$ .
$P_e$	Conjunto de personalidades mais frequentes do evento $e \in E$ .
$P_u$	Personalidade e/ou traços da personalidade do usuário $u \in U$ .
$H_{he}$	Conjunto de eventos organizados pelo organizador $h$ .
$SFMD$	<i>Feature</i> Espacial Distância Média.
$SFUD$	<i>Feature</i> Espacial Distância do Usuário.
$SOF$	<i>Feature</i> Social.
$PSEF$	<i>Feature</i> de Similaridade Psicológica entre Usuário e Evento.
$PSEF$	<i>Feature</i> de Similaridade Psicológica entre Usuários.
$SEF$	<i>Feature</i> Emocional e Sentimental.

atual, devido ao gasto de tempo e dinheiro para deslocar de um local a outro; e (2) indivíduos tendem a frequentar eventos em certas áreas específicas, devido ao conforto de conhecimento da área e do que ela tem a oferecer. Além disso, como notado por [Macedo, Marinho e Santos \(2015\)](#), o padrão comportamental relacionado à localização indica também a possibilidade de usuários preferirem deslocar grandes distâncias. Isto pode ocorrer por vários motivos, como a busca de maior diversidade de eventos, privacidade e novas experiências.

Neste trabalho, propõe-se o uso de 2 subcategorias, distância média entre eventos e distância entre eventos e usuários. A primeira, em suma, representa a distância entre o local de ocorrência de cada evento, auxiliando na identificação de preferências de locais para participação

dos eventos, como o centro de uma cidade. A Equação 3.1 detalha o cálculo da primeira subcategoria para um dado usuário  $u$  e um novo evento  $e$ , que se resume à média aritmética da distância euclidiana entre dois eventos. A escolha pela distância euclidiana se dá pelo fato de indivíduos tenderem a participar de eventos em uma mesma cidade ou cidades próximas, e pelo fato de a distância euclidiana necessitar de poucas computações. A última, representada na Equação 3.2, busca um padrão na distância entre o usuário e os eventos participados. Para fins de clareza, a Tabela 3.1 resume os símbolos usados nesta seção e seus significados.

$$SFMD(u, e) = \frac{1}{|A_u|} \times \sum d(e, e_i), \forall e_i \in A_u \quad (3.1)$$

$$SFUD(u, e) = d(u, e) = \sqrt{(e_{long} - u_{long})^2 + (e_{lat} - u_{lat})^2} \quad (3.2)$$

### 3.2.1.2 Features Temporais

As *features* temporais buscam por padrões de preferências de um indivíduo quanto ao momento e duração de ocorrência do evento. Neste trabalho, dado um evento  $e$  participado pelo usuário, as *features* temporais são extraídas da seguinte forma:

1. Inicialmente, calcula-se o tempo de duração do evento em horas;
2. Divide-se o dia em 4 grupos de intervalos iguais de 6 horas (Alvorecer, Manhã, Tarde e Noite). Cada grupo recebe um valor binário que indica se o evento ocorre durante aquele grupo; por exemplo, supondo um evento que comece às 10h e termine às 18h, o resultado seria [0, 1, 1, 0];
3. Por meio uma lista binária de 7 itens, indica-se os dias da semana no qual o evento ocorre.

### 3.2.1.3 Feature Semântica

O aspecto semântico analisado busca padrão no tema e conteúdo do evento. Para tal, é considerada a semelhança entre as categorias às quais o evento  $e$  pertence e as categorias dos eventos frequentados por  $u$ . Por fim, as categorias do evento  $e$  são comparadas às categorias de interesse do perfil do usuário  $u$ . Todas as comparações são feitas seguindo a similaridade de Jaccard, representada na Equação 3.3, onde  $X$  e  $Y$  são conjuntos de dados numéricos que representam os dados a serem comparados quanto à similaridade.

$$jaccard\_sim(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (3.3)$$

### 3.2.1.4 Feature Social

As *features* sociais buscam a influência que o usuário sofre quanto à presença de certas pessoas no evento. Por exemplo, um evento, no qual amigos com um mesmo gosto estarão presentes, tem grande chance de atrair o usuário. Para isso, usa-se uma ideia similar à proposta

por Li *et al.* (2020), na qual é utilizada a similaridade do cosseno de interesses de usuários conectados. Neste caso, como é tida por verdade a possibilidade de usuários possuírem uma quantidade de categorias de interesse diferentes, utiliza-se a similaridade de Jaccard. A *feature* é representada pela Equação 3.4, sendo a média ponderada da similaridade entre o usuário  $u$  e sua lista de amigos que participarão do evento  $e$ . Vale ressaltar que, a lista de participação de amigos de um usuário é tida utilizando a relação de conexão entre usuários, existente na EBSN e no *dataset*.

$$SOF(u, e) = \frac{1}{|B_u|} \times \sum jaccard\_sim(u, u_i) \times jaccard\_sim(l_{u_i}, l_e), \forall u_i \in B_u \quad (3.4)$$

### 3.2.1.5 Features Psicológicas

Os traços psicológicos a serem analisados com as *features* psicológicas são traços inerentes à personalidade do usuário e ao efeito de eventos passados nos sentimentos e emoções do usuário. O objetivo da proposta da comparação de traços de personalidade está no fato de indivíduos frequentarem locais e eventos com pessoas com as quais ela se assemelha, não apenas quanto ao gosto mas também quanto às ações dos participantes e o estilo de conversa, por exemplo. Para isso, algo semelhante às *features* sociais é feito. Para um evento  $e$ , as principais personalidades participantes do evento têm os traços extraídos, criando uma lista binária de 4 posições. Cada posição indica o par de traços identificados<sup>3,4</sup> como é possível ver no exemplo da Tabela 3.2.

Tabela 3.2 – Amostra do resultado da extração de traços de personalidades.

Type	E/I	S/N	T/F	J/P
ENTP	1	0	1	0
INTP	0	0	1	0
INTJ	0	0	1	1
ENTJ	1	0	1	1

A Tabela 3.2, como mencionado, apresenta um exemplo de extração de traços da personalidade de um usuário. No caso do primeiro, “ENTP”, o valor “[1, 0, 1, 0]” indica que o usuário possui traços de extroversão, intuição, decisões lógicas e abertas a novas oportunidades.

Esta lista é então comparada quanto à similaridade média com os traços da personalidade do usuário. O mesmo é feito comparando a similaridade média entre o usuário e as conexões

<sup>3</sup> Neste trabalho foram utilizados os traços MBTI: Extroversão (E) ou Introversão (I), Sensação (S) ou Intuição (N), Pensamento (T) ou Sentimento (F), e Julgamento (J) ou Percepção (P).

<sup>4</sup> Saiba mais sobre o indicador MBTI em: <https://www.themyersbriggs.com/en-US/Products-and-Services/Myers-Briggs>

frequentando o evento  $e$ , quanto aos traços de personalidade. As Equações 3.5 e 3.6 detalham, respectivamente, o cálculo das médias aritmética e ponderada. A notação da Equação 3.7 segue a nomenclatura proposta na Equação 3.3.

$$PSEF(u, e) = \frac{1}{|P_e|} \times \sum \cos\_sim(P_u, P_{e_i}), \forall P_{e_i} \in P_e \quad (3.5)$$

$$PSUF(u, e) = \frac{1}{|B_u|} \times \sum \cos\_sim(P_u, P_{u_i}) \times PSEF(u_i, e), \forall u_i \in B_u \quad (3.6)$$

$$\cos\_sim(X, Y) = \frac{X \cdot Y}{\|X\| \|Y\|} \quad (3.7)$$

A proposta do uso de sentimentos e emoções está na direta relação entre a participação do usuário e como ele se sente no evento. Supondo que o usuário  $u_a$  sempre demonstre felicidade na participação de um evento organizado por um usuário  $u_b$ , existe uma grande chance do usuário se interessar por eventos futuros organizados por  $u_b$ . Para a consideração desta *feature*, usa-se a Equação 3.10. Considerando um usuário  $u$  e um novo evento  $e$ , as *features* são as médias ponderadas tanto do sentimento quanto da emoção do usuário durante a participação dos eventos organizados pelo mesmo organizador de  $e$ . Para sentimentos, considera-se 1 como sendo o valor representante da categoria positiva, 0 neutra e  $-1$  negativa. Para emoção, “felicidade” possui valor 2, “surpresa” possui valor 1 e o restante, por serem emoções negativas, possuem valor  $-1$  e  $-2$ . O cálculo da *feature* é o mesmo para sentimento e emoção, sendo assim, o termo  $X_{ue_i}$  na Equação 3.10 assume um dos valores presentes nas Equações 3.8 e 3.9

$$S_{ue} = [-1, 0, 1], \text{ Valor do sentimento do usuário } u \text{ no evento } e \quad (3.8)$$

$$E_{ue} = [-2, -1, 1, 2], \text{ Valor da principal emoção do usuário } u \text{ no evento } e \quad (3.9)$$

$$SEF(u, e) = \frac{1}{|A_u \cap H_{he}|} \times \sum X_{ue_i} \times jaccard\_sim(u, e_i), \forall e_i \in H_{he} \cap A_u, H_{he} \cap A_u \neq \emptyset \quad (3.10)$$

