

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

GABRIEL DALZOTO SALLES

**APLICAÇÃO DE TÉCNICAS DE ANÁLISE DE DADOS EM UM JOGO SÉRIO
SOBRE DATAS COMEMORATIVAS**

PONTA GROSSA

2024

GABRIEL DALZOTO SALLES

**APLICAÇÃO DE TÉCNICAS DE ANÁLISE DE DADOS EM UM JOGO SÉRIO
SOBRE DATAS COMEMORATIVAS**

**Application of Data Analysis techniques in a serious game about
Commemorative Dates**

Trabalho de conclusão de curso de graduação apresentada como requisito para obtenção do título de Bacharel em Ciência da Computação da Universidade Tecnológica Federal do Paraná (UTFPR).
Orientador(a): Helyane Bronoski Borges.

PONTA GROSSA

2024



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

GABRIEL DALZOTO SALLES

**APLICAÇÃO DE TÉCNICAS DE ANÁLISE DE DADOS EM UM JOGO SÉRIO
SOBRE DATAS COMEMORATIVAS**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do título de
Bacharel em Ciência da Computação da Universidade
Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 28/maio/2024

Helyane Bronoski Borges
Doutorado
Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa

Simone Nasser Matos
Doutorado
Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa

Simone de Almeida
Doutorado
Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa

PONTA GROSSA

2024

Dedico este trabalho à minha família, pelos
momentos de ausência.

AGRADECIMENTOS

Certamente estes parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre essas palavras, mas elas podem estar certas que fazem parte do meu pensamento e de minha gratidão.

Agradeço à minha família, que sempre me incentivou (e incentiva) a seguir no caminho dos estudos e compreenderam a minha ausência, em alguns momentos, na realização neste trabalho.

Agradeço também aos meus amigos e colegas de curso por toda a trajetória, desde o primeiro período até agora. É gratificante ver que todos os momentos e conversas até aqui valeram a pena.

Também agradeço às professoras e aos alunos do LESIC, por proporcionar a oportunidade de criar o jogo sério utilizado neste trabalho e de auxiliar em trabalhos publicados em eventos, como o CBIE e o SEI/SICITE.

Além disso, agradeço aos professores e alunos da instituição parceira do LESIC, por trazer um espaço para a coleta de dados do jogo.

Enfim, a todos os que por algum motivo contribuíram para a realização desta pesquisa.

RESUMO

A aplicação de jogos sérios é uma ótima prática de ensino, pois faz com que o aluno possa aprender determinado conteúdo de forma divertida e rápida. Muitos destes jogos possuem métodos de avaliação tradicionais, como questionários. Com o uso de algoritmos de aprendizagem de máquina e análise de dados, é possível gerar hipóteses e encontrar padrões não só a respeito do jogo, mas também das partidas realizadas pelos alunos, com a finalidade de propor melhorias ao jogo e descobrir características dos jogadores. Este trabalho aplicou técnicas de análise de dados e algoritmos de aprendizagem de máquina em um jogo sério com tema Datas Comemorativas. Primeiramente, foram levantadas as variáveis do jogo que seriam importantes na análise: após esta etapa, o jogo sério foi aplicado para alunos com deficiência intelectual. Após a coleta, estes dados foram pré-processados, para eliminar dados incompletos. A análise destes dados contou com a aplicação de algoritmos de aprendizagem de máquina, como a regressão linear e o K-Means, e técnicas de visualização de dados, como gráficos de barras, gráficos de setores, gráficos de dispersão e tabelas. Como resultado, foi possível descobrir características das fases do jogo, mostrando quais fases foram mais fáceis ou difíceis para o jogador, além de trazer padrões sobre as partidas de cada jogador, em relação às jogadas realizadas.

Palavras-chave: análise de dados; aprendizagem de máquina; jogos sérios.

