
Rede Social *Gamer* com modelagem do algoritmo K-means
para indicação de novas amizades

Marcus Antonio Rufino Ribeiro



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

Monte Carmelo – MG
2022

Marcus Antonio Rufino Ribeiro

Rede Social *Gamer* com modelagem do algoritmo *K-means*
para indicação de novas amizades

Trabalho de Conclusão de Curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Área de concentração: Sistemas de Informação

Orientador: Prof. Dr. Rafael Dias Araújo

Dedico este trabalho para minha mãe Aparecida de Lurdes Rufino que sempre me incentivou a concluir minha formação para ter um futuro melhor.

Agradecimentos

Agradeço primeiramente a Deus por mais essa etapa concluída em minha vida.

Agradeço a toda minha família em especial a minha mãe Aparecida de Lurdes Rufino e aos meus padrinhos Maria Madalena e Eurípedes Moraes que me ajudaram em todo momento da minha graduação.

Agradeço a todos meus professores em especial ao meu professor e orientador Rafael Dias Araújo pelo apoio e dedicação com a ideia desse projeto.

Agradeço a todos os meus amigos de curso, onde fiz amizades que irei levar para a vida toda.

E, por fim, agradeço a toda comunidade acadêmica pelo ambiente amigável para estudos e atividades de crescimento pessoal que sempre manteve em minha formação.

Resumo

Redes sociais relacionadas a games na Internet já é uma realidade consolidada. Graças à praticidade, ao acesso de ter sempre o celular em mãos e ao fato de se poder jogar em qualquer lugar, os jogos para dispositivos móveis vêm dominando esse setor. No entanto, uma pesquisa de opinião mostrou que jogadores possuem dificuldade em encontrar outros jogadores geograficamente próximos que gostam desse tipo de esporte para se conectarem. Assim, este trabalho tem como o objetivo a criação de uma rede social focada para o público *gamer*. Para o desenvolvimento da rede social foram utilizadas as tecnologias de HTML, CSS, Java Script, PHP, e banco de dados. Para a parte de sugestões de amizades foi utilizado o método de clusterização *k-means* em conjunto com o Método *Elbow* para saber qual seria o valor ideal de K para a aplicação. Experimentos realizados com dados sintéticos mostraram a viabilidade da proposta.

Palavras-chave: Rede Social, Indicação de Amizades, *K-means*, Método *Elbow*.

Abstract

Social networks for the Games domain on the Internet are already a consolidated reality. Thanks to practicality, to mobile access always at hand, and to the fact that we can play anywhere, mobile games are dominating this sector. However, a survey has shown that players have difficulty finding other geographically close players who enjoy this type of sport to connect with. Thus, this work aims to create a social network focused on the gamer audience. For its development, programming languages such as HTML, CSS, Java Script, PHP, and database technologies were used. For the friendship recommendation, the k-means clustering method was used in conjunction with the Elbow Method to find out what would be the ideal value of K for the application. Experiments carried out with synthetic data showed the viability of the proposal.

Palavras-chave: Social Network, Friendship Recommendation, K-means, Elbow Method.

Lista de Siglas e Abreviaturas

ASCII	American Standard Code
CSS	Cascading Style Sheets
DBMS	Data Base Management System
DT	Decision Tree
DTNs	Delay-tolerant networking
HTML	HyperText Markup Language
JS	JavaScript
KNN	K-Nearest Neighbor
LDA	Latent Dirichlet Allocation
NB	Narrowband
NMF	Non-Negative Matrix Factorization
QP	Quadratic Programming
SGBD	Sistema gerenciador de banco de dados
SSE	Sum Squared Error
SVM	Support Vector Machine
TI	Tecnologia da Informação
UC	Caso de Uso
UNIX	Uniplexed Information and Computing System
W3C	World Wide Web Consortium
UML	Unified Modeling Language
WSS	Within Sum of Squares

Lista de Figuras

Figura 1. <i>Clusterização</i> de uma base de dados.....	21
Figura 2. Base de dados modelo íris.....	22
Figura 3. Base de dados clusterizada do modelo iris.....	22
Figura 4. Equação do Agrupamento K-means	24
Figura 5. Repetição do processo até convergir nova média do cluster.....	25
Figura 6. Distância euclidiana para calcular a distância entre o centroide e a observação	26
Figura 7. Cross-tabulation do Pandas para comparar a “clusterização” com a classe de cada grupo	26
Figura 8. Curva do cotovelo	27
Figura 9. Valor da Silhueta.....	28
Figura 10. Diagrama Caso de Uso.....	33
Figura 11. Modelo de Dados	35
Figura 12. Interface de Login	36
Figura 13. Tela de cadastro.....	37
Figura 14. Tela de amigos do usuário logado.....	37
Figura 15. Tela de jogos cadastrados no sistema.	38
Figura 16. Tela de um jogo selecionado.....	39
Figura 17. Tela Principal	40
Figura 18. Tela de um usuário selecionado	40
Figura 19. Tela de solicitações de amizades	41
Figura 20. Resposta formulário 1	42
Figura 21. Resposta formulário 2	43
Figura 22. Resposta formulário 3	43
Figura 23. Resposta formulário 4	44
Figura 24. Trecho da lista usuários cadastrados no sistema Rede Gamer	45
Figura 25. Experimento 1	46
Figura 26. Experimento 2	47
Figura 27. Número de clusters para dois conjuntos de dados aleatórios do sistema Rede Gamer.	48

Figura 28. Experimento 3	49
--------------------------------	----

Lista de Quadros

Quadro 1. Identificação dos Atores	34
Quadro 2. Identificação dos Casos de Uso	34

Sumário

1 INTRODUÇÃO	12
1.1 OBJETIVOS E DESAFIOS DA PESQUISA	12
1.2 QUESTÕES DE PESQUISA	13
1.3 METODOLOGIA	13
1.4 ORGANIZAÇÃO DA MONOGRAFIA.....	14
2 FUNDAMENTAÇÃO TEÓRICA	15
2.1 CONCEITOS BÁSICOS	15
2.1.1 HTML.....	15
2.1.2 CSS	16
2.1.3 JavaScript.....	18
2.1.4 Bootstrap	18
2.1.5 Banco de Dados	19
2.1.6 Agrupamento.....	20
2.2 TRABALHOS SOBRE MODELAGEM E RECOMENDAÇÃO DE USUÁRIOS	29
3 DESENVOLVIMENTO DO SISTEMA REDE GAMER.....	32
3.1 ANÁLISE E PROJETO	32
3.2 PROTÓTIPO FUNCIONAL.....	36
4 EXPERIMENTOS E ANÁLISE DOS RESULTADOS	41
4.1 PESQUISA DE OPINIÃO	41
4.2 MÉTODO PARA A AVALIAÇÃO	43
4.3 EXPERIMENTOS	44
4.4 AVALIAÇÃO DOS RESULTADOS	49
5 CONSIDERAÇÕES FINAIS	50
5.1 PRINCIPAIS CONTRIBUIÇÕES	50
5.2 TRABALHOS FUTUROS	51
REFERÊNCIAS.....	52
APÊNDICE A - FORMULÁRIO DE PESQUISA DE OPINIÃO.....	56

Introdução

Segundo uma pesquisa realizada pela (NEWZOO, 2019), “na última década (2009-2019), os jogos cresceram de um passatempo popular para um fenômeno de entretenimento geral, com o mercado global de jogos chegando a US \$ 148,1 bilhões em 2019”. Os jogos para dispositivos móveis têm dominado esse setor, isso se deve à praticidade, à acessibilidade de estar sempre com o celular na mão e ao fato de poder jogar em qualquer lugar. Além disso, há grandes dificuldades em aproximar as pessoas que gostam desse esporte, falta um ambiente e um local específico onde essas pessoas possam se comunicar e se reunir para trocar informações e jogar.

Uma das maneiras pelas quais essas pessoas se conectam é através das mídias sociais. Segundo Diana (nd), professora de Biologia e doutora em Gestão do Conhecimento, as redes sociais são espaços virtuais em que grupos de pessoas ou empresas interagem enviando mensagens e compartilhando conteúdos, entre outras coisas (NEWZOO, 2019).

Redes sociais de finalidades semelhantes encontram-se disponíveis na Internet, tais como a *Steam*¹, a *Origin*², a *Epic Games*³, exceto pelo fato de que essas vendem o jogo aos *gamers*, diferentemente desta proposta em que os *gamers* apenas utilizam o sistema para conexão e interação com outros jogadores afim de socializar suas percepções sobre os jogos.

Para modelar as sugestões de amizades na rede social proposta neste trabalho, levou-se em consideração a região onde o usuário vive e os jogos que o mesmo adicionou como favorito levando em conta o tipo de gênero que mais gosta, pois, o objetivo é sugerir possíveis amizades entre pessoas da mesma cidade ou cidades e regiões próximas e que gostem dos mesmos jogos ou estilos de jogos semelhantes do usuário que está logado no sistema.

1.1 Objetivos e Desafios da Pesquisa

O objetivo é projetar e implementar uma rede social onde *gamers* possam se conectar, fazer novas amizades, compartilhar informações e se reunir para jogar.

¹ <https://store.steampowered.com/>

² <https://www.origin.com/bra/pt-br/store>

³ <https://store.epicgames.com/pt-BR/>

Para o *gamer* se conectar com outras pessoas, ele pode se conectar buscando pelo nome da pessoa ou encontrando a pessoa na sugestão de amizades que a própria rede irá fornecer, através da estrutura de agrupamento, levando em conta a região e os jogos que o usuário adicionar como favorito.

O desafio consiste em desenvolver a rede social onde os usuários se sintam à-vontade em acessá-la e dar suas opiniões e quais atributos utilizar para a criação do algoritmo para as sugestões de amizades. Além disso, almejou-se chegar a um modelo de página do projeto responsiva tanto para as telas de computadores e para o uso de dispositivos móveis, com funcionalidade de sugestão de novas amizades para os usuários já cadastrados e trazer os primeiros usuários, além de incentivá-los a sempre estar usando a rede social.

Para a parte de sugestões de amizades foi utilizado algoritmo de agrupamento, para agrupar usuários com gostos semelhante de gêneros e de jogos e que estejam em uma região próxima, para que eles possam, além jogar juntos seus jogos em comuns, terem a possibilidade de se conhecerem pessoalmente.

1.2 Questões de Pesquisa

Este trabalho foi guiado pelas seguintes questões de pesquisa (QP):

QP1: A criação de uma rede social específica para *gamers* desperta o interesse dos usuários em acessar e interagir com amigos antes de iniciar algum jogo?

QP2: Quais os atributos dos usuários que são mais relevantes para a recomendação de amizade nesse contexto?

1.3 Metodologia

Em um primeiro momento, ao idealizar a criação da rede social de propósito específico, foi feita uma pesquisa de opinião com potenciais usuários para verificar as necessidades e aceitação da proposta.

Tecnologias para desenvolvimento Web, como HTML, CSS, Java Script, PHP, e banco de dados relacionais foram utilizados para criar um protótipo funcional de uma rede social voltada para o público *gamer*. Para a parte de sugestões de amizades, foram realizados testes experimentais exploratórios com os algoritmos *dbscan* e *k-means*. Após testes com o algoritmo do *dbscan*, ele foi descartado pois sua finalidade é encontrar *outline* e não seria legal que usuários considerados como *outline* não recebessem nenhuma sugestão de amizade.

O método cotovelo foi utilizado para encontrar qual seria o valor ideal de K (quantidade de grupos) para a aplicação. Para saber quais variáveis utilizar no algoritmo de sugestão de amizades, foram feitas análises experimentais exploratórias para identificar qual combinação traria o melhor resultado ao usuário.

1.4 Organização da Monografia

O Capítulo 2 apresenta a fundamentação teórica acerca deste trabalho, com os principais conceitos necessários para o entendimento e os trabalhos encontrados na literatura, bem como apresenta trabalhos sobre modelagem e recomendação de usuários. O Capítulo 3 apresenta o desenvolvimento do sistema. E o Capítulo 4 descreve a forma como ocorreram os experimentos e análise dos resultados, bem como o método para a avaliação dos resultados. Por fim, o Capítulo 5 traz as considerações finais deste trabalho.

Fundamentação Teórica

2.1 Conceitos Básicos

2.1.1 HTML

HyperText Markup Language (HTML), expressão inglesa que significa “linguagem de marcação de hipertexto”, é uma linguagem de marcação para produção páginas na Web, que pode ser lida em qualquer computador e transmitida pela Internet (MILETTO, 2014).

A principal característica da linguagem HTML é formatar textos usando marcações especiais, as chamadas *tags*, para que possam ser convenientemente exibidos por clientes da Web, também conhecidos como *browsers* ou navegadores. Além disso, essa linguagem permite a conexão entre páginas da Web, criando documentos com o conceito de hipertexto (PEDROSO, 2007).

As funções básicas da linguagem HTML são:

- Documentos HTML são arquivos de texto escritos em ASCII⁴;
- HTML não diferencia letras maiúsculas de minúsculas em sua marcação, ou seja, não diferencia maiúsculas de minúsculas;
- Nem todas as *tags* e suas funções correspondentes são suportadas por todos os navegadores. Se um cliente da Web não entender marcação, ele simplesmente a ignorará;
- Os arquivos HTML podem ter a extensão ".html" ou ".htm". O primeiro geralmente é usado em sistemas UNIX e o segundo em sistemas Windows. Os navegadores podem exibir documentos com ambas as extensões (PEDROSO, 2007).

Para escrever um documento HTML, basta um simples editor de texto e conhecimento dos códigos que compõem a linguagem, chamados de *tags*, que indicam a função de cada

⁴ ASCII (*American Standard Code for Information Interchange*) é o formato mais comum usado em arquivos texto em computadores e na Internet. Num arquivo ASCII, cada caractere alfabético, numérico ou especial, é representado por um número binário de sete bits. É possível representar até 128 caracteres com este código.

elemento, os navegadores (*browsers*) identificam as *tags* e apresentam as páginas conforme especificado.

A linguagem HTML pertence a uma classe de linguagens de programação conhecida como “linguagens de *tags*” (*Tag Languages*) ou somente “linguagens de marcação”. Nesse tipo de linguagem, os comandos são escritos na forma de marcadores chamados *tags*. Em geral, as *tags* vêm em pares e delimitam o texto a ser formatado. A *tag* que abre o escopo de um comando pode ser identificada por caracteres.

Por exemplo:

```
<B>Linguagem</B> de Marcação.
```

A sintaxe acima determina que a marcação ‘B’ seja aplicada ao texto ‘Linguagem’, mas não ao texto ‘de Marcação’ (PEDROSO, 2007).

Existem exceções para esta operação de paridade de *tag* onde a *tag* `<tag>` não requer a *tag* `</tag>` correspondente. Mais adiante ambos os tipos se encontram exemplificados. Além do próprio comando, uma *tag* pode conter seus parâmetros. Os parâmetros são características de cada marcação que permitem que ela seja executada de diferentes maneiras. Pode-se dizer que enquanto um comando diz ao navegador o que fazer, seus parâmetros informam como fazê-lo, por meio de suas opções. Assim como cada comando possui parâmetros específicos, cada parâmetro possui opções específicas (PEDROSO, 2007).

Veja o exemplo:

```
<FONT SIZE="4"COLOR="green"> Texto </FONT>
```

O exemplo demonstrado mostra que 'FONT' é o comando e os parâmetros são 'SIZE' e 'COLOR', cujas opções escolhidas pelo programador foram '4' e 'green', respectivamente. Observe que, diferentemente da marcação de abertura, a marcação de fechamento de um comando não precisa mencionar seus parâmetros (PEDROSO, 2007).

2.1.2 CSS

Cascading Style Sheets (CSS), que no português significa “Folhas de Estilo em Cascata”, funciona como a decoração da página html, pois quando o HTML foi criado ele não

suportava *Tags* que ajudariam na formatação das páginas. Assim a relação entre o HTML e o CSS é bem forte pois o CSS foca em toda a estética de um *site* (MILETTO, 2014).

Em outubro de 1994, as CSS foram propostas pela primeira vez por Hakon Lie, que queria simplificar o que era então uma programação de *sites* muito mais complexa. As pessoas tiveram que usar mais código para chegar a um resultado simples como criar uma tabela (MILETTO, 2014).

Em 1995, o World Wide Web Consortium (W3C) – um grupo da indústria de computadores – desenvolveu o CSS1. A linguagem de estilo ganhou destaque entre 1997 e 1999, período em que se tornou familiar para a maioria dos programadores (WEBNATAL, 2007).

Inicialmente, o HTML era a única linguagem usada para construir *sites*. Conforme ganhou popularidade, alguns comandos (*tags*) foram criados pelos navegadores para facilitar o uso da linguagem. Para facilitar ainda mais a criação desses *layouts*, o W3C desenvolveu o CSS e o disponibilizou para *Web designers* (WEBNATAL, 2007).

CSS é uma linguagem que gera o *layout* dos *sites*. Este programa permite ao usuário criar páginas *Web* com códigos mais fáceis de escrever do que códigos HTML. Esses códigos permitem enviar solicitações facilmente (WEBNATAL, 2007).