### 3.2.2 Módulo de Treinamento e Classificação

Com a extração das *feautres* completa, é necessário realizar o treinamento do modelo recomendador de eventos. Para isso, o *dataset* é dividido em duas bases, treinamento e validação, seguindo a distribuição  $80\% \times 20\%$ , respectivamente. Ambas divisões possuem uma lista de eventos participados e não participados para cada usuário, onde a lista de eventos participados já estava presente na base original mencionada na Seção 3.1 e a lista de eventos não participados foi criada selecionando uma quantia arbitrária de eventos presentes na base original, mas não presentes na lista de participação de um usuário. Toda a base de treinamento é então fornecida às técnicas de aprendizado propostas, que irá ajustar, internamente, parâmetros utilizados durante

a predição do resultado final de forma a minimizar o erro entre a predição e o valor real que deveria ser retornado. A escolha de um algoritmo de aprendizado para realização do serviço de recomendação é uma escolha complexa, visto que diversos fatores como tempo de treinamento, complexidade do algoritmo, limitações, ambiente no qual será usado, entre outros, devem ser considerados. Li *et al.* (2020) propõem o uso do algoritmo de aprendizado de máquina conhecido como *Extreme Learning Machine (ELM)*, um algoritmo da classe de redes neurais proposto inicialmente por Huang, Zhu e Siew (2004) em 2004. O algoritmo, conhecido pela eficiência e velocidade de treinamento, é composto por 3 principais camadas: a primeira é a camada de *input*, composta por  $n$ -neurônios; a segunda é camada oculta (*hidden layer*) com  $m$ -neurônios ( $m$  é um hiperparâmetro); e a terceira é uma camada de *output* de  $k$ -neurônio.

Este trabalho propõe o uso de uma variação do ELM, além da versão padrão, conhecida como *Pruning ELM (P-ELM)*, assim como a utilização da técnica de aprendizado de máquina *XGBoost*, proposta por Chen e Guestrin (2016). O P-ELM é um algoritmo que remove neurônios ao fim do treinamento, daí o nome “*pruning*”; isso ocorre visando diminuir a complexidade do modelo e aumentar o desempenho, diminuindo o número de computações necessárias e generalizando o modelo.

Apesar de alguns trabalhos analisarem não apenas o interesse do usuário por um evento, mas a probabilidade de um usuário participar de um evento (como proposto por Lan *et al.* (2022)), este trabalho propõe apenas a classificação binária de interesse. Isto é feito em busca de simplificar o entendimento do resultado do classificador. Assim sendo, o treinamento ocorre da seguinte forma: para cada usuário  $u_i \in U$  e evento  $e_j \in A_{u_i}$ , as *features* extraídas são transformadas em um vetor de entrada dos algoritmos de aprendizado propostos e o resultado de cada um sugere se o usuário possui interesse, ou não, no evento  $e_j$  em específico. Vale ressaltar que o vetor de *features* é criado para cada usuário e evento.

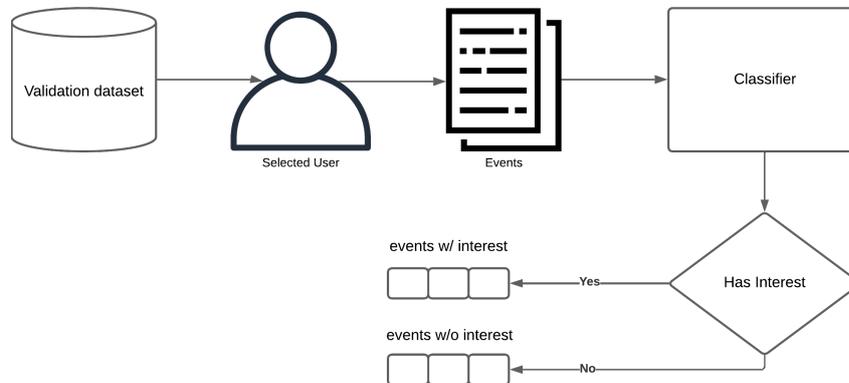
### 3.2.3 Recomendação e *Feedback*

O último dos módulos abordados na arquitetura do serviço de recomendação proposto na Figura 3.4 é o de geração da lista de recomendação e coleção do *feedback*. Como já mencionado, o trabalho não pode ser aplicado em um aplicativo móvel para coleta de *feedbacks* implícitos, como tempo de tela em um evento; sendo assim, uma outra forma de *feedback* foi utilizada: o interesse explícito em um evento, por parte de um usuário. O esquema da Figura 3.5 representa o processo de geração da lista de recomendação, assim como da coleta do *Feedback*.

Inicialmente, para cada evento da base de validação, calcula-se a probabilidade de o usuário ter interesse no evento, e este é considerado como interessante se for superior a um limiar definido. Por fim, aqueles então julgados interessantes para o usuário são adicionados a uma lista que é recomendada ao usuário. No trabalho, não foi aplicado nenhum sistema de ranqueamento à lista recomendada, sendo assim, todos aqueles então julgados interessantes, são recomendados ao usuário. Na base de validação, para cada usuário, existe um campo que define se o usuário tem

ou não interesse no evento; assim, realiza-se a comparação entre a lista de eventos recomendados ao usuário e a lista dos eventos que possuem o campo anteriormente mencionado com o valor definido para interesse; assim, o *feedback* é coletado.

Figura 3.5 – Esquema de recomendação e coleta de Feedback.



Fonte: Autoria própria

### 3.3 Módulo de Integração de Dados

Como uma alternativa à utilização da base de dados criada anteriormente, um *script* para integração de dados públicos provenientes de outras plataformas foi desenvolvida. Esta ferramenta realiza a extração de “nome”, “descrição”, “categorias”, “data” e “URL” de cada evento e, armazena-os em arquivos JSON, por meio de um *web scraper*<sup>5</sup> desenvolvido em **Python** utilizando as bibliotecas *Selenium*<sup>6</sup> e *BeautifulSoup*<sup>7</sup>. Vale ressaltar que, para o desenvolvimento, teste e experimentos neste trabalho, este módulo não foi utilizado, sendo usado apenas o detalhado na Seção 3.1. Este módulo foi abordado por se mostrar útil para futuros trabalhos. O *script* desenvolvido realiza a extração de eventos da plataforma Meetup nas seguintes cidades:

- Nova Iorque, NY, EUA;
- Los Angeles, CA, EUA;
- Paris, Île-de-France, FR;
- São Paulo, SP, BR;
- Londres, RU;
- Barcelona, ES.

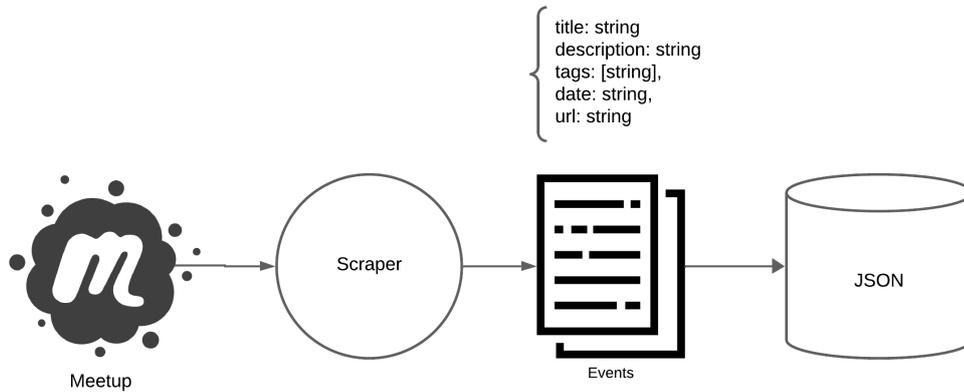
<sup>5</sup> *Web Scraping* é uma técnica usada para extração de dados não estruturados de forma automática em um *website* e armazenamento em forma estruturada (Zhao, 2017).