ABSTRACT

The application of serious games is an excellent teaching practice, as it allows students to learn specific content in a fun and rapid manner. Many of these games employ traditional assessment methods, such as questionnaires. By using machine learning algorithms and data analysis, it is possible to generate hypotheses and identify patterns not only regarding the game itself but also the sessions played by students. This aims to propose improvements to the game and uncover player characteristics. This study applied data analysis techniques and machine learning algorithms to a serious game themed around Commemorative Dates. Initially, the game's variables that would be significant for the analysis were identified. Subsequently, the serious game was applied to students with intellectual disabilities. After data collection, the data were pre-processed to eliminate incomplete entries. The analysis of these data involved the application of machine learning algorithms, such as linear regression and K-Means, as well as data visualization techniques, including bar charts, pie charts, scatter plots, and tables. As a result, it was possible to identify characteristics of the game's stages, highlighting which stages were easier or more difficult for players, in addition to revealing patterns in each player's sessions concerning the moves made.

Keywords: data analytics; machine learning; serious games.

LISTA DE FIGURAS

Figura 1 – IA, Aprendizagem de Máquina e Ciência de Dados	20
Figura 2 – Ciência de Dados como interseção de três grandes áreas	20
Figura 3 – Exemplo de um gráfico de barras	27
Figura 4 – Exemplo de um gráfico de setores	28
Figura 5 – Exemplo de um gráfico de dispersão	29
Figura 6 – Exemplo de um histograma.....	30
Figura 7 – Tela inicial do Jogo das Datas Comemorativas.....	34
Figura 8 – Processo metodológico deste trabalho	36
Figura 9 – Exemplo de uma jogada da base de dados Dominó	47
Figura 10 – Exemplo de uma jogada da base de dados Jogo da Memória	54
Figura 11 – Tela da fase do Dominó, com data comemorativa Páscoa.....	79
Figura 12 – Interface de vitória nas três fases.....	80
Figura 13 – Interface de derrota nas três fases	81
Figura 14 – Tela da fase do Jogo da Memória, com data comemorativa Festa Junina.....	81
Figura 15 – Tela da fase do Labirinto, com data comemorativa Natal.....	82

LISTA DE QUADROS

Quadro 1 – Aplicações práticas de Aprendizagem de Máquina.....	22
Quadro 2 – Métodos utilizados por Algoritmos Não Supervisionados	25
Quadro 3 – <i>Strings</i> de Busca (em Inglês, Português e Espanhol).....	31
Quadro 4 – Relação entre fases e datas comemorativas no jogo.....	35
Quadro 5 – Atributos comuns a todas as telas do jogo sério	37
Quadro 6 – Atributos comuns às fases jogáveis.....	38
Quadro 7 – Atributos específicos da fase 1 do jogo sério – Dominó	38
Quadro 8 – Atributos específicos da fase 2 do jogo sério – Jogo da Memória ..	39
Quadro 9 – Atributos específicos da fase 3 do jogo sério – Labirinto	40
Quadro 10 – Critérios escolhidos para tratamento dos dados da base	42
Quadro 11 – Perguntas propostas na análise geral dos dados	61