Programadores de todo o mundo utilizam este programa. O CSS controla opções de borda, linhas, cores, alturas, larguras, imagens e posicionamento sem ter que codificar em HTML. O CSS também possui alguns códigos pré-construídos que permitem aos usuários economizar tempo criando códigos muito comuns (WEBNATAL, 2007).

Algumas vantagens do CSS:

1. A partir de um único documento CSS podem ser controlados vários documentos .html;
2. Fácil de criar layouts que não precisam de códigos muito complicados;
3. Linguagem de aprendizado rápido (WEBNATAL, 2007).

O CSS é um dos programas que facilitam a vida de quem trabalha na área de TI. Embora este programa tenha algumas falhas - na verdade, as falhas são causadas por navegadores que não implementam CSS corretamente, como algumas versões do *Internet Explorer* - ajuda muito na construção de *sites* e páginas da *Web* (WEBNATAL, 2007).

2.1.3 JavaScript

JavaScript é uma linguagem de programação da Web que serve para controlar o HTML e o CSS, criada em 1995 por Brendan Eich com o propósito de tornar as páginas Web dinâmicas e com o uso mais agradável (FLANAGAN, 2012).

Como o nome sugere, Java Script é uma linguagem de script, amplamente definida como uma linguagem de programação que permite ao programador controlar um ou mais aplicativos de terceiros. Com o JavaScript, é possível controlar alguns comportamentos do navegador por meio de trechos de código enviados na página HTML (CAELUM, 2020, p. 227).

Trata-se de uma linguagem no desenvolvimento Web mais popular. Uma linguagem responsável por praticamente qualquer tipo de dinamismo necessário às páginas, e suportada por todos os navegadores. Se usarmos todo o poder que o JavaScript tem para oferecer, é possível chegar a resultados impressionantes. Exemplos disso são aplicativos complexos da Web como Gmail, Google Maps e Google Docs (CAELUM, 2020, p. 226).

Outro ponto em comum entre as linguagens de *scripting* é que elas são linguagens tipicamente interpretadas, isto é, não dependem de nenhuma compilação para serem realizadas. Esse recurso está presente no JavaScript: o código é interpretado e executado conforme lido pelo navegador, linha por linha, assim como o HTML (CAELUM, 2020, p. 227).

As conversões automáticas do JavaScript são executadas em tempo real o que permite uma alta tolerância a erros. Como se verá no decorrer das explicações, nem sempre essas conversões levam a algo esperado, o que pode ser fonte de muitos erros se não conhecermos bem esse mecanismo (CAELUM, 2020, p. 227).

O script do programador é enviado ao navegador com o HTML, mas para que o navegador diga a diferença entre o script e o HTML, o script deve estar dentro da *tag* `<script>` (CAELUM, 2020, p. 227).

2.1.4 Bootstrap

Bootstrap⁵ é um framework que auxilia no desenvolvimento do front-end na parte do desenvolvimento com HTML, CSS e JS, melhorando a experiência do usuário (DUARTE, 2017) para a criação de sites e aplicações responsivas de forma rápida e simples. Além disso, pode lidar com sites de desktop e páginas de dispositivos móveis da mesma forma.

⁵ <https://getbootstrap.com/>

Originalmente, o Bootstrap foi desenvolvido para o Twitter por um grupo de desenvolvedores liderados por Mark Otto e Jacob Thornton Logo e se tornou uma das estruturas de front-end e projetos de código aberto mais populares do mundo.

Antes do Bootstrap ser um framework de código-fonte aberto, ele era conhecido como Twitter Blueprint. Após alguns meses de desenvolvimento, o Twitter realizou sua primeira Hack Week, que foi um projeto que ganhou ampla visibilidade à medida que desenvolvedores de todos os níveis de habilidade usavam a estrutura sem orientação externa. Após este fato, o projeto serviu como um guia de estilo para o desenvolvimento de ferramentas internas na empresa por mais de um ano antes de ser lançado ao público usuário (DUARTE, 2017).

Segundo Hesterberg (2011), o bootstrap fornece inferências estatísticas - erro padrão e estimativas de viés, intervalos de confiança, e testes de hipóteses - sem suposições, como distribuições normais ou variações iguais. Como tal, os métodos bootstrap podem ser notavelmente mais precisas do que as inferências clássicas baseadas em distribuições Normal ou t. O bootstrap usa o mesmo procedimento básico, independentemente da estatística ser calculado, sem exigir o uso de fórmulas específicas de aplicação.

2.1.5 Banco de Dados

Banco de dados é um conjunto de dados que se relacionam e criam uma informação, esses dados são controlados por um Sistema de Gerenciamento de Banco de Dados (SGBD), que são programas que permitem a criação, manipulação e compartilhamento de banco de dados entre usuários e aplicações (SILBERSCHATZ *et al.*, 2012).

O sistema gerenciador de bancos de dados (SGBD) – do inglês Data Base Management System (DBMS) – é um conjunto de programas que gerenciam a estrutura do banco de dados e controlam o acesso aos dados armazenados. Até certo ponto, o banco de dados se assemelha a um arquivo eletrônico com conteúdo muito bem-organizado com a ajuda de um software poderoso, conhecido como sistema gerenciador de banco de dados (ROB; CORONEL, 2011).

A primeira parte do quinto capítulo da obra de Lev Manovich “The Language of New Media” (2001) aborda o banco de dados. O referido capítulo esclarece seu conceito como fundamental para a compreensão das novas mídias, argumenta sobre a relação entre dados e algoritmo, faz a distinção entre banco de dados e narrativa e propõe uma análise do cinema de Peter Greenaway e Dziga Vertov sob a lógica do banco de dados (MANOVICH, 2015).

Dados são diferentes de informações. Os dados são fatos brutos, isto é, os fatos ainda não foram processados para revelar seu significado. Portanto, é preciso transformar os dados brutos em um resumo de dados. Já as informações são o resultado do processamento de dados brutos para revelar seu significado. Esse processamento pode ser simples, como a organização dos dados para revelar padrões, ou complexos, como a realização de previsões ou a extração de inferências utilizando modelagem estatística. Para revelar seu significado, as informações exigem um contexto. Por exemplo, uma leitura de temperatura média de 105° não tem muito significado, a menos que saibamos seu contexto: está em graus Fahrenheit ou Celsius? Trata-se da temperatura de uma máquina, de um corpo ou atmosférica? As informações podem ser utilizadas como o fundamento para a tomada de decisões. Por exemplo, o resumo de dados de cada questão do formulário de entrevista pode apontar os pontos fortes e fracos do laboratório, ajudando-o a tomar decisões confiáveis para melhor atender às necessidades de seus clientes (ROB; CORONEL, 2011).

Em regra, o gerenciamento de dados eficiente requer o uso de um banco de dados computadorizado. Uma base de dados (ou banco de dados) é uma estrutura de computação compartilhada e integrada que armazena:

- Dados do usuário final, ou seja, dados brutos de interesse desse usuário.
- Metadados, ou dados por meio dos quais integram e gerenciam dados do usuário final (ROB; CORONEL, 2011).

Os metadados fornecem uma descrição das características dos dados e o conjunto de relacionamentos que vinculam os dados encontrados no banco de dados. O componente de metadados, por exemplo, armazena informações como o nome de cada item de dados, o tipo de valor armazenado (numérico, datas ou texto), se esse item deve ou não ser deixado em branco etc. Logo, os metadados fornecem informações que complementam e aprimoram o valor e o uso dos dados. Em suma, os metadados acarretam uma representação mais completa dos dados no banco de dados. Devido às propriedades dos metadados, é possível ouvir a definição de banco de dados como “um conjunto de dados auto descritivos” (ROB; CORONEL, 2011, p. 7).

2.1.6 Agrupamento

Agrupamento, ou *clusterização*, é uma técnica de aprendizado de máquina não-supervisionada. Na criação do modelo de *clusterização* não se utiliza a classe alvo, por isso é chamado de não-supervisionado.

Também não se faz a separação do *dataset* entre treino e teste. Utiliza-se todos os dados, sem fazer divisões das linhas dos dados. A Figura 1 mostra uma base de dados que através de suas características semelhantes o algoritmo de agrupamento dividiu ela em dois *clusters*.



Figura 1. *Clusterização* da base de dados.
Fonte: Kaggle (2019).

No aprendizado não-supervisionado, as colunas do *dataset* são chamadas de dimensões e não mais de atributos ou variáveis, conforme se vê nas Figuras 2 e 3. Por exemplo, no *dataset* de Flores (*iris dataset*) tem-se 4 dimensões: altura da pétala, largura da pétala, altura da sépala e largura da sépala (GUEDES, 2019).

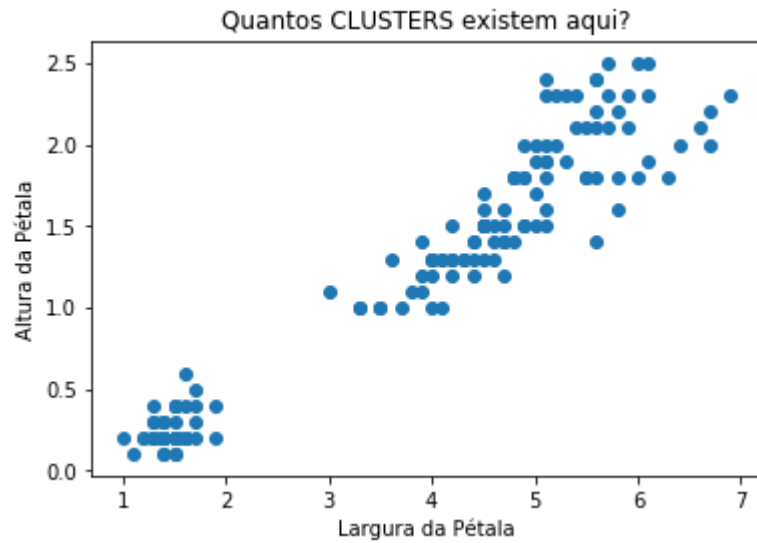


Figura 2. Base de dados modelo iris.
Fonte: Kaggle, 2019.

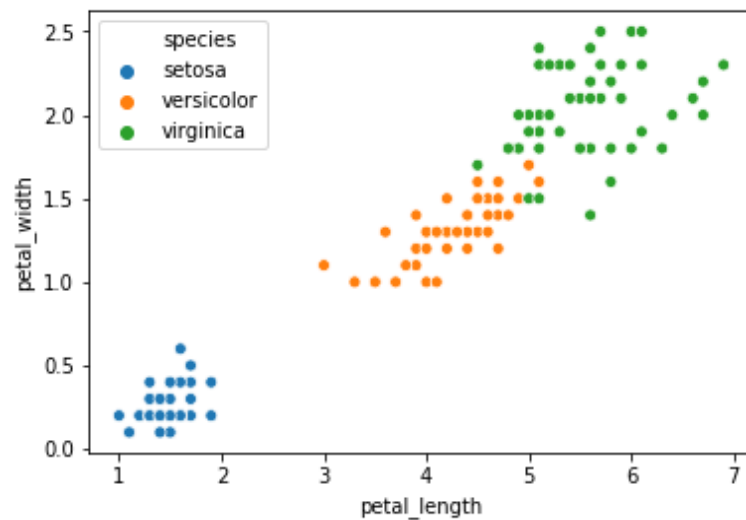


Figura 3. Base de dados clusterizada do modelo iris.
Fonte: Kaggle, 2019.

Dadas várias imagens não rotuladas, pode-se utilizar o agrupamento (a *clusterização*) para agrupar as imagens por similaridade. Seria até possível identificar as anomalias, ou seja, aqueles casos que diferem de todos os outros.

O *clustering* (agrupamento) pode ser usado tanto para preparar dados com finalidade de encontrar padrões que ainda não são conhecidos, quanto para construir modelos para criar *clusters*, separando os dados em grupos que ainda não são conhecidos. O *clustering* já é aplicado em várias situações, por exemplo, no reconhecimento de padrões, análise de imagens, agrupamento de espécies, detecção de anomalias, compressão de dados, agrupamento de clientes, classificação de documentos, agrupamento de notícias (GUEDES,

2019). Por esse motivo o algoritmo de agrupamento foi utilizado na parte de sugestões de amizades, para agrupar usuários com gostos semelhante de gêneros e de jogos e que estejam em uma região próxima, podendo assim além jogar juntos seus jogos em comuns, se conhecerem também na vida real. Nesse cenário, as instâncias de dados são as pessoas que não possuem uma classificação prévia, dessa forma, faz sentido usar algum algoritmo de agrupamento, que é uma técnica não supervisionada.

Quando se fala sobre aprendizado de máquinas existem alguns conceitos importantes que devem ser entendidos:

- *Overfitting*: se refere à situação em que o algoritmo se ajusta aos dados de treinamento, mas perdeu a capacidade de generalização para dados novos. Em outras palavras, em vez de aprender, o algoritmo apenas memorizou os dados de treinamento (FREITAS, 1998). Essa situação pode ocorrer quando a rede tem mais parâmetros do que necessário para a resolução do problema (REZENDE, 2005).
- *Underfitting*: se refere à situação em que o algoritmo que não se ajusta nem aos dados de treinamento nem aos novos (FREITAS, 1998). Ocorre quando a rede tem menos parâmetros do que necessário para a resolução do problema (REZENDE, 2005).

2.1.6.1 K-Means

Existem vários algoritmos para o agrupamento. O K-means costuma ser o mais aplicado em dados com poucas dimensões de dados numéricos e contínuos para organizar os dados em grupos categóricos (KAGGLE, 2019).

O agrupamento K-means tem por objetivo subdividir um conjunto de observações em um conjunto de *clusters* (k), resultando na subdivisão dos dados⁶ em células de Voronoi. K é o número de *clusters*. O K-means pode ser pensado como um método para desvendar a qual grupo um determinado objeto de fato pertence.

Ele faz isso usando a Soma de Erros Quadrados (SSE, do inglês *Sum Squared Error*), que tenta minimizar a distância entre os pontos e seu centroide. Após a criação do modelo, quando aplicado, o K-Means lembra a média de cada cluster, também chamada de centroides, e a utiliza para encontrar o centroide mais próximo de qualquer novo dado. O centroide representa o centro do cluster. A Figura 4 representa a equação do agrupamento *K-means*.

⁶ Esta divisão dos dados entre os clusters tem que ser de forma que os diferentes clusters fiquem mais separados entre eles, enquanto as observações dentro de cada cluster fiquem mais próximas entre elas.

The diagram shows the objective function equation for K-means clustering: $J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2$. Annotations include:

- 'number of clusters' pointing to the variable k .
- 'number of cases' pointing to the variable n .
- 'case i ' pointing to the variable i in the inner sum.
- 'centroid for cluster j ' pointing to the variable c_j .
- 'Distance function' pointing to the term $\|x_i^{(j)} - c_j\|^2$.
- 'objective function' pointing to the variable J .

Figura 4. Equação do Agrupamento K-means.

Fonte: Kaggle, 2019.

É usado sobretudo em estatística e pode ser aplicado a quase todos os campos de estudo. Em marketing, exemplificando, pode ser aproveitado pelos profissionais de marketing para agrupar diversos dados demográficos de pessoas em grupos simples que facilitam a divisão em segmentos.

Para resolver problemas de agrupamento, utiliza-se um algoritmo simples de aprendizado não supervisionado, o agrupamento K-means. Um procedimento simples de agrupar um determinado conjunto de dados em um número de clusters definidos pela letra “K” que é definida antecipadamente é seguido (TECHOPEDIA, 2022).

Os clusters são então posicionados como pontos e todas as observações ou pontos de dados são associados ao cluster mais próximo, computados, ajustados e então o processo recomeça utilizando os novos ajustes até que um resultado desejado seja alcançado.

O agrupamento K-means tem uso em mecanismos de busca, segmentação de mercado, estatísticas e até astronomia (TECHOPEDIA, 2022). Os astrônomos o usam para filtrar enormes quantidades de dados astronômicos; como eles não podem analisar cada objeto um por um, eles precisam de uma maneira de encontrar estatisticamente pontos de interesse para observação e investigação. O algoritmo “K pontos” é colocado no espaço de dados do objeto representando o grupo inicial de centroides. Cada objeto ou ponto de dados é atribuído ao K mais próximo.

Depois que todos os objetos são atribuídos, as posições dos K centroides são recalculadas. O algoritmo repetido até que as posições dos centroides não se movam mais (TECHOPEDIA, 2022)⁷.

⁷ <https://www.techopedia.com/definition/32057/k-means-clustering>

A clusterização com K-means compreende as seguintes etapas:

- Inicialização: a localização dos centroides é gerada aleatoriamente
- Designação: os K clusters são criados associando cada observação ao seu centroide mais próximo

Atualização: o centroide de cada *cluster* se torna a nova média do cluster. Este processo é repetido várias vezes até convergir, conforme mostra a Figura 5.