<sup>6</sup> Disponível em: <https://www.selenium.dev/>

<sup>7</sup> Disponível em: <https://beautiful-soup-4.readthedocs.io/en/latest/>

Para facilitar o entendimento da arquitetura proposta para a organização dos dados, a Figura 3.6 esquematiza em forma visual a coleção dos dados. Vale ressaltar que, foi escolhido o armazenamento em arquivos em JSON para facilitar a disponibilidade do *script*, contendo os dados dos eventos<sup>8</sup>.

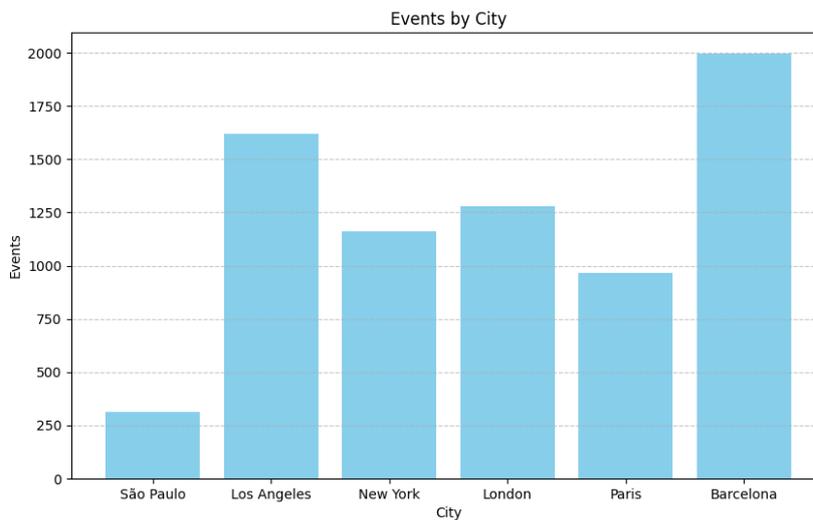
Figura 3.6 – Esquema da coleção de dados provenientes de outras plataformas.



**Fonte:** Autoria própria

No total, o *script* fez uma coleta de 7331 eventos, onde a distribuição de eventos por cidade é mostrado no Gráfico 3.7

Figura 3.7 – Distribuição de cidade por eventos.



**Fonte:** Autoria própria

<sup>8</sup> Disponível em: <https://1drv.ms/f/s!Ar5b2WXuinYWpi6CkLAVXuKbKA-X?e=c7roh>

## 4 Experimentos e Resultados

Este capítulo tem, como foco, a apresentação dos experimentos realizados nos módulos “Treinamento e Classificação” e “Recomendação e *Feedback*”, apresentados anteriormente no Capítulo 3, assim como uma análise de resultados experimentais obtidos. O intuito destes experimentos é a validação dos módulos e dos modelos aplicados nestes, assim como possibilitar a análise da eficácia do serviço de recomendação proposto e a determinação de ser ou não um caminho viável para novos trabalhos. Para tanto, encontra-se organizado da seguinte forma: Seção 4.1 detalha as métricas utilizadas para avaliação dos modelos classificadores detalhados no capítulo anterior, e conseqüentemente para avaliação do serviço de recomendação; Seção 4.2 detalha o processo de ajuste dos hiper-parâmetros dos dois modelos ELM mencionados; Seção 4.3, assim como a Seção 4.2, detalha os ajustes realizados nos hiper-parâmetros do modelo *XGBoost*; por fim, a Seção 4.4 detalha e analisa os resultados experimentais obtidos. Os ajustes de hiper-parâmetros foram realizados utilizando a abordagem sistemática *Grid Search*, implementada na biblioteca “GridSearchCV”, com foco na acurácia <sup>1</sup>.

Todo o trabalho foi implementado utilizando a linguagem **Python**<sup>2</sup> em sua versão **3.11.6**. Os experimentos foram realizados utilizando a base de dados desenvolvida e fornecida na Seção 3.1. Quanto à semente de aleatoriedade, foi utilizada a mesma em todas as execuções, com o valor 42. A execução dos testes foi realizada em um ambiente que segue a seguinte configuração de máquina: processador Intel i7-13700T 4.90Ghz e 32GB de memória RAM. No total, de todos os 66.013 usuários disponíveis, por questões de limitações temporais e de uso da máquina, foram avaliados e utilizados 100 usuários, com o maior número de eventos participados.

### 4.1 Avaliação dos modelos de recomendação

Para a avaliação dos modelos e *features*, foram utilizadas as seguintes métricas tendo como base as utilizadas nos trabalhos relacionados:

- **Acurácia:** representa a proporção de classificações corretas sobre todas as instâncias, sendo calculada de acordo com a Equação 4.1;

$$Acuracia = \frac{| \text{classificações corretas} |}{| \text{todas as classificações} |} \quad (4.1)$$

- **Precisão Média:** representa a acurácia das classificações verdadeiramente positivas (TP), em relação às classificações verdadeiramente positivas somadas às falsamente positivas

<sup>1</sup> *GridSearchCV* é uma biblioteca que implementa uma busca exaustiva sobre valores de parâmetros definidos para um estimador (Scikit-learn, 2024)

<sup>2</sup> Disponível em: <https://www.python.org/>

(FP), sendo calculada de acordo com a Equação 4.2;

$$Precisao = \frac{1}{N} \sum_{i=1}^N \frac{|TP_i|}{|TP_i + FP_i|} \quad (4.2)$$

- **Revocação (Recall):** representa a proporção de classificações TP em relação à soma das classificações TP e falsas negativas (FN), sendo calculada de acordo com a Equação 4.3. Em outras palavras, valores altos de revocação indicam, no contexto deste trabalho, que o modelo teve um alto índice de acerto ao identificar eventos que devem ser recomendados ao usuário;

$$Recall = \frac{|TP|}{|TP + FN|} \quad (4.3)$$

- **F1-Score:** representa a união de precisão e revocação. A união ocorre por meio da média harmônica entre precisão e revocação, como apresentado na Equação 4.4;

$$F1 = 2 \times \frac{Precisao \times Recall}{Precisao + Recall} \quad (4.4)$$

As métricas acima serão comparadas entre os 3 diferentes modelos de classificação: ELM<sup>3</sup>, Pruning ELM, XGBoost. Como não houve a possibilidade da aplicação do modelo em um ambiente real, não foi possível a coleta de *feedbacks* implícitos, como a abertura do evento e o tempo gasto na tela do evento.

## 4.2 Ajustes dos modelos baseado no ELM

Os ajustes de hiper-parâmetros nos modelos *ELM* e *Pruning-ELM* são semelhantes, já que o último é baseado no primeiro. O ELM depende de três hiper-parâmetros: Número de Neurônios Ocultos (*HN*), Taxa de Regularização L2 Alpha ( $\alpha$ ) e Limiar de Classificação (*THR*). O *Pruning*, necessita de ajustar, além dos três parâmetros iniciais, o seguinte: Limiar de Poda (*PTHR*). Vale ressaltar que, a versão implementada do *Pruning-ELM* foi uma versão básica, onde a poda ocorre apenas baseada no limiar de importância do neurônio.

A Tabela 4.1 detalha os valores utilizados e comparados durante a fase de ajuste do modelo.

## 4.3 Ajustes do modelo XGBoost

O modelo XGBoost, ao contrário do ELM, é mais complexo, e necessita de mais ajustes; no trabalho, os hiper-parâmetros ajustados foram: Taxa de Aprendizado ( $LR / \eta$ ), Número de

<sup>3</sup> Vale ressaltar que, para assegurar a integridade do modelo *ELM* e suas variação *Pruning-ELM*, por serem implementados em sua totalidade, a base *Breast Cancer Wisconsin*<sup>4</sup> foi utilizada para análise de performance.