LISTA DE GRÁFICOS

Gráfico 1 – Distribuição de Vitórias e Derrotas no Dominó.....	45
Gráfico 2 – Placares mais frequentes no Dominó	46
Gráfico 3 – Histograma da razão de <i>clicks</i> no Dominó.....	46
Gráfico 4 – Modelo de regressão linear para uma partida que teve pior jogada – Dominó.....	48
Gráfico 5 – Modelo de regressão linear para uma partida que não teve pior jogada – Dominó	49
Gráfico 6 – Contagem de piores jogadas, por partida – Dominó	49
Gráfico 7 – Resultado da aplicação do K-Means para o Dominó.....	51
Gráfico 8 – Distribuição de Vitórias e Derrotas no Jogo da Memória	52
Gráfico 9 – Placares mais frequentes no Jogo da Memória.....	52
Gráfico 10 – Histograma da razão de <i>clicks</i> no Jogo da Memória.....	53
Gráfico 11 – Modelo de regressão linear para uma partida que teve pior jogada – Jogo da Memória.....	55
Gráfico 12 – Modelo de regressão linear para uma partida que não teve pior jogada – Jogo da Memória	55
Gráfico 13 – Contagem de Pontuações a partir da Regressão Linear – Jogo da Memória.....	56
Gráfico 14 – Contagem de ordem das jogadas a partir da Regressão Linear – Jogo da Memória.....	57
Gráfico 15 – Resultado da aplicação do K-Means para o Jogo da Memória.....	58
Gráfico 16 – Distribuição de Vitórias e Derrotas no Labirinto.....	59
Gráfico 17 – Quantidade de jogadas por segundo no Labirinto	59
Gráfico 18 – Resultado da aplicação do K-Means para o Labirinto	60
Gráfico 19 – Tempo médio de cada jogador, por fase	62
Gráfico 20 – Tempo total de uso em cada tela do jogo.....	63
Gráfico 21 – Quantidade de vitórias e derrotas por fase	64
Gráfico 22 – Tempo total de cada interface das fases	64
Gráfico 23 – Média de Jogadas por fase, entre o jogador e o robô	65
Gráfico 24 – Resultado da aplicação do K-Means em relação a todos os jogadores	66
Gráfico 25 – Quantidade de vitórias e derrotas por fase	68
Gráfico 26 – Distribuição de alunos por <i>cluster</i> em cada fase.....	71

LISTA DE TABELAS

Tabela 1 – Trabalhos retornados no mapeamento sistemático	31
Tabela 2 – Resultado da coleta e do tratamento dos dados do jogo sério	42
Tabela 3 – Ordem das piores jogadas no Dominó, por partida	50
Tabela 4 – Pontuação das jogadas por tipo no Jogo da Memória	54
Tabela 5 – Resultado da coleta e do tratamento dos dados do jogo sério	62
Tabela 6 – Placares mais frequentes em cada fase	68

LISTA DE SIGLAS E ACRÔNIMOS

API	Application Programming Interface
CSV	Comma-Separated Values
DI	Deficiência Intelectual
JSON	JavaScript Object Notation
LESIC	Laboratório de Engenharia de Software e Inteligência Computacional
MSE	Mean Squared Error
UTFPR	Universidade Tecnológica Federal do Paraná

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Definição do problema e motivação	16
1.2	Objetivos	17
1.3	Justificativa.....	17
1.4	Organização do trabalho	18
2	REFERENCIAL TEÓRICO.....	19
2.1	Análise de Dados.....	19
2.1.1	Aprendizagem de Máquina.....	22
<u>2.1.1.1</u>	<u>Aprendizagem Supervisionada.....</u>	<u>23</u>
<u>2.1.1.2</u>	<u>Aprendizagem Não Supervisionada</u>	<u>24</u>
2.1.2	Técnicas de visualização de dados	26
<u>2.1.2.1</u>	<u>Gráfico de Barras</u>	<u>26</u>
<u>2.1.2.2</u>	<u>Gráfico de Setores.....</u>	<u>27</u>
<u>2.1.2.3</u>	<u>Gráfico de Dispersão.....</u>	<u>28</u>
<u>2.1.2.4</u>	<u>Histograma</u>	<u>29</u>
2.2	Trabalhos Correlatos	30
2.3	Considerações Finais do Capítulo.....	33
3	PROCESSO METODOLÓGICO PARA A ANÁLISE DE DADOS	34
3.1	Jogo Educacional sobre Datas Comemorativas.....	34
3.2	Metodologia	35
3.2.1	Levantamento de Variáveis	35
3.2.2	Coleta de Dados.....	40
3.2.3	Pré-Processamento dos Dados.....	41
3.2.4	Análise dos Dados.....	42
3.3	Considerações finais	43
4	ANÁLISE DE DADOS.....	44
4.1	Resultados	44
4.1.1	Resultados da Análise no Dominó.....	44
<u>4.1.1.1</u>	<u>Técnicas de Visualização de Dados.....</u>	<u>45</u>
<u>4.1.1.2</u>	<u>Algoritmos de Aprendizagem de Máquina.....</u>	<u>46</u>
4.1.2	Resultados da Análise no Jogo da Memória	51
<u>4.1.2.1</u>	<u>Técnicas de Visualização de Dados.....</u>	<u>51</u>
<u>4.1.2.2</u>	<u>Algoritmos de Aprendizagem de Máquina.....</u>	<u>53</u>