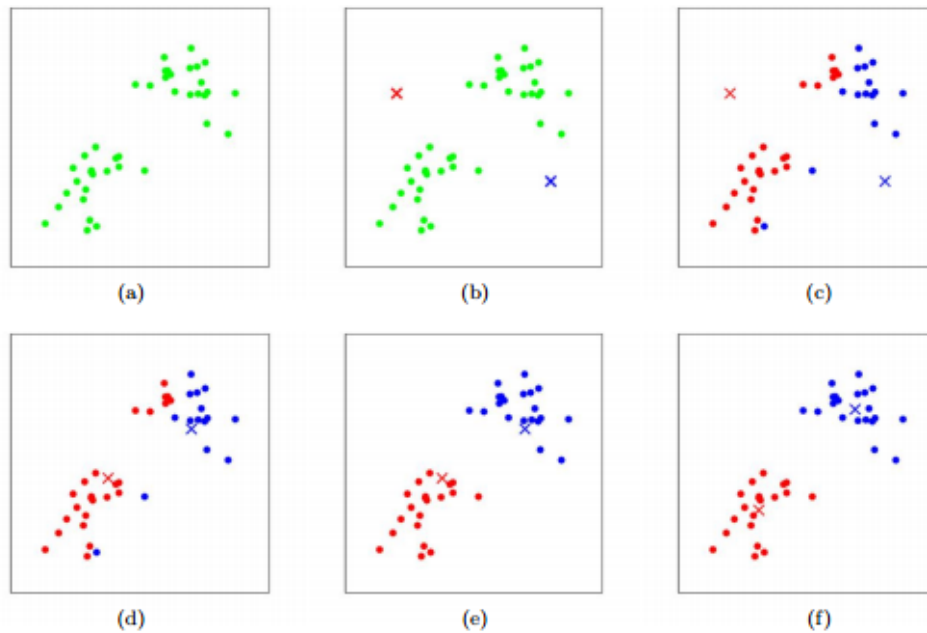
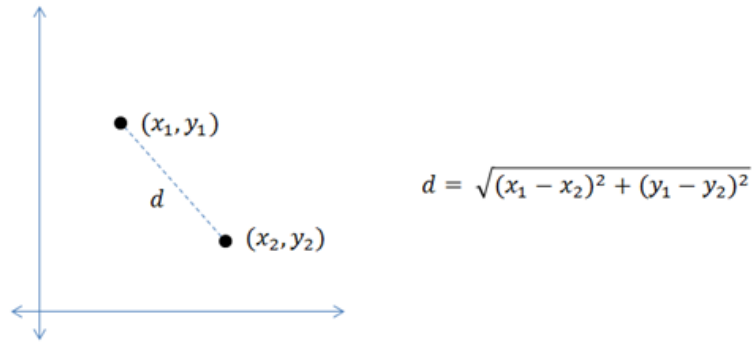


Figura 5. Repetição do processo até convergir nova média do cluster.

Fonte: Kaggle, 2019.

A similaridade é uma métrica usada para determinar a distância que mostra quão semelhantes são as observações de um mesmo grupo, ou seja, o quanto elas são mais parecidas. Ao mesmo tempo, demonstra quão distantes estão as observações de vários grupos.

Uma das métricas utilizadas para o cálculo da similaridade é a distância Euclidiana, usada para calcular a distância entre as observações e o centroide mais próximo. A Figura 6 mostra a fórmula matemática utilizada para o cálculo da distância Euclidiana e um exemplo considerando as coordenadas $D1(2,0)$ e $D2(1,3)$, assim como $D4(2,0)$ $(2,2)$.



Distance between D1 and D2	Distance between D1 and D4
$\sqrt{(2 - 1)^2 + (0 - 3)^2}$	$\sqrt{(2 - 2)^2 + (0 - 2)^2}$
$= \sqrt{(1)^2 + (3)^2}$	$= \sqrt{(0)^2 + (-2)^2}$
$= \sqrt{1 + 9}$	$= \sqrt{0 + 4}$
$= \sqrt{10} = 3.17$	$= \sqrt{4} = 2$

Figura 6. Distância euclidiana para calcular a distância entre o centroide e a observação.
 Fonte: Kaggle, 2019.

Para medir a qualidade do modelo de cluster, para avaliar o modelo, ou seja, o crosstab pandas⁸ é usado para comparar o agrupamento (ou *clustering*) com a classe de cada grupo, conforme apresenta a Figura 7.

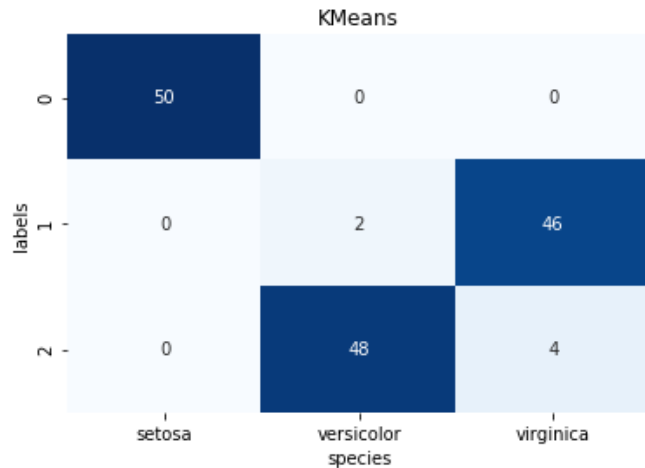


Figura 7. Cross-tabulation do Pandas para comparar a “clusterização” com a classe de cada grupo.
 Fonte: Kaggle, 2019.

⁸ Crosstab Pandas - Este método é usado para calcular uma tabulação cruzada simples de dois (ou mais) fatores. Por padrão, calcula uma tabela de frequência dos fatores, a menos que uma matriz de valores e uma função de agregação sejam passadas.
 Sintaxe: pandas.crosstab (índice, colunas, valores = Nenhum, rownames = Nenhum, colnames = Nenhum, aggfunc = Nenhum, margins = False, margins_name = 'All', dropna = True, normalize = False)

Outras estratégias, como utilização da Inércia, devem ser usadas para conjuntos de dados não rotulados. A soma dos erros quadráticos das instâncias de cada cluster é igual à inércia, e tem o objetivo de:

- Medir o quanto os clusters estão separados entre eles;
- Medir a distância de cada dado para o centroide do seu cluster;
- Calcular o erro médio (SSE) em busca de minimizar a inércia na escolha dos clusters;
- Quanto mais próximos entre si e do centroide, menor a inércia (KAGGLE, 2019).

2.1.6.2 Método *Elbow*

Um bom cluster tem o menor número de clusters e tem um baixo erro médio (SSE). A curva do cotovelo, também conhecida como abordagem da curva do cotovelo, é um método para determinar o número ideal de K clusters. Este método examina a variação dos dados em termos do número de clusters. O valor ideal de K é aquele que tem a menor *Within Sum Of Squares* (WSS) que é a soma de todos os dados até o centroide do agrupamento que eles pertencem, e também possui o menor número de clusters. Dá-se a esse conceito o nome de curva de cotovelo, pois não há tanta flutuação quanto haveria no ponto “cotovelo”. O número ideal de K-clusters seria exatamente onde o cotovelo estaria (ROUSSEEUW, 1987). Na Figura 8, o “cotovelo” acontece no número de *clusters* igual a 2.

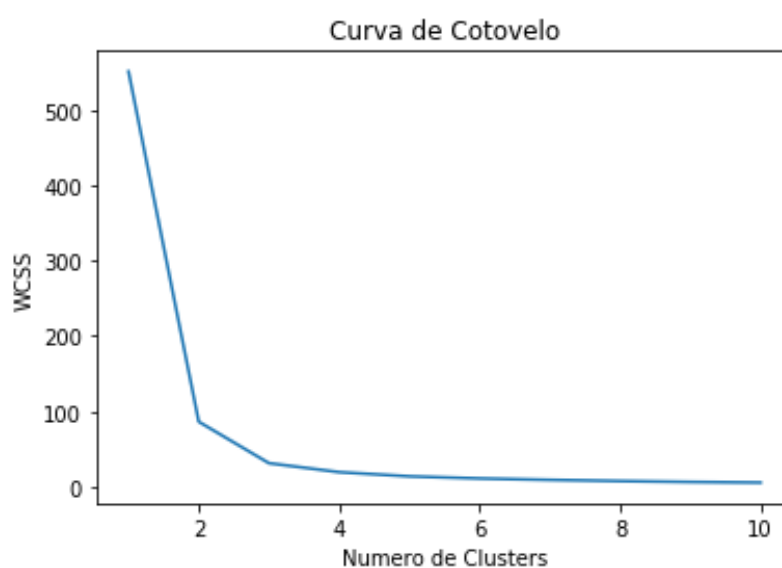


Figura 8. Curva do cotovelo.

Fonte: Kaggle, 2019.

A silhueta simula o quanto perto as observações mais semelhantes estão umas das outras enquanto são separadas de outros clusters distintos. Visa uma melhor coesão dos clusters, bem como um maior isolamento entre os clusters.

O valor do índice de silhueta varia no intervalo de -1 a 1 e é calculado para cada elemento do cluster, revelando quais elementos estão bem situados no mesmo e quais seriam melhor situados em outro cluster. (FADEL et al. 2014).

Quando o valor da silhueta está próximo de 1 indica que os elementos foram classificados de maneira adequada, quando o valor da silhueta está próximo de -1 indica que os elementos foram mal classificados e se o valor da silhueta for próximo de 0 indica que o elemento está situado em uma situação intermediária entre os clusters. Usando a distância euclidiana pode-se medir também pela silhueta a distância entre as observações (ROUSSEEUW, 1987), conforme é visto na Figura 9.

Exemplificando:

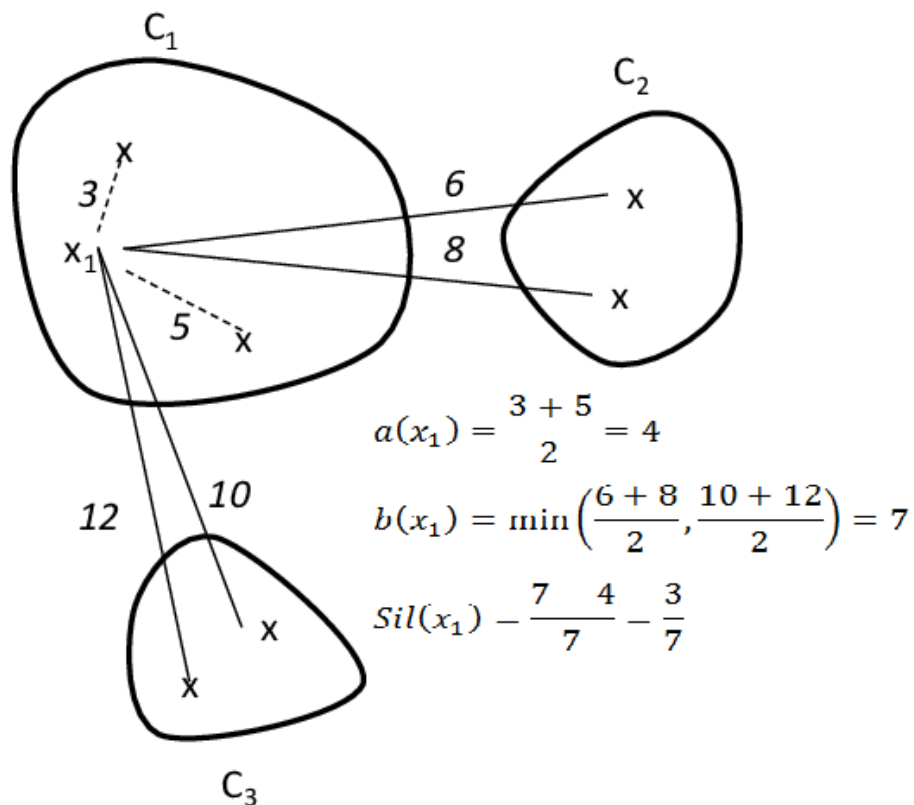


Figura 9. Valor da Silhueta.

Fonte: Kaggle, 2019.

- *Model Tuning* (ajuste do modelo): Não há garantia de que os pontos de acesso começarão nas posições mais favoráveis. Como os centroides iniciais, que foram atribuídos aleatoriamente, começaram em uma posição subótima, podemos ter um resultado que convergiu, mas não resultou no melhor resultado. O algoritmo foi executado várias vezes para corrigir esse problema, que é controlado pelo parâmetro “n_init”, que é definido como 10 por padrão. Isso significa que o algoritmo K-Means é executado dez vezes, cada vez com diferentes posições iniciais (ANASTACIO, 2020).

E preciso tomar cuidado ao usar clusterização com dados que não foram padronizados. Como o K-means emprega a distância euclidiana para estimar a distância, ele é particularmente sensível à escala dos dados. Como resultado, o dimensionamento deve sempre ser feito antes de empregar K-means. Outro problema é que com um grande número de observações, o K-Means pode ser bastante lento. Como resultado, em vez de usar a coleção completa de dados, às vezes é empregada uma amostra por motivos de desempenho. Se os dados contiverem ruído ou valores discrepantes, o agrupamento poderá produzir resultados indesejáveis. Embora ter um grande número de clusters (K) ajude a reduzir erros, já que a distância das instâncias dentro do cluster tende a diminuir, também aumenta o risco de *overfitting*, que é quando o modelo tem um desempenho excelente quando se utiliza uma base de treino, mas quando se utiliza os dados de testes o resultado é ruim, ou seja, o modelo não encontra padrões para instâncias ainda não conhecidas. Além disso, o *clustering* tem a vantagem de não exigir dados pré-ordenados (ANASTACIO, 2020).

2.2 Trabalhos sobre Modelagem e Recomendação de Usuários

Segundo Neves (2018), as *Delay-tolerant networking* (DTNs) surgiram como uma solução para comunicação em cenários onde a Internet não tem suas premissas básicas atendidas para um bom funcionamento. As DTNs dependem diretamente de seus nós para um bom desempenho, pois utilizam a mobilidade dos nós para enviar mensagens aos seus destinos. No entanto, devido a fatores como economia de recursos, falta de interesse na mensagem ou simplesmente a negação de colaboração, isso afeta negativamente o desempenho da rede. Dessa forma, é fundamental considerar os fatores sociais, estendidos dos usuários aos nós da rede, para que se encontre a melhor estratégia de encaminhamento, aumentando as chances de entrega das mensagens. Neves (2018) realizou um estudo onde é proposto um novo protocolo para a disseminação de mensagens em redes DTN, utilizando os

interesses dos nós da rede para a formação do nível de passagem de mensagens. Segundo o autor, esses níveis são formados agrupando os nós através de técnicas de aprendizado de máquina, K-MEANS e em clusters, de acordo com o nível de interesse dos nós pela mensagem gerada, direta ou indiretamente, passando a mensagem pelos grupos formados até o destinatário do nó. Os resultados deste estudo apontam que a proposta é promissora, apresentando resultados superiores aos protocolos bem qualificados na literatura.

Segundo Souza (2017), o crescente uso das redes sociais fez com que a quantidade de informações que são compartilhadas diariamente (opiniões pessoais, notícias e eventos do tipo natural ou social) aumentasse, tornando estas redes uma fonte de informação sobre eventos. No Brasil, acontecimentos recentes, como a Operação Lava Jato conduzida pela polícia Federal e o processo de impeachment da Presidente são exemplos de eventos que geraram repercussões nos meios de comunicação. Esses fatos marcantes, podem ser utilizados para caracterizar quais são os acontecimentos relevantes de um determinado evento. Nas redes sociais, esses acontecimentos geram discussões, compartilhamentos e novas notícias, onde informações relevantes acabam tendo mais repercussões. Contudo, processar uma elevada massa de dados a fim de eliminar ruídos e reconhecer informações torna-se muito custoso.

Neste contexto, este trabalho teve como objetivo demonstrar uma abordagem para caracterização de informações relevantes de eventos, através da extração de tópicos em dados compartilhados no Twitter, onde avaliamos o desempenho de três métodos de aprendizagem de máquina (K-means, Latent Dirichlet Allocation - LDA e Non-Negative Matrix Factorization - NMF) usados para extrair tópicos sobre as bases de dados da Operação Lava Jato e do processo de impeachment da presidente do Brasil em duas arquiteturas de pré-processamento diferentes (tradicional e com reconhecimento de entidade), demonstrando que é possível utilizar uma rede social como fonte de dados para descobrir os tópicos relevantes através do sensoriamento de usuários que observam um evento. Nos experimentos de Souza (2017), foi observado que as técnicas de pré-processamento têm influência direta sobre o resultado da extração de tópicos. Além disso, observou-se que a técnica silhueta ajudou a encontrar o melhor valor de clusters para uma determinada amostra de dados. Nos resultados o NMF apresentou o melhor desempenho nas duas bases de dados, tanto na tarefa de extração de tópicos quanto no tempo de execução (SOUZA, 2017).