Tabela 4.1 – Tabela de valores utilizados para ajuste fino do *ELM* e *Pruning-ELM*

	<b>ELM</b>	<b>Pruning-ELM</b>
<b>HN</b>	10, 19, 38, 76, 152	10, 19, 38, 76, 152
$\alpha$	0, 1e-3, 1e-2, 5e-2, 1e-1	0, 1e-3, 1e-2, 5e-2, 1e-1
<b>THR</b>	0.3, 0.5, 0.7	0.3, 0.5, 0.7
<b>PTHR</b>	$\times$	1e-3, 1e-2, 5e-2, 1e-1

Rodadas de Reforço (*NBR*), Profundidade Máxima (*MD*), Subamostra (*SS*), Colsample\_bytree (*CBT*), Taxa de Regularização L2 ( $\lambda$ ) e Gamma ( $\gamma$ ).

A Tabela 4.2 mostra os valores utilizados e comparados durante a fase de ajuste do modelo.

Tabela 4.2 – Tabela de valores utilizados para ajuste fino do *XGBoost*

	<b>XGBoost</b>
<b>LR / <math>\eta</math></b>	0.005, 0.01, 0.05
<b>NBR</b>	9, 19, 38
<b>MD</b>	2, 3, 4
<b>SS</b>	0.4, 0.5, 0.6, 0.7
<b>CBT</b>	0.6, 0.7, 0.8
$\lambda$	2, 5, 10, 20
$\gamma$	3, 4

## 4.4 Resultados

Nesta seção são apresentados os resultados alcançados no trabalho quanto à qualidade da criação de listas de recomendação de eventos para diferentes usuários. Ao final, é feita uma análise dos resultados levando em consideração comparações entre os diferentes modelos classificadores. Os resultados apresentados foram obtidos a partir de experimentos que utilizaram apenas os parâmetros após a execução dos ajustes detalhados nas Seções 4.2 e 4.3 do Capítulo 4.

Para facilitar a comparação entre os modelos, assim como a interpretação realizada no Capítulo 5, esta seção foi organizada da forma seguinte: Subseção 4.4.1 apresenta os valores adquiridos para cada métrica mencionada, para cada um dos 100 usuários, assim como o valor médio das métricas para o algoritmo *ELM*; e as Subseções 4.4.2 e 4.4.3 apresentam o mesmo para os algoritmos *Pruning-ELM* e *XGBoost*, respectivamente. A Subseção 4.4.4 apresenta a performance média de cada uma das técnicas de aprendizado mencionadas para cada métrica e

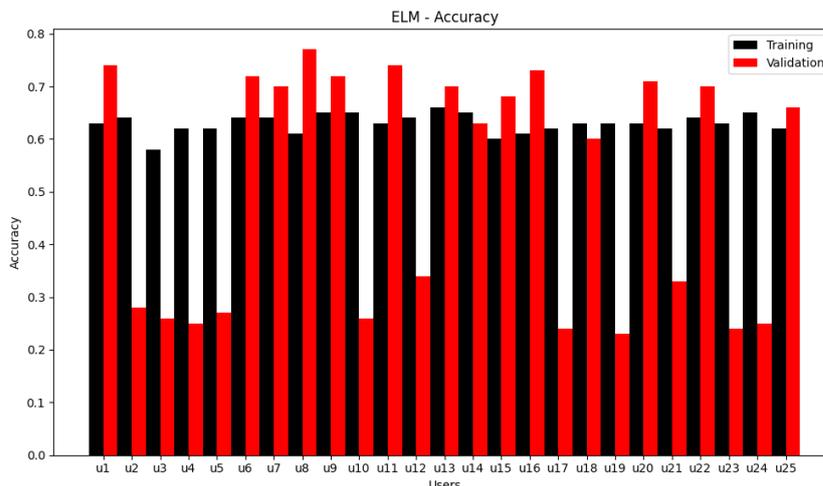
com diferentes *features*. Todos os modelos foram inferidos a partir do treinamento nos melhores parâmetros encontrados nas subseções 4.2 e 4.3, para cada usuário.

Dado que os algoritmos foram executados com 100 usuários, foram gerados gráficos de cada um para cada usuário e métricas; no entanto, para simplificar a apresentação de resultados, são apresentados um gráfico por métrica e algoritmo para os primeiros 25 usuários; após, é apresentado um gráfico com a média dos valores. Nesta seção, a cor preta indica métricas durante treinamento e vermelhas indicam validação do modelo inferido.

#### 4.4.1 Resultados ELM

As Figuras 4.1, 4.2, 4.3 e 4.4 apresentam, separadamente, os valores das 4 métricas em treinamento e validação para os primeiros 25 usuários para o *ELM*.

Figura 4.1 – Gráfico de acurácia por usuário.



Fonte: Autoria própria

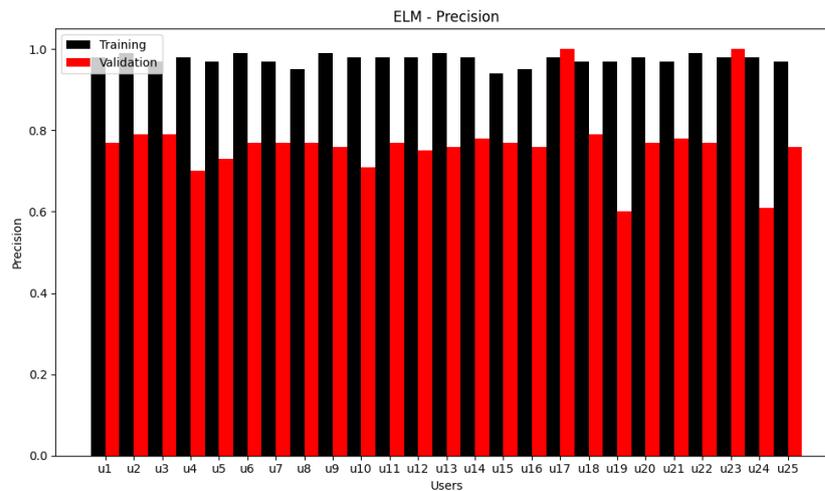
De acordo com os gráficos das Figuras 4.1 e 4.4, observa-se que o ELM tem uma performance boa para alguns usuários; no entanto, para uma grande parcela dos usuários, mostrou-se sinais de sobreajuste (*overfitting*), mesmo com hiper-parâmetros considerados baixos.

#### 4.4.2 Resultados Pruning-ELM

As Figuras 4.5, 4.6, 4.7 e 4.8 apresentam, separadamente, os valores das 4 métricas em treinamento e validação para os primeiros 25 usuários para o *Pruning-ELM*.

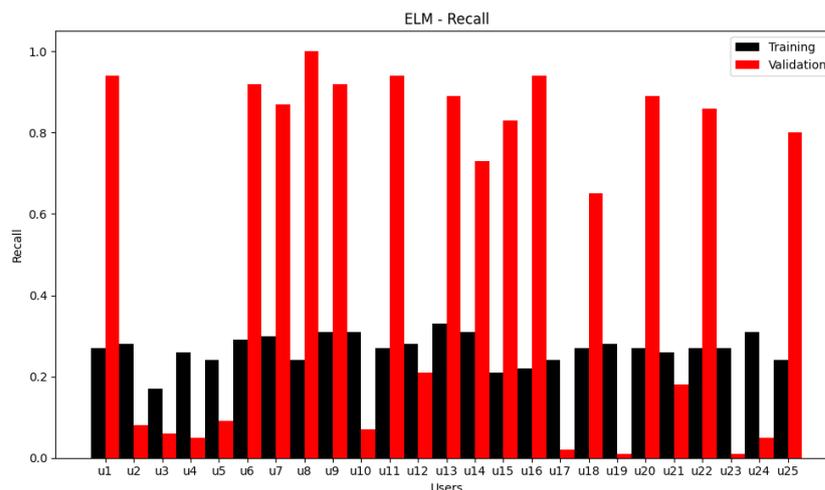
Pelas Figuras 4.5 e 4.8, nota-se uma similaridade com o *ELM*, anterior, o que é previsível, dada a similaridade entre as técnicas; no entanto, de um modo geral, o *Pruning-ELM* tem uma performance melhor que o *ELM*. Isto indica que, como esperado, a generalização do modelo removendo neurônios ajuda na tendência de *overfitting* do algoritmo tradicional ao inferir o modelo classificador.

Figura 4.2 – Gráfico de precisão por usuário.