4.1.3	Resultados da Análise no Labirinto	58
<u>4.1.3.1</u>	<u>Técnicas de Visualização de Dados.....</u>	<u>58</u>
<u>4.1.3.2</u>	<u>Algoritmos de Aprendizagem de Máquina.....</u>	<u>60</u>
4.1.4	Resultados da Análise Geral do Jogo	61
<u>4.1.4.1</u>	<u>Técnicas de Visualização de Dados.....</u>	<u>61</u>
<u>4.1.4.2</u>	<u>Algoritmos de Aprendizagem de Máquina.....</u>	<u>65</u>
4.2	Discussão e Comparação dos Resultados	66
4.3	Considerações finais	70
5	CONCLUSÃO	73
5.1	Trabalhos futuros	74
	REFERÊNCIAS.....	75
	APÊNDICE A – Jogo das Datas Comemorativas.....	78

1 INTRODUÇÃO

A aplicação de jogos sérios na educação tem ganhado destaque nos últimos anos, mostrando-se como uma abordagem inovadora para o ensino e a aprendizagem em diversas áreas do conhecimento (SAVI; UBRICHT, 2008). De acordo com Sá, Teixeira e Fernandes (2007), esses jogos possuem o potencial de oferecer ao aluno momentos lúdicos ao aprender sobre determinado conteúdo, a partir de uma proposta interativa, além de dar suporte aos professores no processo de ensino-aprendizagem.

Para que um jogo possa ser definido como sério ou educacional, ele deve possuir objetivos pedagógicos e seu uso deve estar inserido em uma metodologia que esteja alinhada com os objetivos de ensino do professor e da Base Nacional Comum Curricular (BNCC, 2018), a qual reúne os conteúdos que o aluno deve conhecer em várias etapas da sua educação.

Além disso, a inclusão de jogos educacionais no currículo escolar traz benefícios tanto para o aluno, como o desenvolvimento de habilidades cognitivas, motivação para resolver problemas e aprendizado por descoberta (SAVI; UBRICHT, 2008), quanto para o professor, ao facilitar a aprendizagem de um conteúdo a partir de mais uma ferramenta de ensino (PRIETO *et al.*, 2005).

Para alunos com Deficiência Intelectual (DI), caracterizada por limitações nas habilidades mentais gerais de um indivíduo, a aplicação de jogos sérios também tem se mostrado como uma alternativa importante às metodologias tradicionais de ensino. Além de ajudar na dinamização do processo de aprendizagem (SOUZA, 2018), a utilização de jogos sérios contribui para a construção da autonomia, criatividade e cooperação dos alunos (MORATORI, 2003).

Além disso, é importante que, antes que um jogo seja disponibilizado, seja feita sua avaliação. Para avaliar a qualidade de um jogo sério, várias estratégias podem ser utilizadas. A mais comum é a criação de questionários, que podem ser respondidos tanto por alunos quanto professores antes ou após a utilização do jogo (SALLES *et al.*, 2023).

Alguns jogos fazem o uso da análise de dados, que são capturados enquanto o aluno joga, com a finalidade de obter análises mais precisas sobre cada jogador e seu desempenho no jogo sério (LUZ, 2021).

Na área de jogos sérios, técnicas de análise de dados são utilizadas, a fim de detectar não só o grau de satisfação e engajamento do jogador a partir de árvores de

decisão e algoritmos de classificação (DAOUDI *et al.*, 2021), mas também fatores como o tempo empregado em determinada fase ou a dificuldade que os jogadores têm com as mecânicas do jogo, utilizando modelos multinível e aprendizagem de máquina (TAUB *et al.*, 2017). Após a obtenção desses dados, são feitas previsões e correlações a partir de aprendizagem de máquina e modelos estatísticos, com a finalidade de refinar os dados e chegar a conclusões acertadas sobre o desempenho dos jogadores (KIGUCHI; SAEED; MEDI, 2022).