O artigo de Mallmann et al. (2018) expõe um trabalho de desenvolvimento que utiliza técnicas de *Machine Learning* para realizar a detecção de cibercrime em mensagens publicadas na rede social Twitter. As mensagens são pré-processadas, constrói-se um

dicionário, e no uso do software WEKA são aplicadas técnicas de agrupamento e de classificação (K-means, SVM, DT e NB) para detecção de cibercrimes. Em resultados experimentais afirma-se que o uso de *Machine Learning* foi essencial para o sucesso deste trabalho, sendo que o classificador SVM apresentou 98,77% de acertos na classificação de cibercrime.

Desenvolvimento do Sistema Rede Gamer

A proposta do trabalho é desenvolver uma rede social focada no público *gamer*, onde os usuários possam interagir entre si no envio de mensagens e dicas sobre jogos, socializando suas percepções sobre os jogos com outras pessoas. Denominada de “Rede Gamer”, a rede social visa criar um ambiente onde seus usuários possam se conectar com jogadores de regiões próximas, preferivelmente com aqueles que gostam dos mesmos estilos de jogo, de forma a permitir a criação de redes de contatos que possam, eventualmente, serem extrapoladas para encontros presenciais.

3.1 Análise e Projeto

Aqui se encontra o diagrama de caso de uso do projeto, identificando os atores do sistema e seus respectivos casos de uso, se encontra também o modelo de dados do sistema.

O diagrama de classes é a parte central da Linguagem de Modelagem Unificada ou, em inglês, *Unified Modeling Language* (UML). Ela representa os principais propósitos da UML e tem a função de separar os elementos de design da codificação do sistema (BLAHA, M.; RUMBAUGH, 2006).

Essa linguagem geralmente é usada por engenheiros para documentar a arquitetura de software. Ele ajuda a modelar vários subconjuntos de diagramas, incluindo diagramas comportamentais, de interação e estruturais (BLAHA, M.; RUMBAUGH, 2006).

O Diagrama de Casos de Uso mostrado na Figura 10 apresenta uma visão geral das funcionalidades do sistema Rede Gamer e os atores que desempenham cada uma delas (visitante, usuário e administrador).

Cada ator tem um caso de uso e o que ele faz dentro do sistema, que são as ações descritas dentro destes círculos.

A única coisa que o visitante faz inicialmente é o cadastro e o *login*. Ou seja, a partir do cadastro, ele se torna usuário que tem acesso aos jogos através do *login*, e todas as demais ações descritas nos círculos da figura abaixo.

O usuário tem acesso ao index, visualiza sua lista de jogos, acessa seu perfil, visualiza perfil de outro usuário etc. Em resumo, cada esfera desta figura (abaixo) representa um caso de uso, ou o que o ator pode realizar dentro do sistema.

O criador do sistema é o próprio administrador, o responsável por acionar novo jogo ou remover os jogos cadastrados no sistema.

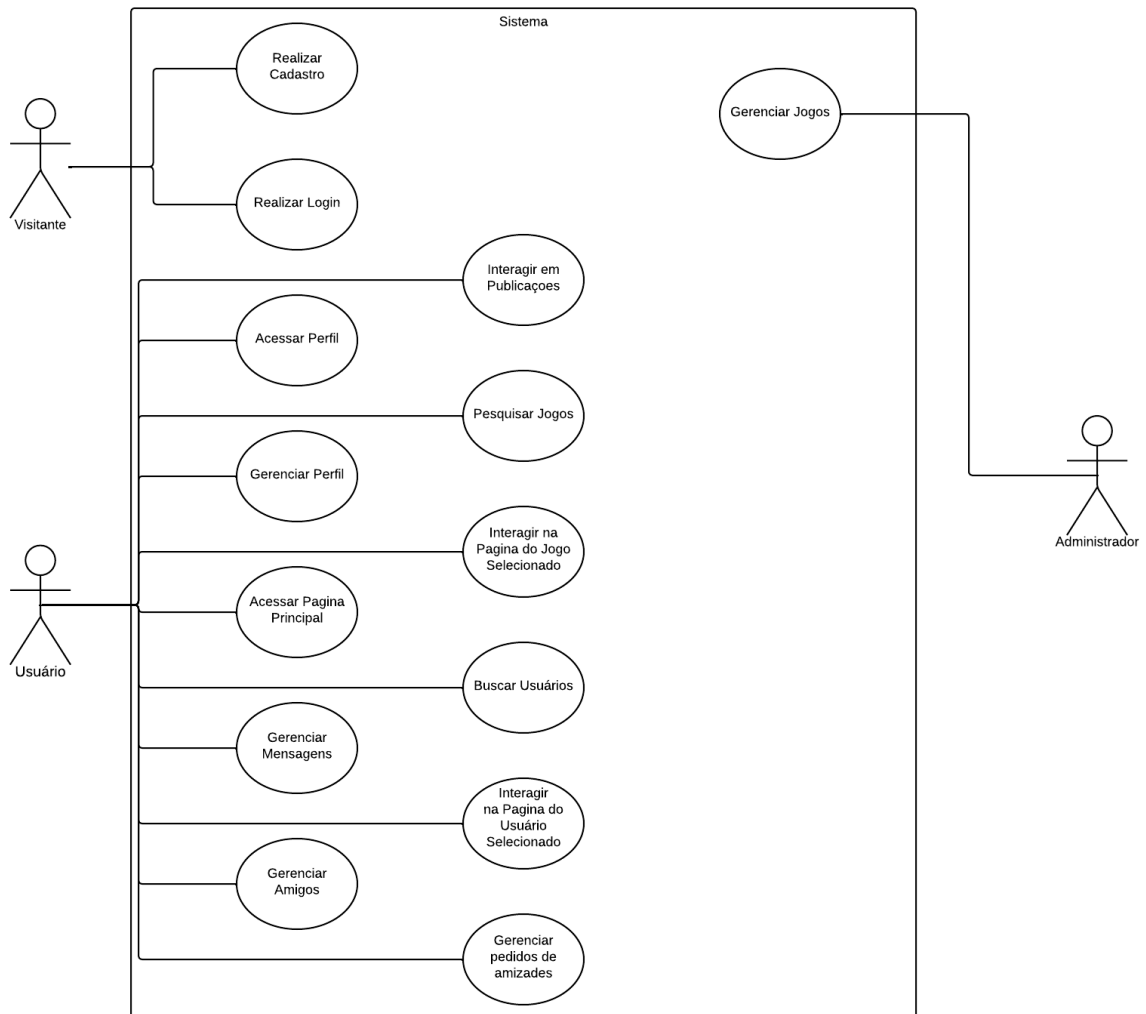


Figura 10. Diagrama Caso de Uso.

Fonte: Autor, 2022.

O Quadro 1 trata da Especificação dos atores do sistema, encontra-se a identificação dos atores, onde são enumerados os atores, o nome do ator (visitante, usuário e administrador) e a descrição das ações de cada um destes atores dentro do sistema.

Quadro 1. Identificação dos Atores

Ator	Nome do ator	Descrição ator
Ator-01	Visitante	Qualquer pessoa que visita o sistema, sem estar autenticado por login/senha
Ator-02	Usuário	Qualquer pessoa que já tenha efetuado o cadastro no sistema
Ator-03	Administrador	A pessoa responsável por cadastrar os jogos e gerenciar o Sistema

Fonte: Autor, 2022.

O Quadro 2 trata da Especificação dos casos de uso do sistema, onde encontra-se a identificação dos Casos de Uso que são enumerados, o nome do Caso de Uso e a descrição de cada Caso de Uso. O que é preciso para aquele Caso de Uso acontecer? Por exemplo, no Caso de Uso 1 Efetuar Login, a descrição é “Autenticação de usuários cadastrados no sistema, permitindo acesso na parte restrita do sistema”, e demais especificações conforme lê-se no quadro 2 “Identificação dos Casos de Uso” do quadro abaixo.

Quadro 2. Identificação dos Casos de Uso

#uc	Nome UC	Descrição UC
UC-01	Realizar Login	Autenticação de usuários cadastrados no sistema, permitindo acesso na parte restrita do sistema
UC-02	Realizar Cadastro	Usuário preenche o formulário com seus dados que são salvos no banco de dados permitindo acesso na parte restrita do sistema
UC-03	Acessar Perfil	O usuário consegue visualizar seu perfil e de outros usuários da rede onde será mostrado suas publicações seus jogos e seus amigos
UC-04	Gerenciar Perfil	O usuário pode alterar seus dados que foram salvos no início do cadastrado assim como alterar ou excluir suas publicações
UC-05	Acessar Página Principal	O usuário tem acesso a página principal onde é possível ver suas sugestões de amizades e publicações de amigos
UC-06	Gerenciar Mensagens	O usuário tem acesso a sua página de mensagem onde pode enviar ou visualizar suas mensagens com seus amigos
UC-07	Gerenciar Amigos	O usuário pode visualizar sua lista de amigos assim como remover usuários que já sejam amigos
UC-08	Interagir Em Publicações	O usuário pode curtir ou comentar suas publicações e de seus amigos
UC-09	Pesquisar Jogos	O usuário através de um campo de pesquisa pode buscar pelo nome dos jogos cadastrado no sistema
UC-10	Interagir na Página do Jogo Selecionado	O usuário pode adicionar o jogo como favorito ou como jogo jogável assim como comentar na página do jogo selecionado
UC-11	Buscar Usuário	O usuário através de um campo de pesquisa pode buscar usuários cadastrado no sistema
UC-12	Interagir na Página do Usuário Selecionado	O usuário pode enviar o pedido de amizade, ver suas publicações, suas informações e jogos marcados como favorito do usuário selecionado
UC-13	Gerenciar Pedidos de Amizades	O usuário pode aceitar ou remover pedidos de amizades de outros usuários
UC-14	Gerenciar Jogos	O Administrador da aplicação tem a possibilidade adicionar e remover as páginas de jogo no sistema

Fonte: Autor, 2022.

Na Figura 11, localiza-se o Modelo de Dados do sistema as tabelas de (usuários, publicações, mensagens etc.), e a partir de cada uma delas, seus respectivos atributos (no usuário, por exemplo, tem-se: nome, id público, e-mail, senha, foto, cidade/estado, senha, data de nascimento., etc.) e na frente de cada atributo encontram-se seu tipo de dados, “nome” um *varchar*, um campo que contém letras e números e pode conter até 100 caracteres (que se encontra entre parênteses). Outro exemplo é o “Id público”, um *varchar* que pode conter letras e números com tamanho de até 50 caracteres.

As tabelas possuem relacionamentos com outras tabelas, indicados por uma linha que as conectam. Exemplificando, o usuário é ligado com “Municípios”, com “Jogosadddfavoritos”, com “Mensagens”, e com “Publicação”, que por sua vez interliga-se com “Reações” e “ComentárioPublicações”. Pois é o usuário quem faz uma publicação, é quem vai adicionar um jogo como favorito.

No usuário há um campo chamado “Id Público”, que possui uma chave e significa “chave-primária”, ou seja, significa dizer que nenhum usuário terá um ‘Id Público’ igual, pois através do ‘Id Público’, o administrador tem acesso à mensagem enviado pelo usuário, o tipo de publicações etc.

No item “Usuário” há no campo “Cidade” que é ligada à tabela do município que contém o campo, o código do IBGE, o nome (que é um *varchar* de no máximo 100 caracteres), a latitude, a longitude, se o município é ou não capital daquele Estado.

No item “publicação”, o administrador tem acesso à chave primária, onde nenhuma publicação possui o mesmo Id da outra, e consegue-se ver quem é o usuário que fez a publicação, quem reagiu à publicação. O mesmo que acontece a parte do item “Comentário”.

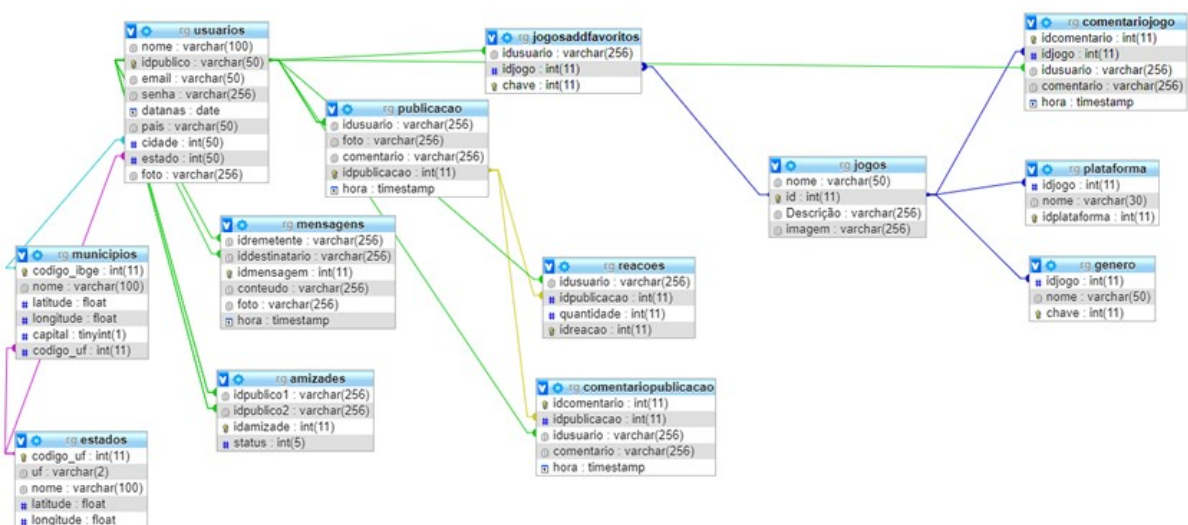


Figura 11. Modelo de Dados.

Fonte: Autor, 2022.

3.2 Protótipo Funcional

Em se tratando da parte da interface, a Figura 12 representa a área de “Login” que possui apenas dois campos onde o usuário digita o E-mail e a senha e clica em “Entrar”, que o sistema vai validar se aquela senha está correta ou se aquele E-mail já tem cadastro no sistema. Na página de Login, pode-se fazer o cadastro clicando no item “Clica aqui” em cor azul, conforme mostrado na Figura 12.



REDE GAMER

E-mail
Senha

Entrar

Fazer cadastro? [Clique aqui](#)

Figura 12. Interface de Login.
Fonte: Autor, 2022.

Figura 13 representa a “Tela de cadastro”, onde considerou-se possuir apenas os atributos relevante apenas o nome, o Id Público (como ele quer ser chamado dentro da rede), e-mail e senha.

O usuário tem a opção de inserir foto no campo do perfil, se ele achar necessário no primeiro momento, ou posteriormente. O usuário informa a data de nascimento com intuito de unir pessoas com faixas etárias próximas.

Também se considerou relevante atribuir aos usuários nesta tela: o país, o Estado e a cidade, com intenção de aproximá-los por região.



[← Voltar](#)

Crie sua conta na RG

Nome ID publico

E-mail

Senha Confirmar Senha

Adicionar Foto:
Escolher arquivo Nenhum arquivo selecionado

Data de nascimento:
dd/mm/aaaa

Nacionalidade:
País

-----Estado----- -----Cidade-----

Figura 13. Tela de cadastro.
Fonte: Autor, 2022.

A Figura 14 mostra a “Tela de Amigos” do usuário que se encontra logado/cadastrado. Nesta tela, são exibidos todos os amigos daquele usuário que já estão cadastrados no banco de dados. Exemplificando, a Figura 14 demonstra o usuário com três amigos denominados “alguém”, “naldido” e “andressa”. Essa parte onde a denominação aparece em fonte maior trata-se do Id Público da pessoa, o nome como ele quer ser chamado (“alguém”, “naldido” e “andressa”). Abaixo, em fonte menor, trata-se do nome real da pessoa. Nos campos à direita também são publicados a cidade e o Estado onde a pessoa mora.

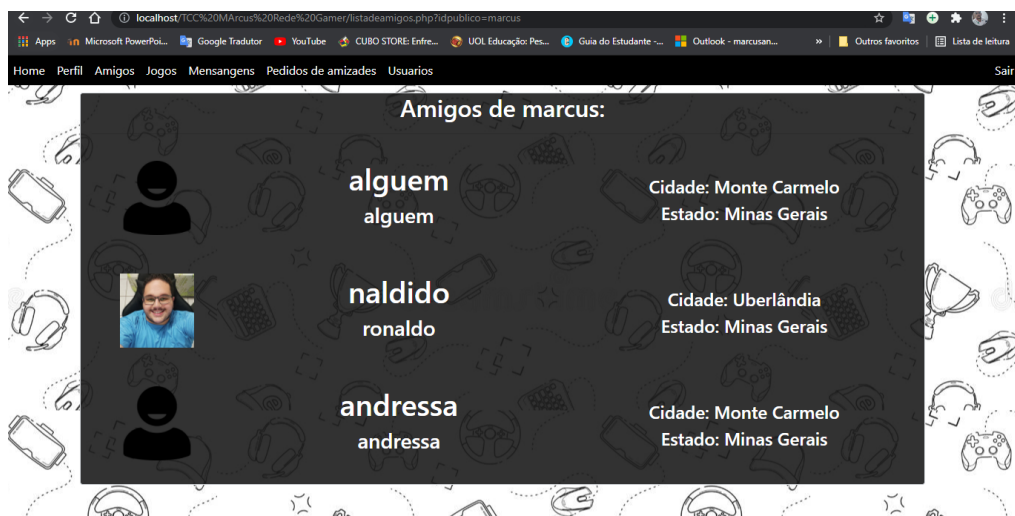


Figura 14. Tela de amigos do usuário logado.
Fonte: Print da tela realizado pelo autor, 2022.