Fonte: Autoria própria

Figura 4.3 – Gráfico de recall por usuário.



Fonte: Autoria própria

### 4.4.3 Resultados XGBoost

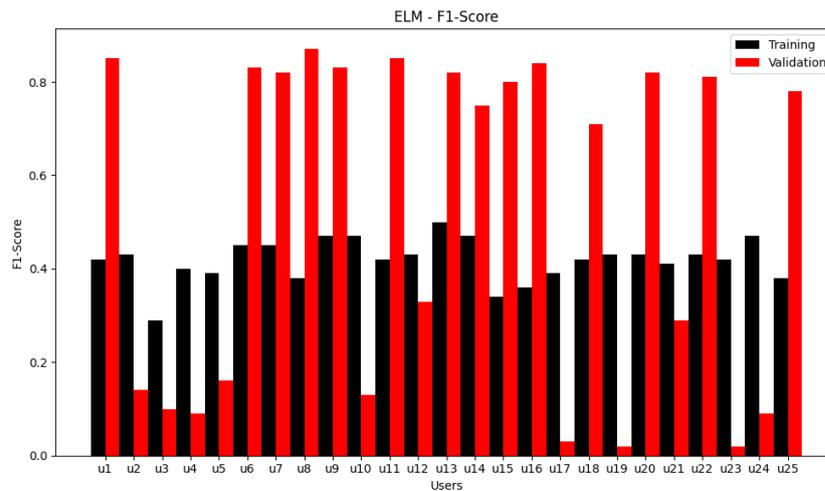
As Figuras 4.9, 4.10, 4.11 e 4.12 apresentam, separadamente, os valores das 4 métricas em treinamento e validação para os primeiros 25 usuários para o *XGBoost*.

Como observa-se nas Figuras 4.9, 4.10, 4.11 e 4.12, o *XGBoost* apresenta os melhores valores para as métricas analisadas, indicando ser o algoritmo que melhor se adapta ao problema de classificação de interesse. No entanto, assim como os dois anteriores, mostrou-se uma tendência de sobreajuste aos ruídos da base de treinamento.

### 4.4.4 Resultados Médios

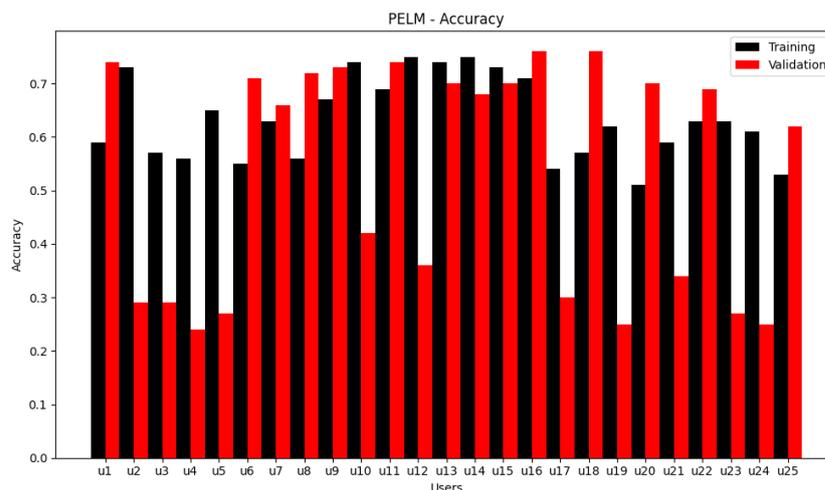
As Figuras 4.13, 4.14, 4.15 e 4.16 apresentam, separadamente, os valores médios das 4 métricas em treinamento e validação para todos os 100 usuários utilizando todas as *features*

Figura 4.4 – Gráfico de F1 por usuário.



Fonte: Autoria própria

Figura 4.5 – Gráfico de acurácia por usuário.



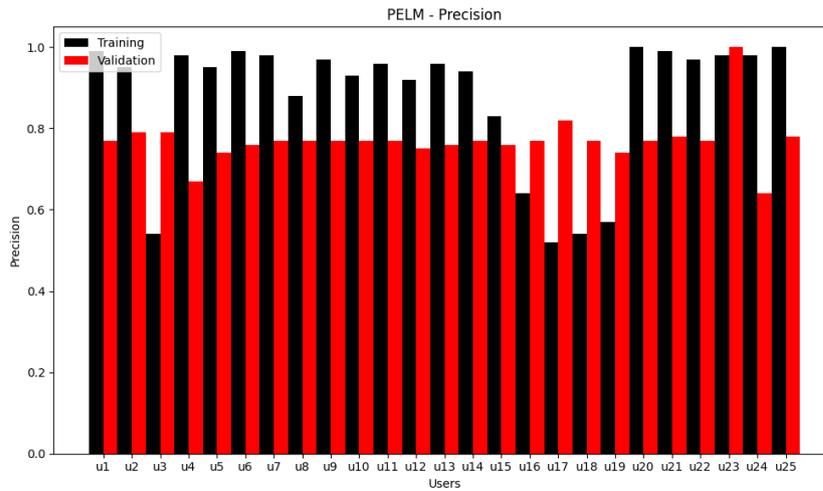
Fonte: Autoria própria

detalhadas no Capítulo 3.

Ao comparar os modelos inferidos quanto à média de suas métricas, observa-se os seguintes pontos:

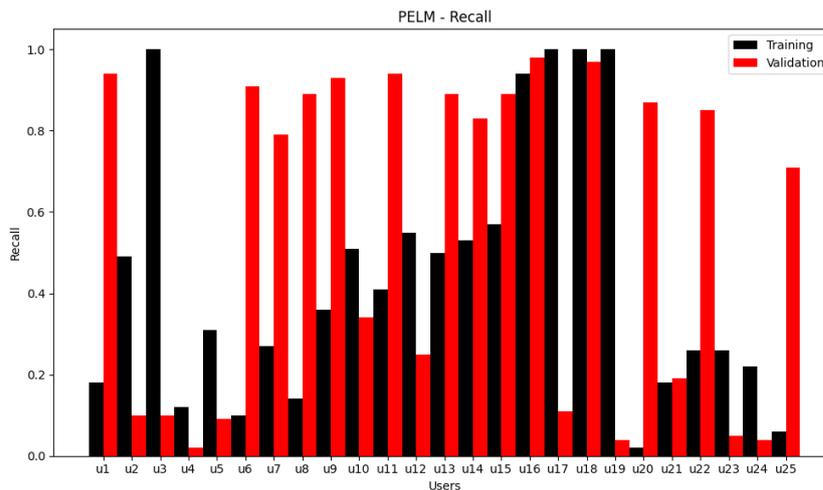
- Como mencionado e observado nas Figuras 4.13, 4.14, 4.15 e 4.16, o XGBoost apresenta uma melhor performance geral quando comparado aos outros dois modelos inferidos utilizando *ELM* e Pruning-ELM, isso é, utilizando todas as features propostas. No entanto, ao remover as feautres que consideram a emoção e sentimento do usuário, o modelo é incapaz de performar bem, ao contrário do *ELM* e Pruning-ELM, que melhoram a performance.
- A feature *PSEF*, afeta em maior escala a performance do algoritmo Pruning-ELM. O

Figura 4.6 – Gráfico de precisão por usuário.



Fonte: Autoria própria

Figura 4.7 – Gráfico de Recall por usuário.

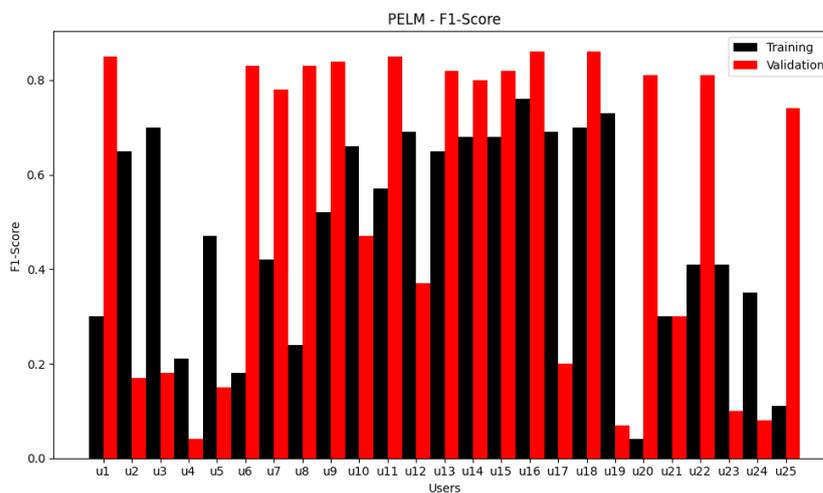


Fonte: Autoria própria

mesmo não ocorre no XGBoost; isto provavelmente está relacionado com a robustez e complexidade do modelo gerado pelo XGBoost, que em alguns casos é menos sensível à adição e remoção de features.