Neste contexto, este trabalho coletou dados de um jogo sério desenvolvido em um projeto de extensão da Universidade Tecnológica Federal do Paraná e, posteriormente, aplicou técnicas de análise de dados. A partir desta análise, pretende-se identificar características relevantes sobre o jogo sério utilizado pelos alunos com deficiência intelectual, em relação às jogadas e às fases realizadas pelos jogadores.

1.1 Definição do problema e motivação

Tendo em vista a importância da utilização de jogos educacionais, especificamente para o público de pessoas com deficiência intelectual, foi criado um jogo sério como ferramenta para o ensino de algumas datas comemorativas celebradas no Brasil, como Natal, Páscoa, Festa Junina, entre outras. Este jogo foi criado a partir do projeto de extensão intitulado Uso de Inteligência Artificial em Aplicações Computacionais, do Laboratório de Engenharia de *Software* e Inteligência Computacional (LESIC), da Universidade Tecnológica Federal do Paraná (UTFPR), Campus Ponta Grossa.

Este jogo foi desenvolvido a fim de expor os alunos a algumas curiosidades e informações sobre cada data comemorativa trabalhada na escola, além de ser uma ferramenta lúdica e divertida de ensino.

Inicialmente, este jogo sério não utilizava técnicas de análise de dados para avaliação e validação deste jogo. A partir do trabalho realizado por Silva (2023), foi percebido que é possível aplicar algoritmos de aprendizagem de máquina utilizando dados das “jogadas” dos alunos com a finalidade não apenas de avaliar o jogo educacional, mas também propor melhorias ao jogo em questão.

Conforme Silva (2023) a realização do processo de aplicação dos algoritmos de aprendizagem de máquina demonstrou ser efetiva na obtenção de informações úteis para propor atualizações e melhorias para o *software*. Além disso, os processos

de atualização do jogo sério, coleta e análise de dados de partidas permitiram identificar dificuldades enfrentadas pelo público-alvo ao jogá-lo.

1.2 Objetivos

O objetivo geral deste trabalho foi aplicar técnicas de análise de dados a um jogo sério desenvolvido para pessoas com deficiência intelectual. Os objetivos específicos definidos para atingir o objetivo geral são:

- Identificar variáveis do jogo sério para compor a análise de dados;
- Aplicar o jogo para alunos com deficiência intelectual;
- Utilizar técnicas de pré-processamento dos dados;
- Empregar técnicas de visualização de dados;
- Utilizar algoritmos de aprendizagem de máquina;
- Comparar os resultados obtidos.

1.3 Justificativa

A utilização da análise de dados em um jogo sério apresenta vantagens que tornam essa abordagem um grande diferencial na avaliação desse tipo de jogo (ALONSO-FERNÁNDEZ *et al.*, 2021). Uma dessas vantagens é o acompanhamento do desempenho dos alunos durante o jogo, fornecendo uma visão detalhada de seus pontos fortes e áreas de melhoria. Isso possibilita uma avaliação mais objetiva e individualizada do progresso de cada aluno, além de permitir uma avaliação mais adequada do jogo em si, aliada a estratégias já utilizadas para avaliar um jogo sério, como a aplicação de questionários (SALLES *et al.*, 2023).

Além disso, a análise de dados em jogos educacionais permite identificar padrões de comportamento e preferências dos alunos, o que possibilita a adaptação e customização do conteúdo do jogo de acordo com as características individuais de cada aluno, além de propor melhorias em certos aspectos do jogo (LUZ, 2023). Isso é especialmente relevante quando se trata de um jogo sério para crianças e pessoas com deficiência intelectual, pois cada estudante pode ter necessidades específicas de aprendizado e estilos de aprendizagem diferentes. Com a análise de dados, pode ser possível identificar essas particularidades e ajustar o jogo para proporcionar uma experiência de aprendizado mais inclusiva e significativa (ALONSO-FERNÁNDEZ *et al.*, 2021).