A Figura 15 apresenta a “Tela de jogos cadastrados no sistema”, isto é, todos os jogos cadastrados no sistema. Neste espaço aparece uma imagem do jogo e à sua direita o nome do jogo. Ali o usuário pode entrar e obter qualquer informação acerca do jogo em questão.

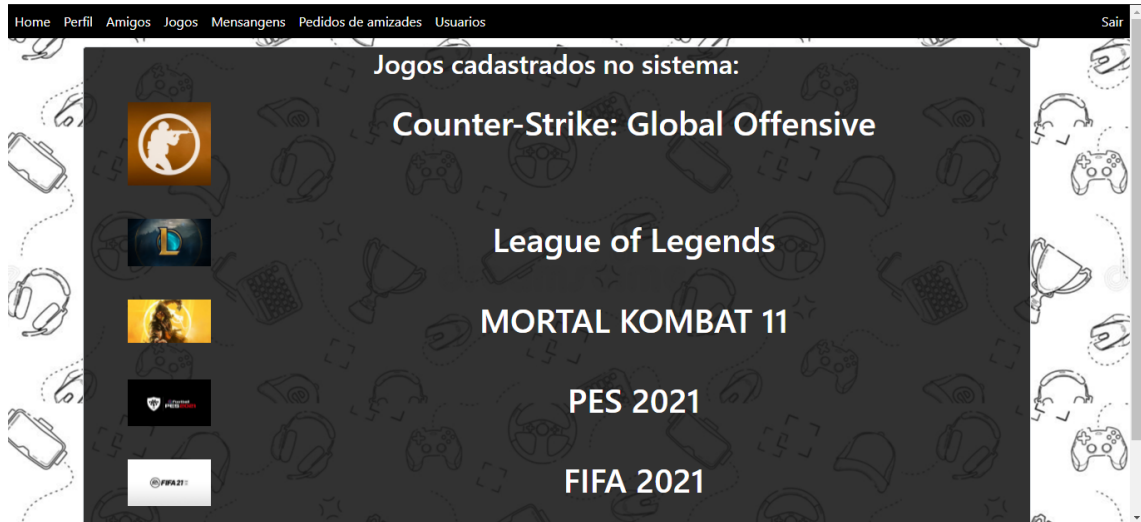


Figura 15. Tela de jogos cadastrados no sistema.
Fonte: Autor, 2022.

A Figura 16 apresenta a “Tela de um jogo selecionado”, ou seja, o usuário selecionou o jogo o qual ele quer ter acesso a informações. No *print* da Figura 16, já se encontra selecionado o jogo “Counter-Strike” como favorito, mostra que o usuário joga este jogo, por exemplo.

Na página do jogo encontra-se os gêneros (neste caso é um FPS) e as plataformas (o jogo roda no Microsoft, Xbox One, Xbox 360, macOS, PlayStation 3, Linux). E em seguida, tem-se uma descrição do jogo, um texto breve sobre como se joga o jogo, os objetivos etc.

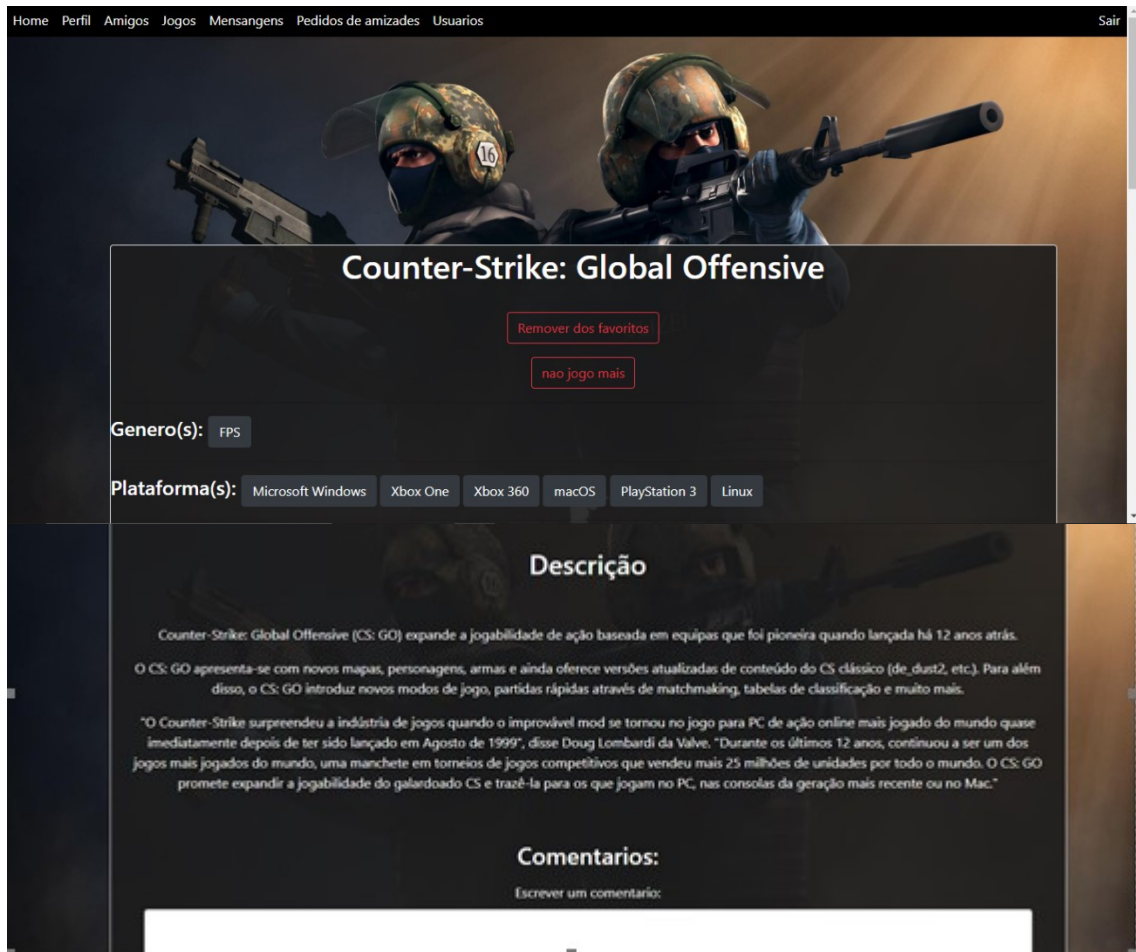


Figura 16. Tela de um jogo selecionado.
Fonte: Autor, 2022.

Em seguida, tem-se na Figura 17 abaixo a “Tela Principal” onde é possível ver o usuário que está logado, o nome dele, a cidade dele. No meio, há um campo onde o usuário pode publicar um novo conteúdo, selecionar a foto, escrever alguma coisa, ou apagar. E no canto direito tem a opção de sugestão de amizade que apareceu para aquele usuário que está logado.

Caso tenha um usuário novo em uma região muito distante dos demais usuários que já estão cadastrados no sistema e não tenha colocado nenhum jogo como favorito ou jogo que gosta de jogar, é provável que o usuário não receberá nenhuma sugestão de amizade inicialmente, pois, considerando a dinâmica do algoritmo de agrupamento, possivelmente o usuário será colocado em um *cluster* separado, esperando novos usuários que estejam em regiões próximas ou que gostem do mesmo estilo de jogo deste usuário.



Figura 17. Tela Principal.
Fonte: Autor, 2022.

A Figura 18 abaixo representa a “Tela de um usuário selecionado” que nada mais é que uma espécie de um perfil do Facebook, por exemplo. Contendo os dados do usuário (nome, Id, cidade, Estado etc.), onde pode ser publicado um novo conteúdo. Se o usuário estiver logado, os dados do perfil podem ser modificados.

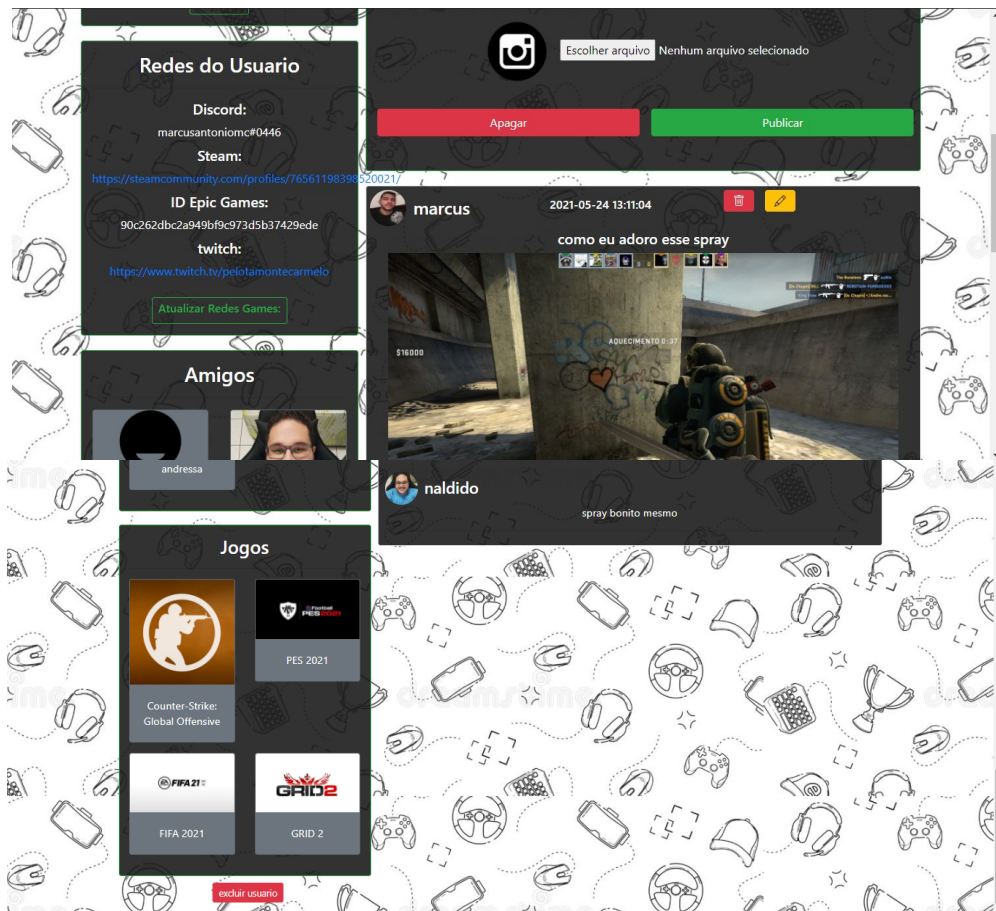


Figura 18. Tela de um usuário selecionado.
Fonte: Autor, 2022.

Na “Tela de solicitações de amizades”, conforme demonstra a Figura 19, onde pode ser feita a solicitação de amizade no item “Pedido de Amizade” e a pessoa pode aceitar o recusar o pedido.

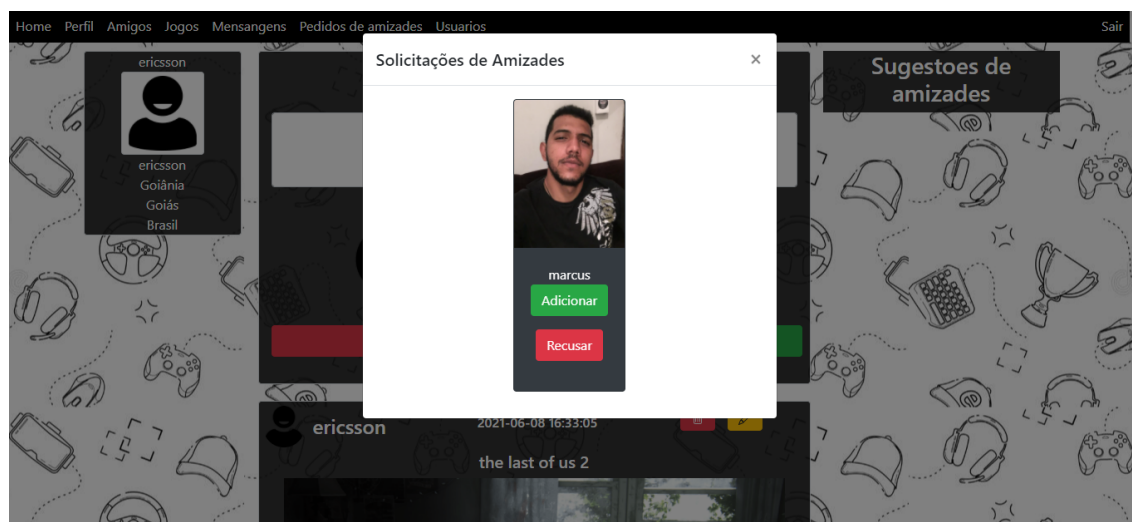


Figura 19. Tela de solicitações de amizades.
Fonte: Autor, 2022.

Experimentos e Análise dos Resultados

A ideia de usar o algoritmo de clusterização para as sugestões de amizades se dá ao fato de que os usuários poderão conhecer novas pessoas da mesma cidade ou regiões próximas de sua cidade onde poderão sair e se interagir fora do mundo virtual. Se um usuário se logar na rede e não receber nenhuma sugestão de amizade basta apenas recarregar a página, pois o algoritmo é executado e atualizado toda vez que o usuário entra na página principal.

Para a implementação da prova de conceito do sistema Rede Gamer, fez-se necessário desenvolver a primeira versão da rede social, e os atributos que foram considerados como relevante para a recomendação de amizade foram a quantidade de jogos em cada gênero que o usuário adicionou como favorito e a latitude e longitude da cidade e estado que o usuário está cadastrado no sistema.

4.1 Pesquisa de Opinião

Após a ideia da criação da rede social foi feita uma pesquisa de opinião para saber se o projeto seria bem aceito para os possíveis usuários do sistema. As respostas do formulário completo se encontram no apêndice A. Grande parte das pessoas que responderam a pesquisa são estudantes que gostam de jogar algum jogo online, como mostram os gráficos da Figura 20.

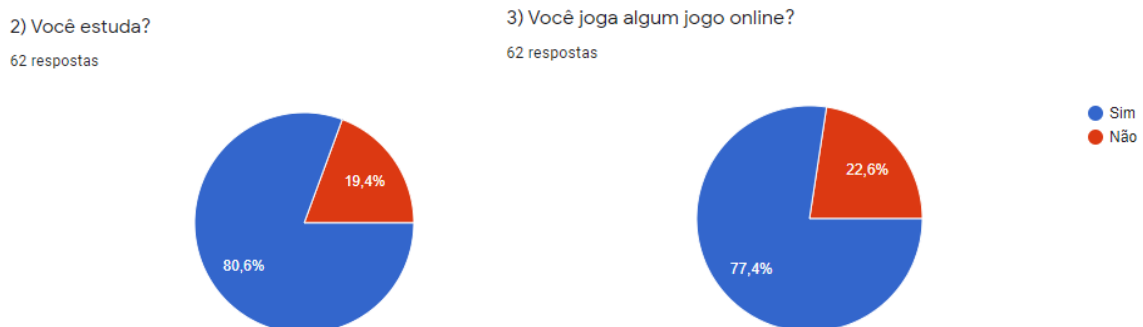


Figura 20. Resposta formulário 1.

Fonte: Print da tela do formulário feita pelo autor, 2022.

Foi pedido para que as pessoas respondessem algumas afirmativas marcando 1- discordo totalmente 2- discordo parcialmente 3- não sei opinar 4- concordo parcialmente 5- concordo totalmente.

A pesquisa mostrou que a maioria das pessoas não gostam de jogar sozinhas conforme mostra a Figura 21.

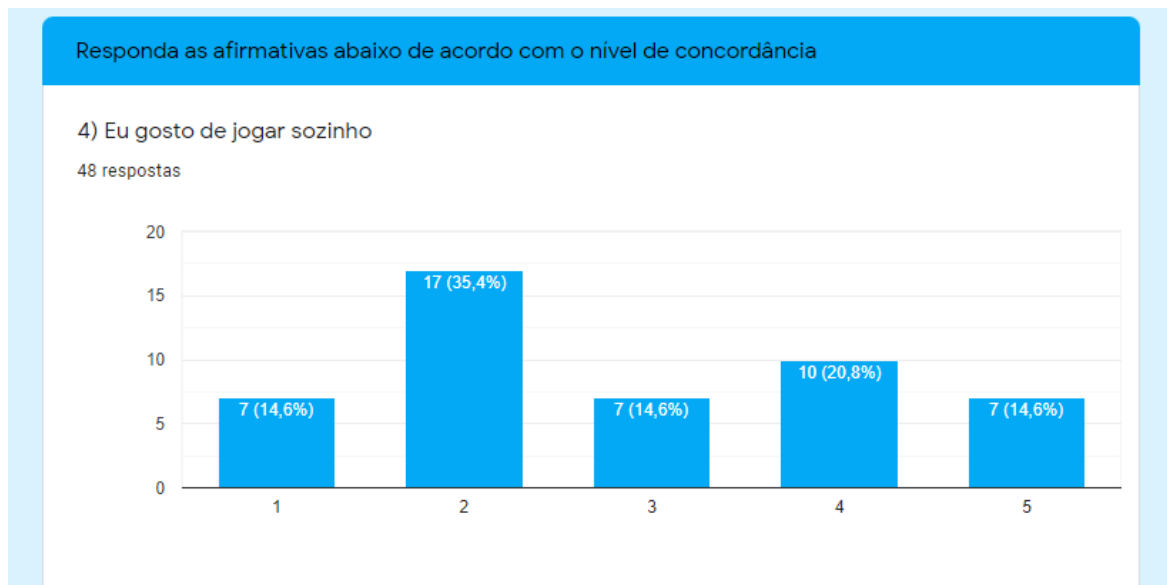


Figura 21. Resposta formulário 2.
Fonte: Print da tela do formulário feita pelo autor, 2022.