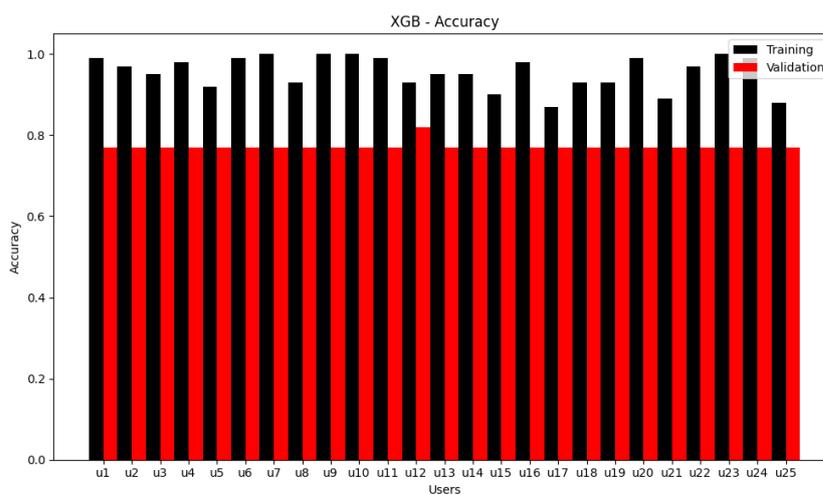
- A comparação entre o modelo do ELM e seu derivado indica que, apesar de serem semelhantes, a derivação apresenta melhor performance de uma forma geral, independente das features adicionadas ou removidas. Indica-se, portanto, que uma generalização do ELM ao remover neurônios, ajuda na classificação final para o problema.

Figura 4.8 – Gráfico de F1 por usuário.



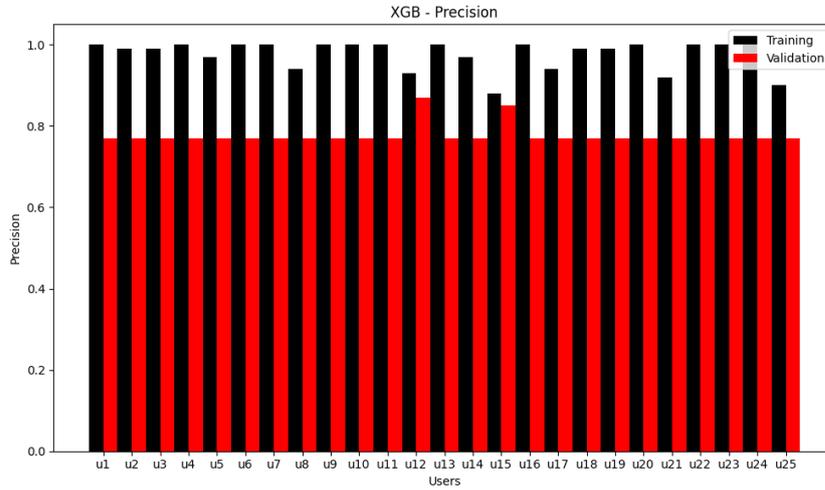
Fonte: Autoria própria

Figura 4.9 – Gráfico de acurácia por usuário.



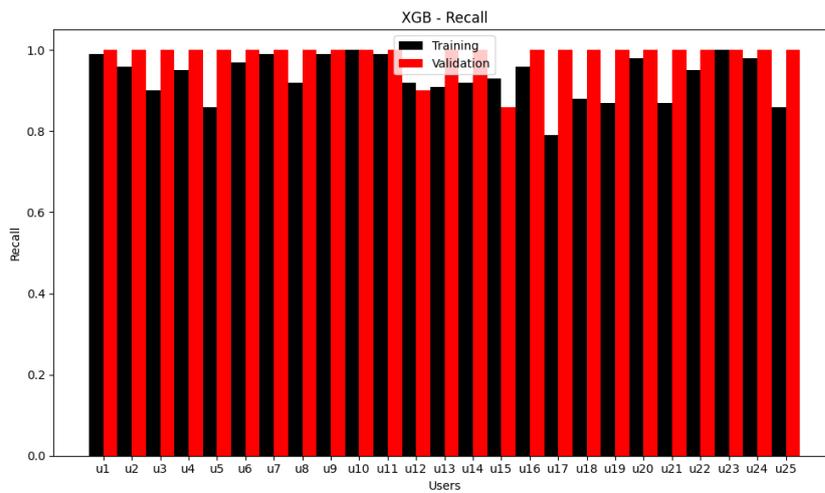
Fonte: Autoria própria

Figura 4.10 – Gráfico de precisão por usuário.



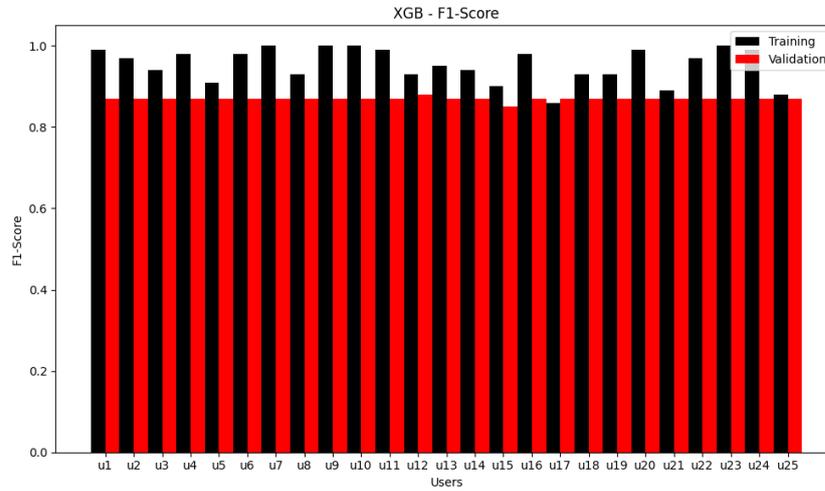
Fonte: Autoria própria

Figura 4.11 – Gráfico de Recall por usuário.



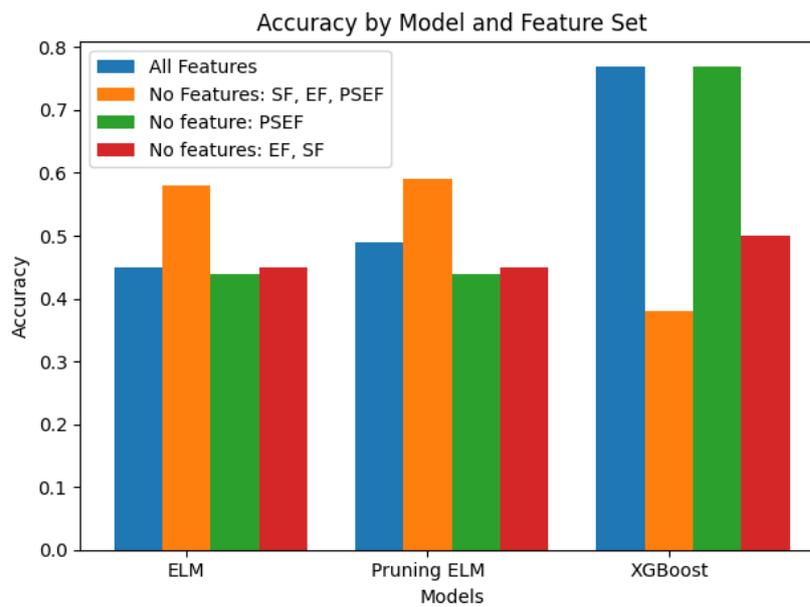
Fonte: Autoria própria

Figura 4.12 – Gráfico de F1 por usuário.