Levando em consideração estes fatores, a coleta e análise de dados relacionados ao jogo procura mostrar algumas características de jogadores com DI, a fim de traçar um perfil da jogabilidade desses alunos. Esta análise contribuirá na avaliação deste jogo educacional, além de propor melhorias para este jogo e para futuros jogos educacionais que visam ao público com deficiência intelectual.

1.4 Organização do trabalho

Este trabalho está dividido em cinco capítulos. O Capítulo 2 apresenta o referencial teórico que sustenta a pesquisa, explorando conceitos essenciais de análise de dados, suas etapas e alguns algoritmos de aprendizagem de máquina que foram utilizados para realizar a análise dos dados do jogo educacional.

O Capítulo 3 detalha a metodologia definida para a realização do trabalho: como funcionou a preparação do jogo para a coleta de dados, o tratamento destes e a utilização dos algoritmos citados no referencial teórico.

O Capítulo 4 apresenta os resultados obtidos na análise de dados, agrupados por algoritmo, por fase do jogo e a discussão sobre esses resultados. Por fim, o último capítulo apresenta a conclusão deste trabalho e explora possibilidades de trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este Capítulo tem como finalidade apresentar alguns conceitos importantes, que foram utilizados na execução deste trabalho e na apresentação dos resultados. A Seção 2.1 apresenta conceitos relacionados a Análise de Dados abordando alguns algoritmos de aprendizagem de máquina e técnicas de visualização de dados utilizados neste trabalho. A Seção 2.2 discorre sobre trabalhos correlatos, que também apresentam técnicas de análise de dados a partir de jogos educacionais. Por fim, a Seção 2.3 aborda as considerações finais deste capítulo.

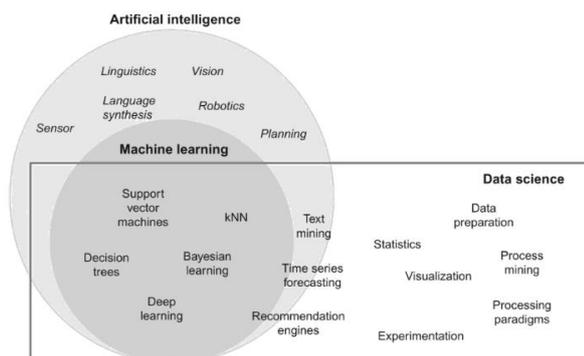
2.1 Análise de Dados

Nas últimas décadas, percebe-se um avanço notável não só no número de usuários em redes sociais, *sites* e na *Internet* de forma geral, como também na quantidade de dados que esses usuários geram ao interagir com alguma ferramenta digital. Este cenário de abundância de dados ajudou a aumentar a quantidade de estudos sobre áreas relacionadas ao tratamento e análise de dados, como *Big Data*, Ciência de Dados e Mineração de Dados (ZADEH; SHAHBAZY, 2020).

Kotu e Deshpande (2020) definem Ciência de Dados como “a aplicação no mundo dos negócios da aprendizagem de máquina, inteligência artificial e outros campos quantitativos como estatística, visualização de dados e matemática”. Portanto, é possível afirmar que a Análise de Dados e Ciência de Dados, tratados neste trabalho como sinônimos, é uma área interdisciplinar que extrai valor e informação a partir de dados.

Devido a este caráter interdisciplinar, a Análise de Dados é uma área que não depende apenas de habilidades de computação e programação, mas também conceitos matemáticos e científicos, como mostra a Figura 1. Esta figura mostra que a área da Ciência de Dados engloba conceitos de estatística, visualização de dados e preparação de dados, mas utiliza também conceitos da área de Aprendizagem de Máquina, como aprendizagem profunda, árvores de decisão, mineração de texto, entre outras classes de algoritmos utilizados. Também é possível perceber que a área de Aprendizagem de Máquina é um subconjunto da área de Inteligência Artificial.