A pesquisa revelou que as pessoas têm dificuldade em encontrar informações sobre os jogos que elas gostam em um lugar específico, conforme mostrado na Figura 22.

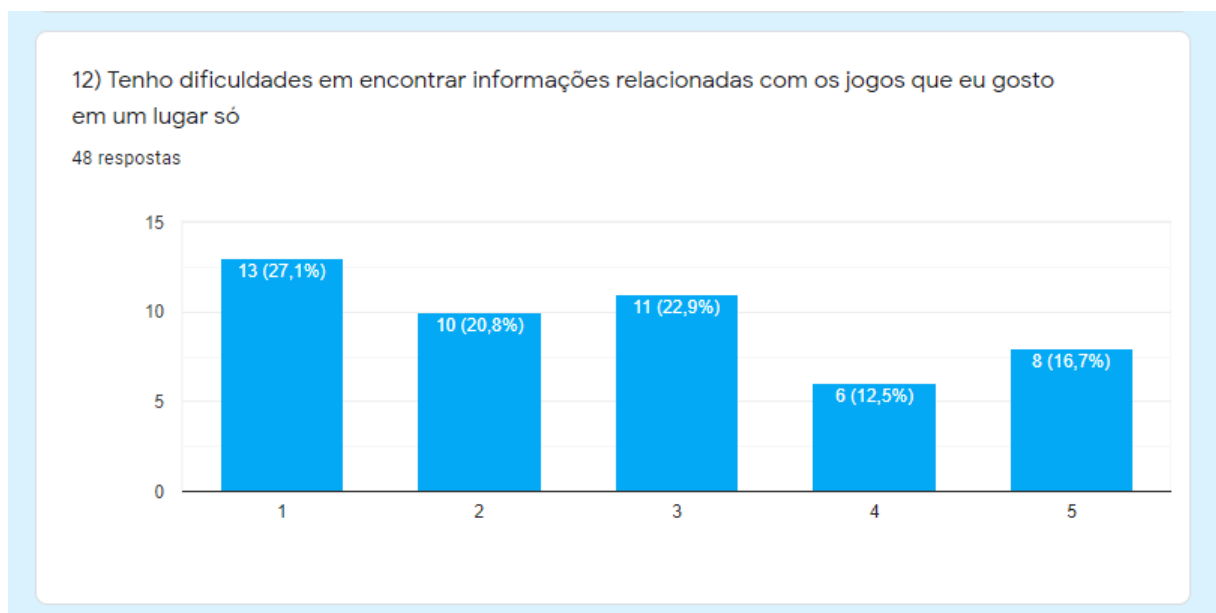


Figura 22. Resposta formulário 3.
Fonte: Print da tela do formulário feita pelo autor, 2022.

A pesquisa também mostrou que as pessoas gostariam de conhecer pessoas da mesma cidade e que gostem de jogar os mesmos jogos, como mostra a Figura 23.

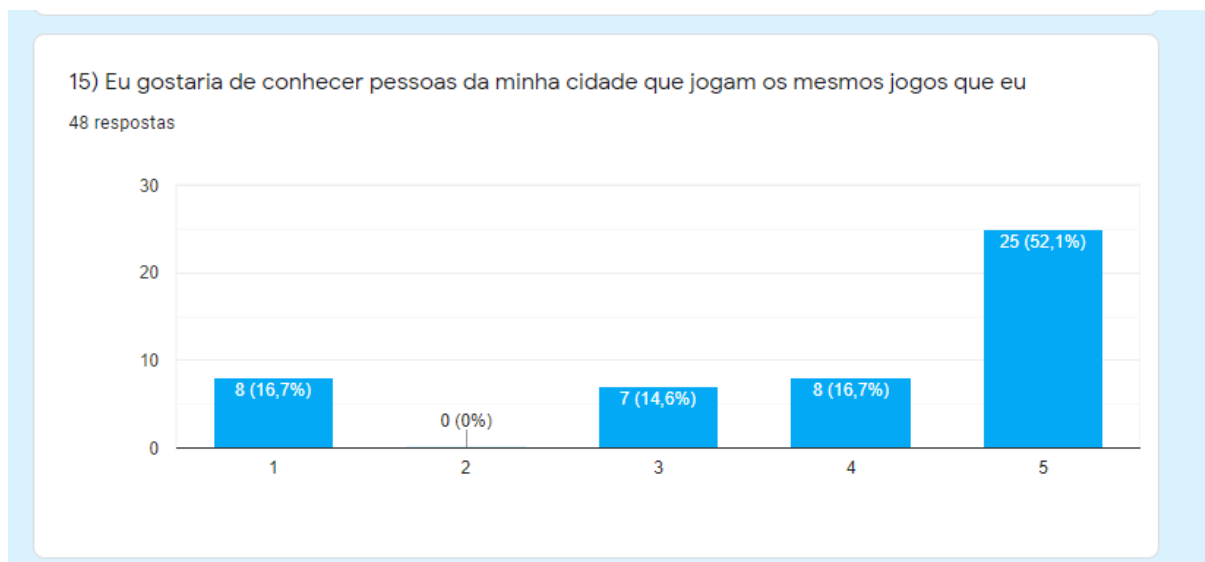


Figura 23. Resposta formulário 4.
Fonte: Print da tela do formulário feita pelo autor, 2022.

4.2 Método para a Avaliação

O sistema Rede Gamer ainda não foi lançado oficialmente na Internet, pois encontra-se sendo experimentado com objetivo de avaliar o seu funcionamento e a satisfação dos usuários.

Embora ainda não tenha sido ainda lançado, foi feita uma divulgação nas redes sociais do próprio programador – o que alcançou cerca de 12 usuários espontâneos (que conheceram o sistema por meio do anúncio), e outros três amigos pessoais do autor do trabalho.

Isto é, o sistema encontra-se em desenvolvimento, sobretudo com uma testagem a partir do uso de 515 pessoas cadastradas, sendo 15 delas pessoas reais convidadas a fazerem parte desta testagem e os demais – cerca de 500 robôs (usuários falsos, criados pelo desenvolvedor do sistema) – nos experimentos sobre a sugestão de amizade (quanto aos resultados esperados).

A Figura 24 aponta os dados cadastrais de alguns destes usuários.

<input checked="" type="checkbox"/>				marcão	1010	marcaopena@gmail.com	6d071901727aec1ba6d8e2497ef5b709	1990-09-22	brasil	3170107
<input checked="" type="checkbox"/>				123123	123123	123123@gmail.com	4297f44b13955235245b2497399d7a93	1996-04-02	brasil	3170107
<input checked="" type="checkbox"/>				Pedro Farias	123456	pedrobfarias@gmail.com	f9ead0b3b5bd3dbb36e6e6ac2473f142	1995-02-25	Brasil	4106902
<input checked="" type="checkbox"/>				tulio araujo	2323	tulioaraujo23@hotmail.com	fb1628b1a099c02e8701dd548819f7fa	1997-08-22	Brasil	3143104
<input checked="" type="checkbox"/>				andressa	ac.rocha	andressaaac.rocha@gmail.com	e10adc3949ba59abbe56e057f20f883e	1996-07-14	Brasil	3100104
<input checked="" type="checkbox"/>				DanieleOliveira	daniOliveira	danieleoliveira@ufu.br	e10adc3949ba59abbe56e057f20f883e	1984-07-02	Brasil	5300108
<input checked="" type="checkbox"/>				Rafael Luan	edorafa	edorafa@hotmail.com	f886b914d22dab1f7f5866cb1ce3c52	1992-05-12	Brasil	3143104
<input checked="" type="checkbox"/>				Marcus Antonio Rufino Ribeiro	marcusasdasantoniomc@gmaiaa.coml.com	marcusantoniomasqdasdac@gmail.com	4297f44b13955235245b2497399d7a93	1996-04-02	brasil	3100302
<input checked="" type="checkbox"/>				123123	marcu123123santoniomc@gmail.com	123121231233@gmail.com	4297f44b13955235245b2497399d7a93	1196-04-02	123123	2100055
<input checked="" type="checkbox"/>				Marcus Antonio Rufino Ribeiro	marcusasdasantoniomc@gmaiaa.coml.com	marcusantoniomasdasdac@gmail.com	49c167d7cd66dc64a474c261860ba50f	1996-04-02	brasil	3100302
<input checked="" type="checkbox"/>				Marcus Antonio Rufino Ribeiro	marcusasdasantoniomc@gmail.com	marcusantoniomasdac@gmail.com	49c167d7cd66dc64a474c261860ba50f	1996-04-02	brasil	3100302
<input checked="" type="checkbox"/>				testemisterb	misterbean	misterbean@gmail.com	e10adc3949ba59abbe56e057f20f883e	1988-12-05	Brasil	4100103
<input checked="" type="checkbox"/>				Marcus Antonio	mmmm	mmarcusantoniomc@gmail.com	49c167d7cd66dc64a474c261860ba50f	1996-04-02	brasil	2400109
<input checked="" type="checkbox"/>				Marcus Antonio Rufino Ribeiro	Pelota	marcusantoniomc@gmail.com	49c167d7cd66dc64a474c261860ba50f	1996-04-02	brasil	3143104
<input checked="" type="checkbox"/>				teste	testebean	testebean@gmail.com	e10adc3949ba59abbe56e057f20f883e	1994-05-07	Brasil	3100104

Figura 24. Trecho da lista usuários cadastrados no sistema Rede Gamer.

Fonte: Print da tela feita pelo autor, 2022.

4.3 Experimentos

É importante ressaltar que os testes realizados são com uma base de dados sintéticos, diferentes em cada experimento com o objetivo de explorar as dimensões e decidir quais serão os melhores atributos para utilizar na parte de sugestão de amizade.

Para os referidos experimentos, já se constatou que pode ser usado o método K-means com o método do cotovelo (algoritmo *elbow*),

Uma maneira de modelar a sugestão de amizade seria o algoritmo do *dbscan*, mais como seu principal objetivo e encontrar *outline* foi descartada essa possibilidade, pois não seria legal para os usuários se eles não recebessem nenhuma sugestão de amizade.

Como o algoritmo do *k-means* precisa passar um valor de N para criar a quantidade de *clusters* a solução para gerar esse número ideal foi utilizar o algoritmo de *Elbow*.

Quanto às execuções dos algoritmos de *clustering*, apresentamos 3 experimentos.

- **Experimento 1:** Para esta parte do experimento, foi acessado o banco de dados e aplicado o nome do usuário em um *array*, o valor da latitude e longitude da cidade e do estado que o Usuário se cadastrou no banco de dados. Depois, foi colocado o algoritmo do *K-means* para fazer a divisão daquele *array* informando o que o administrador quer: que seja dividido em 7 clusters, como mostra a Figura 25.

```

$sql=mysqli_query($conexao,"SELECT idpublico, cidade, estado FROM usuarios")or
die("erro ao selecionar");
while($dados = mysqli_fetch_array($sql)){
    $codestado=$dados['estado'];
    $codcidade=$dados['cidade'];
    $sqleestado=mysqli_query($conexao,"SELECT latitude,longitude FROM estados
Where codigo_uf=$codestado")or die("erro ao selecionar");
    while($dadoseestado = mysqli_fetch_array($sqleestado)){
        $estadolat=$dadoseestado['latitude'];
        $estadolong=$dadoseestado['longitude'];
    }
    $sqlcidade=mysqli_query($conexao,"SELECT latitude,longitude FROM
municipios Where codigo_ibge=$codcidade")or die("erro ao selecionar");
    while($dadoscidade = mysqli_fetch_array($sqlcidade)){
        $cidadelat=$dadoscidade['latitude'];
        $cidadelong=$dadoscidade['longitude'];
    }
    $users[$dados['idpublico']] = array($estadolat,$estadolong,$cidadelat,$cidadelon
g);
}
$samples=$users;
$kmeans = new KMeans(7);
$ola=$kmeans->cluster($samples);
var_dump($ola);

```

Figura 25. Experimento 1.

Fonte: Print da tela feita pelo autor, 2022.

O algoritmo funcionou colocando usuários de estados próximos no mesmo grupo, mas não foi utilizado pois o objetivo do administrador não era apenas juntar as pessoas da mesma região.

- **Experimento 2:** Foram criados robôs que simulariam os usuários contendo o nome, valor do código do estado que foi cadastrado no banco de dados e valores que representariam o gênero do jogo que o usuário adicionou como favorito. Depois, é colocado o algoritmo do *K-means* para fazer a divisão daquele *array* informando a quantidade de *clusters*, sendo neste teste utilizado o valor 10, como mostra a Figura 26.

```

for($i=0;$i<100;$i++){
    if($i%2==0){
        $users['AA'].$i=array(3143104,31,1,0,0,0,0);
    }
    else{
        $users['BB'].$i=array(3143104,31,0,1,0,0,0);
    }
}
.
.aqui continua a criação de robôs simulando os usuários
.
for($i=0;$i<100;$i++){
    if($i%2==0){
        $users['UU'.($i+1000)]=array(3143104,31,1,0,1,0,1);
    }
    else{
        $users['vv'.($i+1000)]=array(3143104,31,1,0,0,1,1);
    }
}
$samples=$users;
$kmeans = new KMeans(10);
$ola=$kmeans->cluster($samples);

```

Figura 26. Experimento 2.
Fonte: Print da tela feita pelo autor, 2022.

O algoritmo funcionou colocando usuários de estados próximos e que gostavam do mesmo gênero de jogos no mesmo grupo, mas não foi utilizado por desconhecimento acerca de qual seria o número ideal para divisão de grupos.

- **Experimento 3:** Através do algoritmo do cotovelo foi possível obter o número ideal de cluster para este sistema, que é igual a 5 se pegar o valor da latitude e da longitude da cidade e do estado, ou 4 se pegar o valor do id da cidade e do estado que foram cadastrados no banco de dados, como mostra a figura 27.

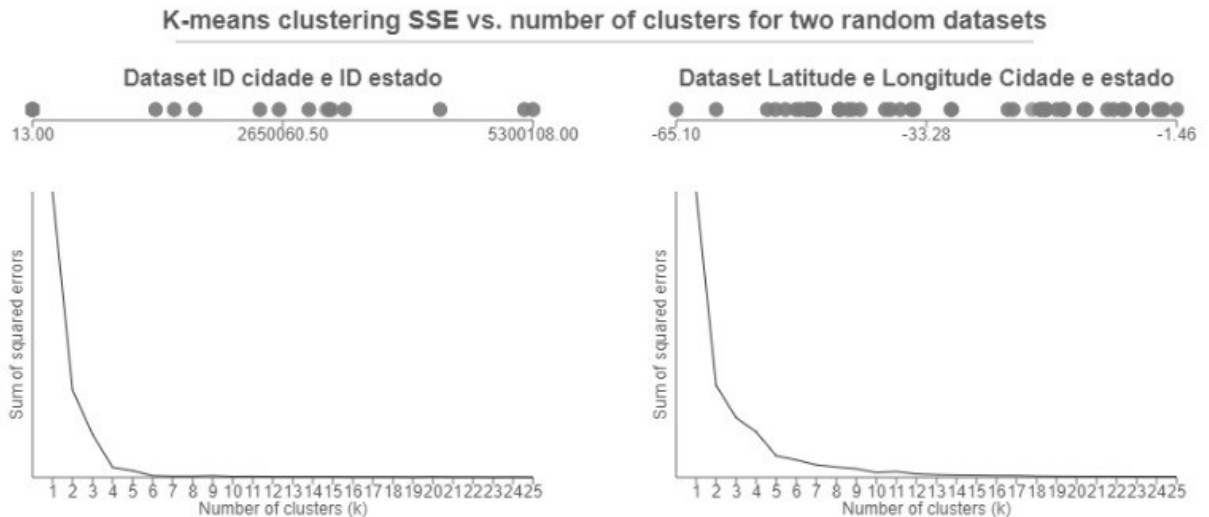


Figura 27. Número de clusters para dois conjuntos de dados aleatórios do sistema Rede Gamer.

Fonte: Print da tela feita pelo autor, 2022.