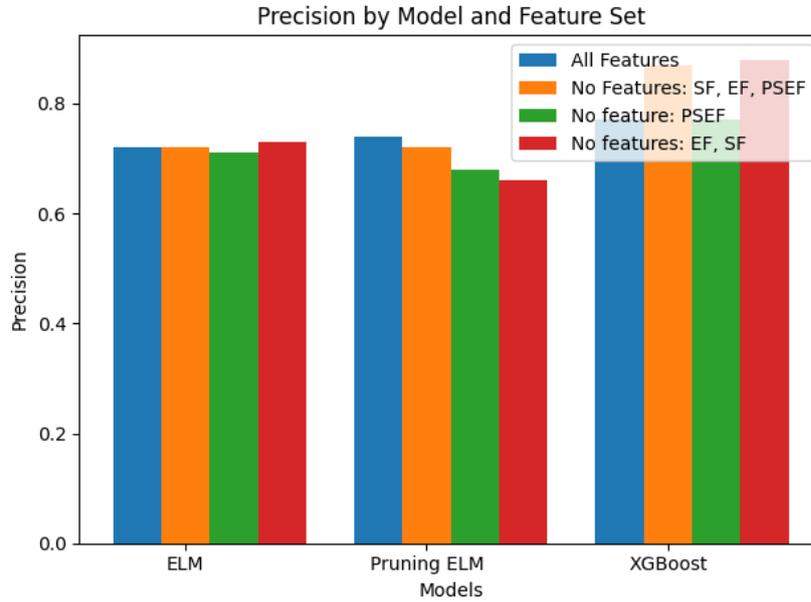
Fonte: Autoria própria

Figura 4.13 – Gráfico de acurácia por usuário.



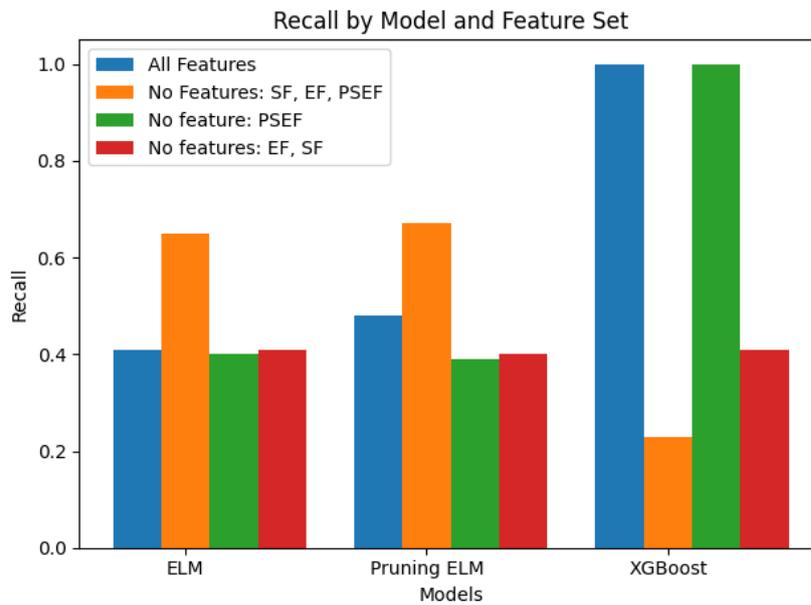
Fonte: Autoria própria

Figura 4.14 – Gráfico de precisão média.



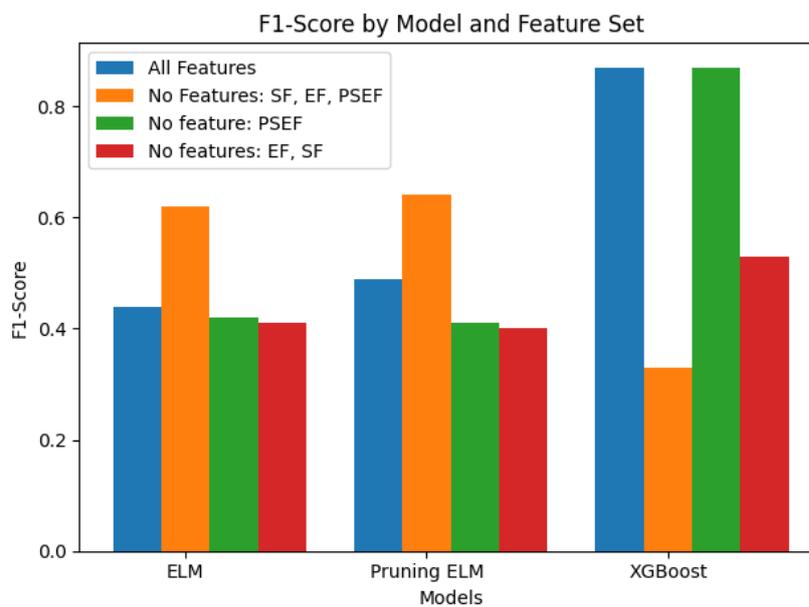
Fonte: Autoria própria

Figura 4.15 – Gráfico de Recall médio.



Fonte: Autoria própria

Figura 4.16 – Gráfico de F1 médio.



Fonte: Autoria própria

# 5 Considerações Finais

Neste capítulo, é são resumidas as conclusões obtidas após a análise dos resultados experimentais obtidos e apresentados no Capítulo 4.4, uma análise dos objetivos inicialmente propostos e atingidos, as contribuições do presente trabalho, e as perspectivas de trabalhos futuros tendo como base os desafios e pontos a melhorar. Desta forma, segue detalhado: Seção 5.1 realiza uma análise dos resultados, pontos positivos e negativos do trabalho, assim como sua contribuição e análise dos objetivos atingidos; Seção 5.2 apresenta perspectivas de trabalhos a serem realizados com base nas faltas apontadas, desafios encontrados e novas hipóteses criadas.

## 5.1 Conclusão

Foi proposta, neste trabalho, a criação de um serviço de recomendação de eventos para Redes Sociais Baseadas em Eventos. Este serviço teve por hipótese, que a possibilidade de análise de características referentes a traços psicológicos, sentimentais e emocionais são benéficas para a definição de interesse por um evento, por parte de um usuário. Além do objetivo geral deste trabalho, criação de um serviço de recomendação de eventos, este teve também como objetivos específicos: a criação de uma base de dados sintética que fosse disponibilizada publicamente para futuros trabalhos, a proposta de um modelo para previsão de interesse em eventos por parte de um usuário, e a proposta de módulo de integração de dados provenientes de outras plataformas. O Capítulo 3 detalhou o processo utilizado para o alcance de cada um dos objetivos, geral e específicos. Ao fim, uma nova base de dados sintética foi criada, assim como os modelos de previsão de interesse e do módulo integrador de dados foram desenvolvidos.

Pelas observações feitas no Capítulo 4.4, pode-se concluir que, entre os técnicas de aprendizado analisadas, o *XGBoost* é um forte candidato quanto a seleção de algoritmos para inferência dos modelos classificadores a integrem o serviço de recomendação. Isso porque as métricas apresentadas e analisadas são satisfatórias, mesmo apesar da visível tendência ao sobreajuste. Observa-se também, pelas Figuras 4.13, 4.14, 4.15 e 4.16 que:

- A *feature PSEF*, isoladamente, não apresenta grande importância para modelos robustos e complexos como o *XGBoost*; no entanto, ao ser utilizada com as *features SF* e *EF*, apresenta uma grande melhoria na performance geral;
- O conjunto de *features SF* e *EF* apresenta um grande impacto na performance de todos os modelos classificadores. Isto indica ser uma boa direção para melhoria de *features* em trabalhos futuros.

Em primeira análise, as observações apresentadas no Capítulo 4.4 podem levar à seguinte hipótese, que necessita de um estudo mais elaborado:

1. Dado que a base de dados utilizada tem diversos fatores parcialmente randômicos, existe a possibilidade de padrões a serem reconhecidos pelos modelos estarem ofuscados pelo caráter aleatório da base (ruídos). Sendo esta hipótese verdade, implicaria que o resultado obtido neste presente trabalho não necessariamente apresenta a real capacidade dos modelos e das *features* extraídas, mas apenas um direcionamento da capacidade. Ademais, estes ruídos podem ter gerado a presença de usuários atípicos além do esperado, que afetaria as métricas de modelos menos complexos e robustos, como o ELM e *Pruning-ELM*.