Para uma melhor precisão foi utilizado o valor da latitude e da longitude das cidades e dos estados que foram cadastrados no banco de dados, então foi criado um *array* passando o nome do usuário, o valor da latitude e da longitude que o usuário se cadastrou no banco de dados e fazendo um implementador para cada gênero de jogo que o usuário curtiu, registrado no banco de dados, e informando que será dividido em 5 *clusters* conforme se vê no experimento 3 descrito na Figura 28.

```
$sqlestado=mysqli_query($conexao,"SELECT latitude,longitude FROM estados Where
codigo_uf=$codestado")or die("erro ao selecionar");
while($dadoestado = mysqli_fetch_array($sqlestado)){
    $estadolat=$dadoestado['latitude'];
    $estadolong=$dadoestado['longitude'];
}
$sqlcidade=mysqli_query($conexao,"SELECT latitude,longitude FROM
municipios Where codigo_ibge=$codcidade")or die("erro ao selecionar");
while($dadoscidade = mysqli_fetch_array($sqlcidade)){
    $cidadelat=$dadoscidade['latitude'];
    $cidadelong=$dadoscidade['longitude'];
}
$sqlidjogo=mysqli_query($conexao,"SELECT idjogo FROM jogosadicionados
WHERE idusuario='$nomeusuario'")or die("erro ao selecionar genero");
while($dadosidjogo = mysqli_fetch_array($sqlidjogo)){
    $idjogo=$dadosidjogo['idjogo'];
    $sqlgenero=mysqli_query($conexao,"SELECT idgenero from generojogo
where idjogo=$idjogo");
```



```

        while($dadosgenero=mysqli_fetch_array($sqlgenero)){
            $idgenero=$dadosgenero['idgenero'];
            $sqlnomegenero=mysqli_query($conexao,"SELECT nome from genero
where id=$idgenero");
            while($dadosnomegenero=mysqli_fetch_array($sqlnomegenero)){
                $nomegenero=$dadosnomegenero['nome'];
                if($nomegenero=='Ação') $acao++;
                if($nomegenero=='Esportes') $esporte++;
                if($nomegenero=='FPS') $fps++;
                if($nomegenero=='RTS') $rts++;
                if($nomegenero=='MOBA') $moba++;
                if($nomegenero=='RPG') $rpg++;
                if($nomegenero=='MMO') $mmo++;
                if($nomegenero=='Battle Royale') $battler++;
                if($nomegenero=='PvP') $pvp++;
                if($nomegenero=='Híbridos') $hibridos++;
                if($nomegenero=='Aventura') $aventura++;
                if($nomegenero=='Estratégia') $estrategia++;
                if($nomegenero=='Corrida') $corida++;
                if($nomegenero=='Simulação') $simulacao++;
                if($nomegenero=='Jogo on-line') $jogool++;
                if($nomegenero=='Luta') $luta++;
            }
        }
    }

    $users[$dados['idpublico']]=array($estadolat,$estadolong,$cidadelat,$cidadelong,$acao,$esporte,$fps,$rts,$moba,$rpg,$mmo,$battler,$pvp,$hibridos,$aventura,$estrategia,$corida,$simulacao,$jogool,$luta);
}
$samples=$users;
echo"<br>";
foreach ($samples as $key => $value) {
    echo $key . " | ";
    foreach ($value as $item)
        echo $item . " | " ;
    echo "<br/>";
}
$kmeans = new KMeans(5);
$ola=$kmeans->cluster($samples);
var_dump($ola);

```

Figura 28. Experimento 3.

Fonte: Print da tela feita pelo autor, 2022.

Este foi o algoritmo utilizado neste sistema, pois agora foi obtido o número ideal de cluster juntando usuários de regiões próximas e que gostavam do mesmo estilo de gênero de jogos.

4.4 Avaliação dos Resultados

À medida em que os resultados dos experimentos foram apresentados, foi possível avaliar o número ideal de grupos dentro do sistema; encontrar a curva de cotovelo para encontrar o número Ideal de Clusters.

Apontando os acertos, partimos do ponto de que a hipótese da criação deste sistema surgiu da ideia de criar um programa de rede social onde os usuários se sentissem à vontade em interagir e conhecer pessoas com interesses em comum, isto é, com finalidade de jogar e trocar informações sobre o jogo, desenvolver relações de afinidades (amizades, troca de dados de interesses) neste campo etc. E tal possibilidade foi testada e aprovada.

Portanto, após diversas testagens, optou-se pela aplicação do *K-means* junto com o método de *Elbow* para gerar a quantidade ideal de *clusters* ideal para o sistema.

A partir da experiência com estes 515 usuários, foi percebido que o sistema é viável, e o que falta é apenas trazer mais usuários, provavelmente através de uma campanha específica para este fim. Isso pode ser promovido através do uso de aplicativos de vídeos, tipo Kwai, Tik Tok, Adobe Premiere Rush, Likee, VHS CamCorder, Cute CUT, Movavi, entre outros, para viralizar/difundir a existência da Rede Gamer.

Considerações Finais

Este trabalho partiu do intuito de implementar uma rede social onde *gamers* ou pessoas que gostam de jogar algum jogo possam se conectar, fazer novas amizades, compartilhar informações e se reunirem para jogar.

Para isso, foi desenvolvido um sistema onde um usuário possa se conectar com outras pessoas buscando pelo nome da pessoa, ou encontrando a pessoa na sugestão de amizades que a própria rede irá fornecer, através da estrutura de agrupamento, levando em conta a região e os jogos que o usuário adicionar como favorito.

Em uma pesquisa de opinião realizada com potenciais usuários da Rede Gamer, foi observado que, no geral, as pessoas não gostam de jogar sozinhas e que elas gostariam de conhecer pessoas da mesma cidade e que gostem de jogar os mesmos jogos. Nos experimentos com dados sintéticos, observou-se dois atributos dos usuários relevantes para a recomendação de amizade nesse contexto: a região do usuário, que leva em consideração a cidade e o estado; e o gênero de jogos que o usuário adicionou como favorito.

Como limitações do trabalho, os experimentos realizados utilizaram dados artificialmente criados para este fim, sendo necessário haver uma experimentação em cenários reais e com mais usuários. Além disso, não foi possível levantar e testar mais atributos dos usuários para efeitos comparativos.

5.1 Principais Contribuições

No âmbito geral, o programa em si pode contribuir no sentido de auxiliar grupos de pessoas que já obtiveram a mesma ideia e necessidade de se ligar a pessoas que jogam os mesmos jogos com objetivo de trocar informações acerca deles, bem como criar ambiente de relações de amizade no âmbito da Internet, e também fora dela. Para esta finalidade, a rede encontra-se disponível para este público.

A análise experimental realizada permitiu identificar as características dos usuários e jogos que são importantes para aplicação do algoritmo *gamers* para realização de sugestões de amizade.

Além disso, este estudo contribuiu para aprofundamento deste autor no acesso a teorias e conceitos relacionados à criação e programação desta rede social, Rede Gamer, bem como na prática deste conjunto de etapas que deu origem ao software.

5.2 Trabalhos Futuros

Como sugestão de trabalhos futuros, é possível aliar-se a grupos de *design* que possam melhorar a marca Rede Game e o *layout* das páginas, alterar o algoritmo de sugestão de amizade utilizando a estrutura de grafos.

Sugere-se também que alunos de Comunicação Social se utilizem deste projeto para a execução de um plano de negócios, incluindo projetos de marketing, divulgação, e/ou associação a outros projetos e produtos de interesses próximos.

Referências

ANASTACIO, Bruno. **K-means: o que é, como funciona, aplicações e exemplo em Python**. Programadores ajudando Programadores [online]. Publicação em: 19 mai. 2020. Disponível em: <<https://medium.com/programadores-ajudando-programadores/k-means-o-que-%C3%A9-como-funciona-aplica%C3%A7%C3%B5es-e-exemplo-em-python-6021df6e2572>>. Acesso em 3 fev. 2022. Citado na página 29.

BLAHA, M.; RUMBAUGH, J. **Modelagem e projetos baseados em objetos com UML 2**. Rio de Janeiro: Campus, 2006. Citado na página 32.

CAELUM. **Curso desenvolvimento Web com html, CSS e Javascript. Caelum**. Escola de Tecnologia. Cursos online ao vivo. apostila, capítulo 38. Disponível em: <<https://www.caelum.com.br/apostila-html-css-javascript/38CA-eventos-com-javascript>>. Acesso em 12 mar. 2022. Citado na página 18.

DIANA, J. **Redes Sociais. nd**. Disponível em: <<https://www.todamateria.com.br/redes-sociais/>>. Acesso em: 28 nov. 2019. Citado na página 12.

DUARTE, L. **Programação Web com Node.js: Completo, do Front-end ao Back-end**. 1. ed. [S.l.]: LuizTools, 2017. Citado na página 18-19.

FLANAGAN, D. **JavaScript: O Guia Definitivo**. 6. ed. [S.l.]: Bookman, 2012. Citado na página 18.

GUEDES, Erivelton Pires. **ML em Python. Curso de Machine Learning em Python**. Publicação em 3 ago. 2019. Disponível em: <https://www.kaggle.com/eriveltonguedes/7-clusteriza-o-k-means-erivelton>. Acesso em 3 fev. 2022. Citado na página 21 e 22.

HESTERBERG, Tim. **Bootstrap**. 2011, 3(6), 497–526. doi:10.1002/wics.182. Citado na página 19.

FADEL, Augusto César; SEMAAN, Gustavo da Silva e BRITO, José André de Moura. Um estudo da aplicação de técnicas de combinação de agrupamentos. 2014. Citado na página 28.

KAGGLE. **Mall Customer Segmentation Data, ML em Python**. Kaggle: Your Machine Learning and Data Science Community. Agosto, 2019. Citado na página 23.

MALLMANN, Jackson; XAVIER, Alex dos Santos; SANTIN, Altair Olivo. **Deteção de Cibercrime em Redes Sociais: Machine Learning**. The Tenth International Conference on Forensic Computer Science and Cyber Law, São Paulo, Brazil, October 29-30, 2018. Citado na página 31.

MANOVICH, Lev. **Banco de Dados**. Arte, Tecnologia e Mediação, v. 18 n. 1, 2015. DOI: <https://doi.org/10.29146/eco-pos.v18i1.2366>. Citado na página 19.

MILETTO, S. d. C. B. E. M. **Desenvolvimento de Software II**: Introdução ao Desenvolvimento Web com HTML, CSS, JavaScript e PHP. 1. ed. [S.l.]: Bookman, 2014. Citado na página 15 e 17.

NEVES, Eric Vieira das. **Agrupamento em dois níveis para disseminação de mensagens em Redes Sociais Móveis tolerantes a atrasos e desconexões**. 2018. 107 f. Dissertação (Mestrado em Informática) - Universidade Federal do Amazonas, Manaus, 2018.
NEWZOO, A. B. M. e. **Betting on Billions: Unlocking the Power of Mobile Games**. [S.l.], 2019. Citado na página 30.

ROUSSEEUW, Peter J. **Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis**. Computational and Applied Mathematics 20, 1987, 53–65p. doi: 10.1016/0377-0427(87)90125-7. Citado na página 27-28.

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. **Sistema de banco de dados**. 7. ed. [S.l.]: GEN LTC, 2016. Citado na página 19.

SOUZA, Bruno Ábia. **Uma abordagem para seleção de tópicos relevantes em redes sociais online**. 2017. 81 f. Dissertação (Mestrado em Informática) - Universidade Federal do Amazonas, Manaus, 2017. Citado na página 30 e 31.

TECHOPEDIA. Disponível em: <<https://www.techopedia.com/definition/32057/k-means-clustering>>. Acesso em 3 fev. 2022. Citado na página 24.

WEBNATAL. **O futuro do CSS(CSS3) e um pouco da sua história**. Publicação: 10 jun. 2007. Disponível em: <<http://Webnatal.wordpress.com/2007/06/10/o-futuro-do-csscss3-e-um-pouco-da-sua-historia/>>. Acesso em 2 jan. 2022. Citado na página 17.

PEDROSO, Robertha Pereira. **Apostila de html**. Programa de Educação Tutorial. Curso de Engenharia de Telecomunicações. Universidade Federal Fluminense. Niterói - RJ, Jun., 2007. Citado na página 15-16.

ROB, Peter; CORONEL, Carlos. **Sistemas de Banco de Dados**: Projeto, implementação e administração. Tradução da 8ª edição norte-americana. Revisão técnica Ana Paula Appel. 1ª edição. São Paulo: Cengage Learning, 2011. Citado na página 19-20.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. [S.l.]: Manole, 2005. ISBN 8520416837. Citado na página 23.

FREITAS, A. A. **Data Mining and Knowledge Discovery with Evolutionary Algorithms**. [S.l.]: Springer, 1998. ISBN 3540433317. Citado na página 23.

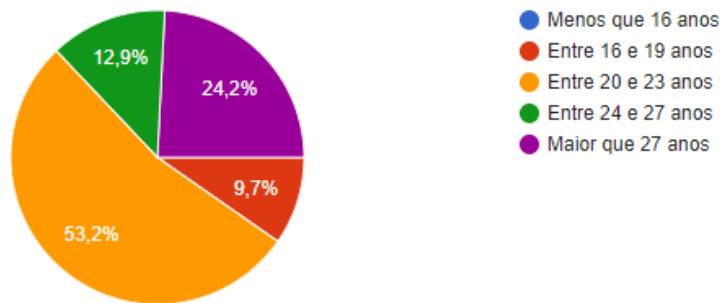
Apêndices

Formulário de Pesquisa de Opinião

1) Indique a sua faixa etária:

62 respostas

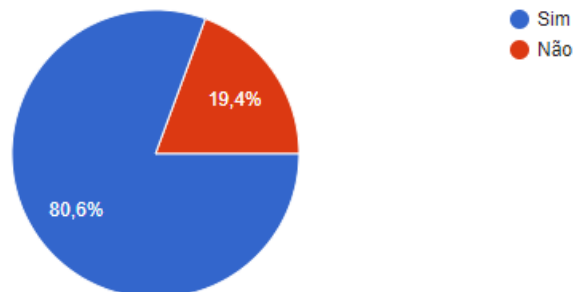
 Copiar



2) Você estuda?

62 respostas

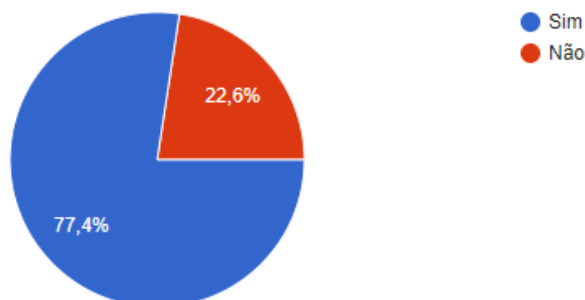
 Copiar



3) Você joga algum jogo online?

62 respostas

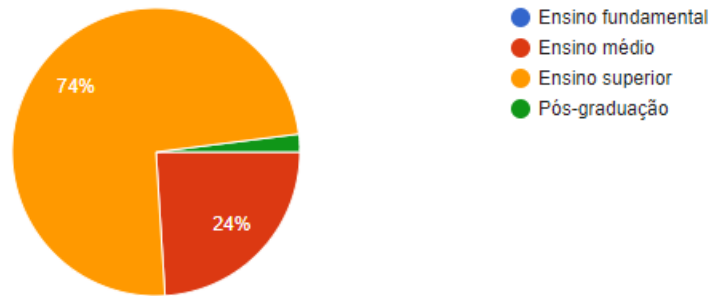
 Copiar



Qual seu nível de Escolaridade?