Em suma, é possível concluir que o serviço proposto para recomendação de eventos, com a utilização das *features* propostas, assim como com o uso do modelo classificador *XGBoost* proposto, possuiu performance satisfatória, no geral. No entanto, há pontos onde mostrou-se necessárias melhorias e estudos mais aprofundados.

## 5.2 Trabalhos Futuros

Tendo em vista as observações, limitações e hipóteses apresentadas anteriormente, propõe-se o seguinte visando a melhoria do serviço de recomendação proposto neste trabalho:

- Análise dos usuários tidos como atípicos com a base atual;
- Avaliação dos modelos citados, assim como das *features* em uma base com menos caráter aleatório, como: categorias de interesse dos usuários, categorias dos eventos e datas dos eventos;
- Criação de modelos para classificação de personalidade, sentimento e emoções baseado em textos;
- Expansão do módulo integrador de dados para outras plataformas e utilização deste módulo para aperfeiçoamento da base proposta;
- Análise de outros valores de hiper-parâmetros para os modelos;
- Uso de diferentes modelos por usuários para análise de métricas de performance geral do classificador;
- Análise da viabilidade dos modelos propostos serem executados em sistemas sem tantos recursos computacionais, como celulares;
- Análise da escalabilidade do serviço.

# Referências

- ACHEAMPONG, F. A.; WENYU, C.; NUNOO-MENSAH, H. Text-based emotion detection: Advances, challenges, and opportunities. **Engineering Reports**, v. 2, n. 7, p. e12189, 2020. Acesso em: 22 out. 2023. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/eng2.12189>>.
- BOBADILLA, J.; ALONSO, S.; HERNANDO, A. Deep learning architecture for collaborative filtering recommender systems. **Applied Sciences**, v. 10, n. 7, 2020. ISSN 2076-3417. Acesso em: 17 out. 2023. Disponível em: <<https://www.mdpi.com/2076-3417/10/7/2441>>.
- BOSLAUGH, S. **Statistics in a nutshell**. 2. ed. [S.l.]: O'Reilly Media, 2012. 182-186 p. ISBN 9781449361143.
- CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. ACM, 2016. (KDD '16). Disponível em: <<http://dx.doi.org/10.1145/2939672.2939785>>.
- DHELIM, S. *et al.* A survey on personalinty-aware recommendation systems. **Artificial Intelligence Review**, 2021.
- HUANG, G.-B.; ZHU, Q.-Y.; SIEW, C.-K. Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE. **2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)**. [S.l.], 2004. v. 2, p. 985–990.
- KO, H. *et al.* A survey of recommendation systems: Recommendation models, techniques, and application fields. **Eletronics - MDPI**, v. 11, p. 141, 2022.
- LAN, T. *et al.* Research on the prediction system of event attendance in an event-based social network. **Wireless Communications and Mobile Computing**, Hindawi, v. 2022, 2022.
- LI, B. *et al.* An event recommendation model using elm in event-based social network. **Neural Computing and Applications**, Springer, v. 32, p. 14375–14384, 2020.
- LIANG, Y. Dfgr: diversity and fairness awareness of group recommendation in an event-based social network. **Neural Processing Letters**, Springer, p. 1–20, 2023.
- LIU, X. *et al.* Event-based social networks: Linking the online and offline social worlds. In: **Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining**. New York, NY, USA: Association for Computing Machinery, 2012. (KDD '12), p. 1032–1040. ISBN 9781450314626. Acesso em: 08 nov. 2023. Disponível em: <<https://doi-org.ez28.periodicos.capes.gov.br/10.1145/2339530.2339693>>.
- LIU, Y. *et al.* A boosting algorithm for item recommendation with implicit feedback. In: **Twenty-Fourth International Joint Conference on Artificial Intelligence**. [S.l.: s.n.], 2015.
- MACEDO, A. Q.; MARINHO, L. B.; SANTOS, R. L. Context-aware event recommendation in event-based social networks. In: **Proceedings of the 9th ACM Conference on Recommender Systems**. New York, NY, USA: Association for Computing Machinery, 2015. (RecSys '15), p. 123–130. ISBN 9781450336925. Acesso em: 20 nov. 2023. Disponível em: <<https://doi.org/10.1145/2792838.2800187>>.

- MAENG, H. Y.; JANG, H. Y.; LI, J. M. A critical review of the motivational factors for festival attendance based on meta-analysis. **Tourism Management Perspectives**, v. 17, p. 16–25, 2016. ISSN 2211-9736. Acesso em: 15 nov. 2023. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2211973615300106>>.
- MAMMONE, A.; TURCHI, M.; CRISTIANINI, N. Support vector machines. **WIREs Computational Statistics**, v. 1, n. 3, p. 283–289, 2009. Acesso em: 22 jan. 2024. Disponível em: <<https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wics.49>>.
- NANDWANI, P.; VERMA, R. A review on sentiment analysis and emotion detection from text. **Social Network Analysis and Mining**, v. 11, n. 1, p. 81, Aug 2021. ISSN 1869-5469. Acesso em: 22 out. 2023. Disponível em: <<https://doi.org/10.1007/s13278-021-00776-6>>.
- NKUNA, G.; COETZEE, M. Social event invitation and recommendation for event based social networks. In: **2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)**. [S.l.: s.n.], 2020. p. 1–5.
- RENDLE, S. *et al.* **BPR: Bayesian Personalized Ranking from Implicit Feedback**. 2012.
- ROY, D.; DUTTA, M. A systematic review and research perspective on recommender systems. **Journal of Big Data**, 2022.
- SAGAN, C. **Pale Blue Dot: A Vision of the Human Future in Space**. New York: Random House, 1994. ISBN 978-0345376596.
- Scikit-learn. **sklearn.model\_selection.GridSearchCV**. 2024. Acesso em: 11 ago. 2024. Disponível em: <[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)>.
- SHAH, L.; GAUDANI, H.; BALANI, P. Survey on recommendation system. **International Journal of Computer Applications**, Foundation of Computer Science, v. 137, n. 7, p. 43–49, 2016.
- SHELKE, N. *et al.* An efficient way of text-based emotion analysis from social media using Irdnn. **Neuroscience Informatics**, v. 2, n. 3, p. 100048, 2022. ISSN 2772-5286. Multimedia-based Emerging Technologies and Data Analytics for Neuroscience as a Service (NaaS). Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2772528622000103>>.
- TABOADA, M. Sentiment analysis: An overview from linguistics. **Annual Review of Linguistics**, v. 2, n. 1, p. 325–347, 2016. Acesso em: 21 out. 2023. Disponível em: <<https://doi.org/10.1146/annurev-linguistics-011415-040518>>.
- TAN, Z.; HE, L. An efficient similarity measure for user-based collaborative filtering recommender systems inspired by the physical resonance principle. **IEEE Access**, v. 5, p. 27211–27228, 2017.
- THORAT, P. B.; GOUDAR, R. M.; BARVE, S. Survey on collaborative filtering, content-based filtering and hybrid recommendation system. **International Journal of Computer Applications**, Foundation of Computer Science, v. 110, n. 4, p. 31–36, 2015.
- WANG, Z. *et al.* Exploiting social influence for context-aware event recommendation in event-based social networks. In: **IEEE INFOCOM 2017 - IEEE Conference on Computer Communications**. [S.l.: s.n.], 2017. p. 1–9.

WU, X.; PUSAKSRIKIT, T. Exploring the relationship between the bigfive personality traits and exhibition shows attendance. **International Journal of Business and Economics**, v. 10, n. 1, p. 95–117, 2018.

WU, Y. *et al.* Group recommendation based on heterogeneous graph algorithm for ebsns. **IEEE Access**, v. 11, p. 1854–1866, 2023.

XU, Q. A.; CHANG, V.; JAYNE, C. A systematic review of social media-based sentiment analysis: Emerging trends and challenges. **Decision Analytics Journal**, v. 3, p. 100073, 2022. ISSN 2772-6622. Acesso em: 21 out. 2023. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2772662222000273>>.

YIN, H. *et al.* Joint event-partner recommendation in event-based social networks. In: **2018 IEEE 34th International Conference on Data Engineering (ICDE)**. [S.l.: s.n.], 2018. p. 929–940.

ZHAO, B. Web scraping. **Encyclopedia of big data**, Springer Living ed. Cham, v. 1, 2017.