[Copiar](#)

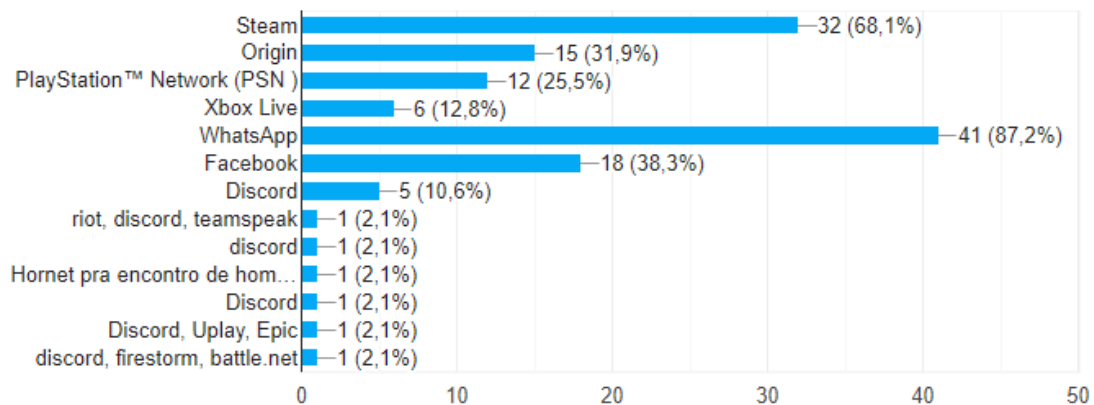
50 respostas



Você utiliza qual(is) comunidades ou sistemas para para reunir os amigos para jogar? Marque uma ou mais alternativas

[Copiar](#)

47 respostas

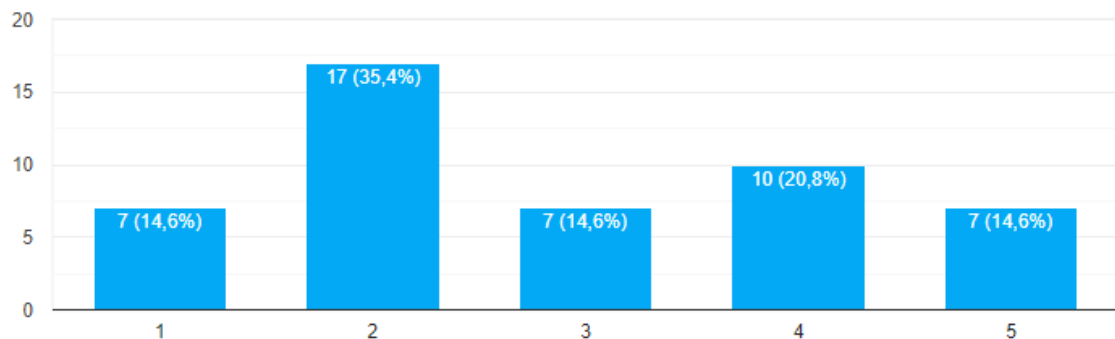


Responda as afirmativas abaixo de acordo com o nível de concordância

4) Eu gosto de jogar sozinho

[Copiar](#)

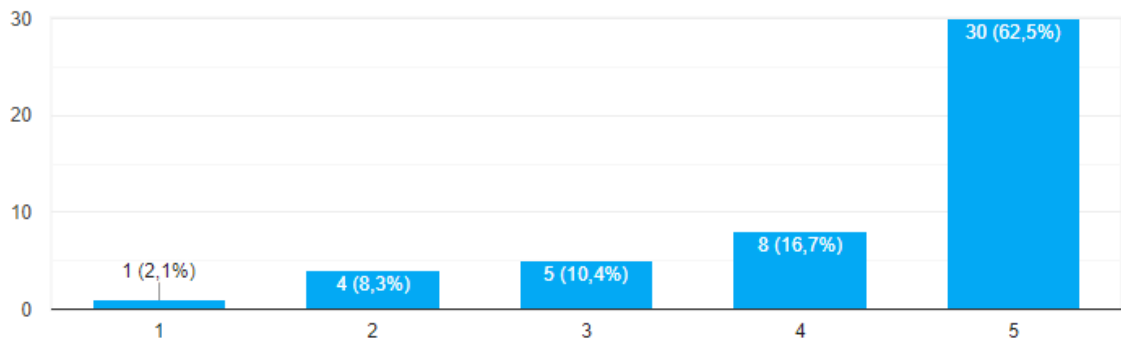
48 respostas



5) Eu gosto de jogar com algum amigo conhecido

 Copiar

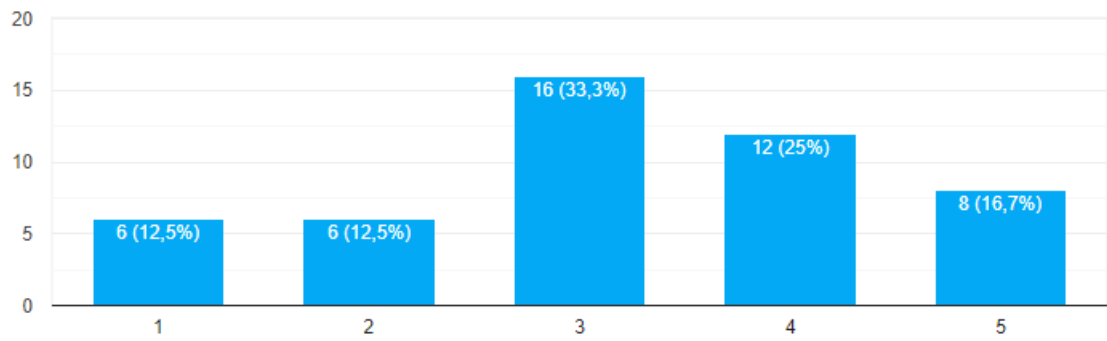
48 respostas



6) Eu gosto de jogar com novas pessoas, mesmo que desconhecidas

 Copiar

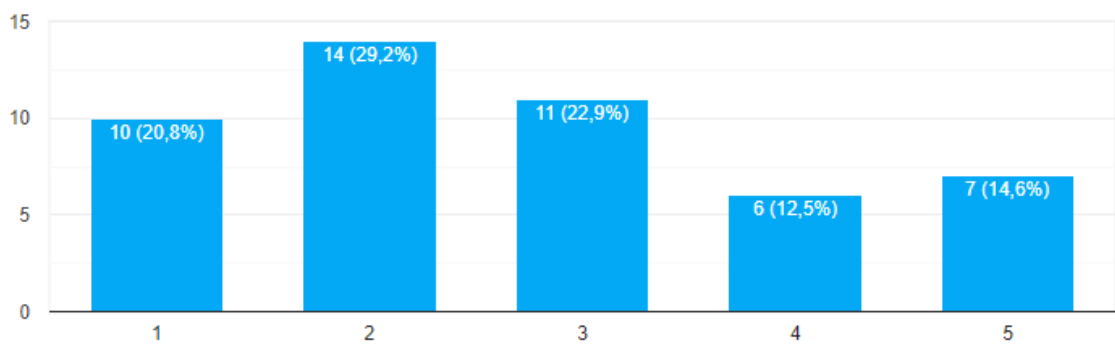
48 respostas



7) Eu acho fácil reunir os amigos para jogar

 Copiar

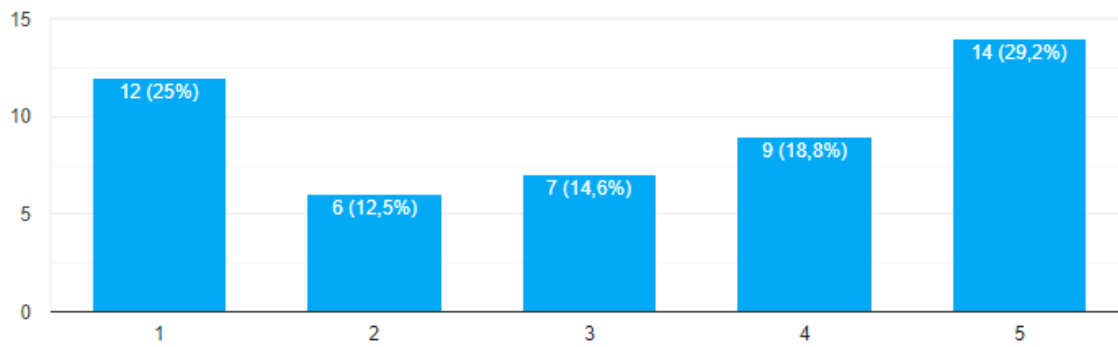
48 respostas



8) Eu gosto de jogos para celular

[Copiar](#)

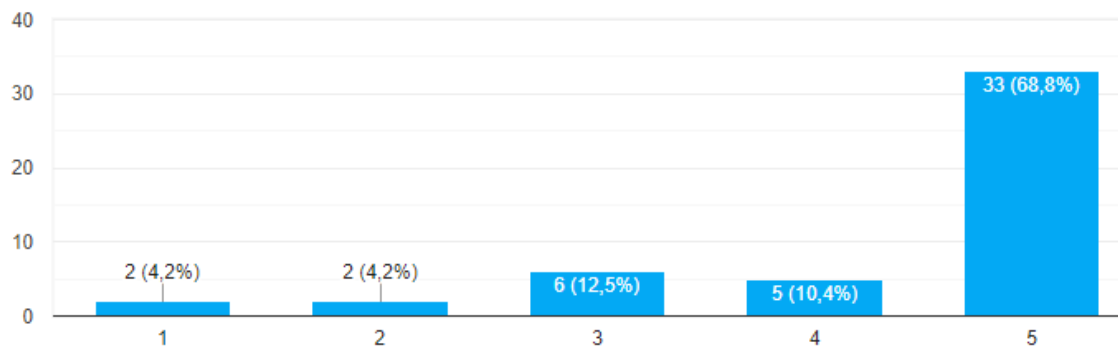
48 respostas



9) Eu gosto de jogos para computador

[Copiar](#)

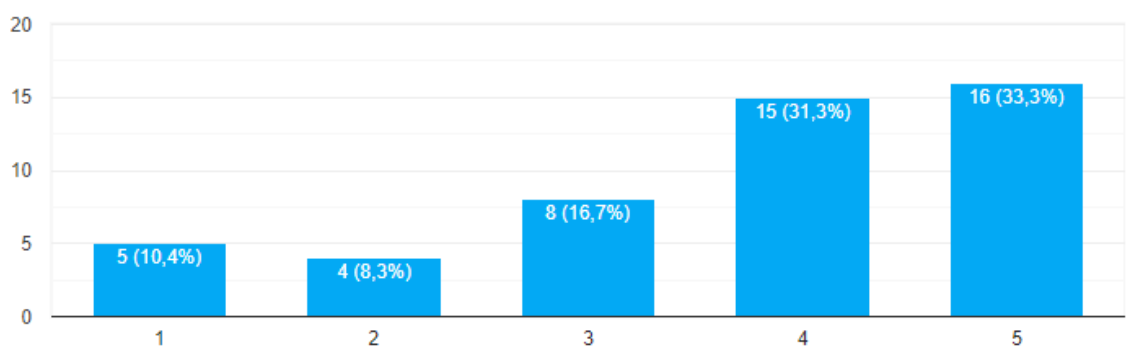
48 respostas



10) Eu gosto de jogos para consoles de video games

[Copiar](#)

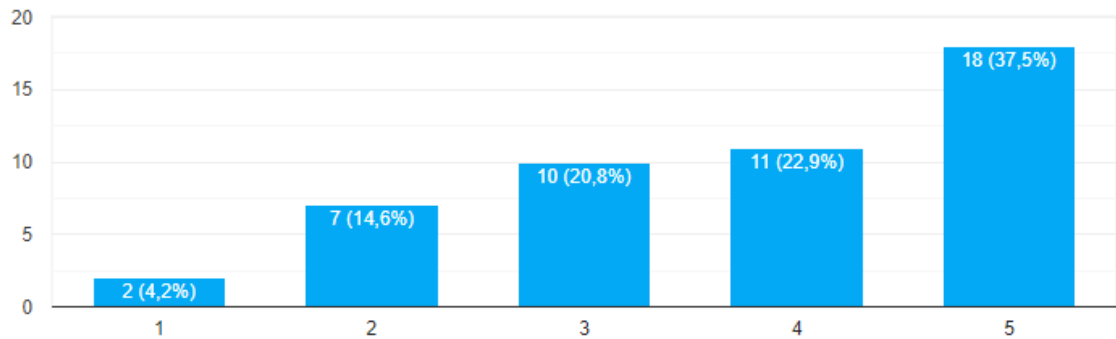
48 respostas



11) Eu sempre acho informações e comentários relacionados com os jogos que eu gosto



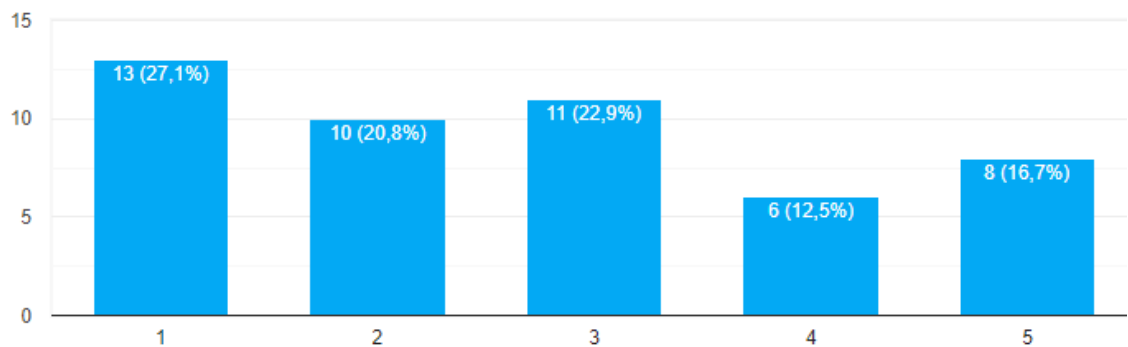
48 respostas



12) Tenho dificuldades em encontrar informações relacionadas com os jogos que eu gosto em um lugar só



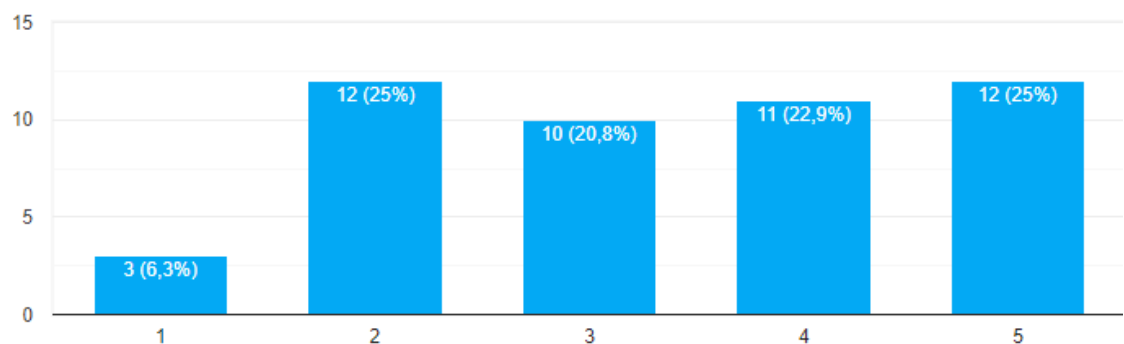
48 respostas



13) Sempre que eu preciso, eu consigo contactar facilmente meus amigos que gostam dos mesmos jogos que eu



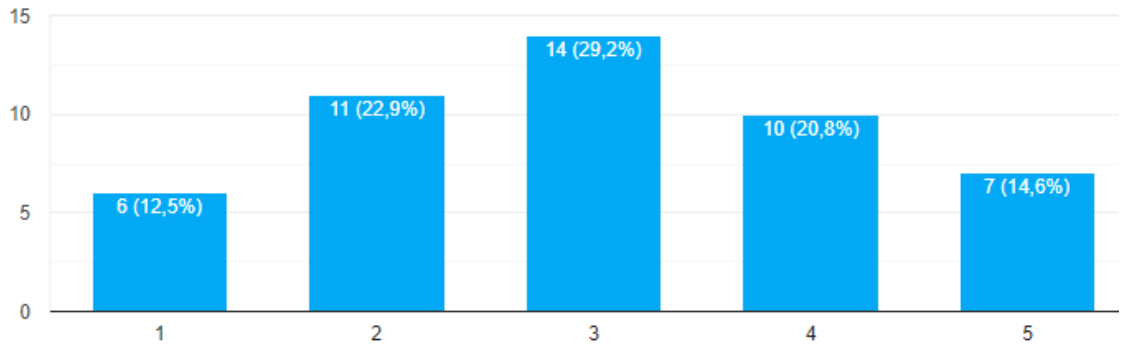
48 respostas



14) Sempre que eu preciso, eu consigo encontrar novas pessoas que gostam dos mesmos jogos que eu

 Copiar

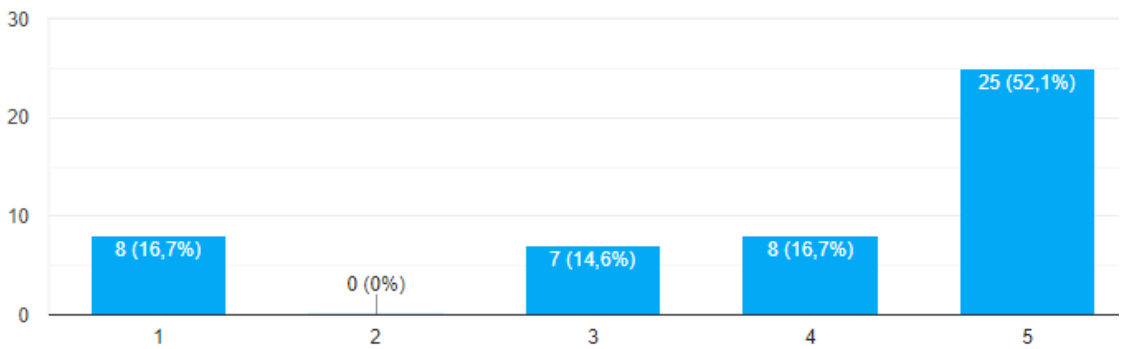
48 respostas



15) Eu gostaria de conhecer pessoas da minha cidade que jogam os mesmos jogos que eu

 Copiar

48 respostas



16) Eu gosto de dar dicas e ajudar meus amigos sobre os jogos que eu gosto

 Copiar

48 respostas