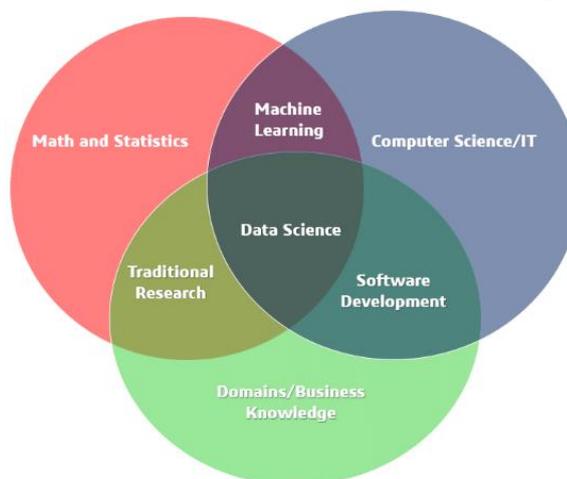
Figura 1 – IA, Aprendizagem de Máquina e Ciência de Dados



Fonte: Kotu e Deshpande (2020)

Segundo Zadeh e Shahbazy (2020), a Análise de Dados também pode ser definida como a interseção de três áreas do conhecimento: a Matemática e a Computação, como já explorado na figura anterior, mas agora junto com a área de Conhecimento de Negócios. Esta área é explorada em situações nas quais a análise de dados é voltada para o mundo empresarial. A Figura 2 mostra a relação entre estas três grandes áreas e a Ciência de Dados, que se apresenta na interseção dos conjuntos.

Figura 2 – Ciência de Dados como interseção de três grandes áreas



Fonte: Zadeh e Shahbazy (2020).

Para realizar uma análise de dados é necessário seguir um conjunto de etapas. Estas etapas contemplam o processo de uma análise, desde a coleta dos dados até o resultado final, como mostra a definição de etapas pela IBM (2023):

- Coleta de dados: utilizando uma variedade de métodos, como extração e transformação de dados, é possível extrair tanto dados estruturados quanto não-estruturados a partir de várias fontes, como dados oriundos

da *Internet*, dados armazenados de um cliente, sensores, redes sociais e muito mais.

- Armazenamento e processamento de dados: como os dados coletados podem ter tipos e estruturas variadas, é necessário escolher uma ferramenta capaz de armazenar os dados obtidos de forma organizada e que possibilite uma visualização funcional destes dados. Aqui, também aplicam-se critérios de inclusão e exclusão de determinada categoria de dados, procurando deixar no sistema apenas os dados essenciais para a análise.
- Análise de dados: nesta etapa, são escolhidos os algoritmos de aprendizagem de máquina que ajudam na criação de modelos estatísticos e na percepção de padrões e distribuição de valores entre os dados coletados. Esta etapa inclui ferramentas de visualização de dados, como gráficos e tabelas.
- Criação de hipóteses: no final do processo, são apresentadas algumas hipóteses e ideias, a partir da análise. É importante ressaltar que não é possível dar 100% de precisão para as conclusões geradas pela análise. Porém, estas conclusões podem servir como um guia para a mudança de alguma atitude tomada pela instituição em questão.

Na educação, a análise de dados também possui um papel importante na melhoria da aprendizagem dos alunos. Alonso-Fernández *et al.* (2021) mostra que dados para a análise do desempenho dos alunos podem ser coletados para organizar informações do perfil do estudante e registros acadêmicos, como as notas obtidas pelos alunos em provas e trabalhos.

Esse tipo de análise também acontece a partir do uso de jogos educacionais, com *Learning Analytics*, definido por Crompton, Bernacki e Greene (2020, v. 36, p. 101-105) como “a medição, coleta e análise de dados relacionados ao sujeito que aprende e a seus contextos, com o propósito de entender e otimizar o processo de aprendizagem”. A metodologia da análise de dados em jogos educacionais é dividida em etapas semelhantes às apresentadas pela IBM (2023), finalizando com a criação e a validação de hipóteses sobre o jogo trabalhado.

Os algoritmos utilizados para analisar os dados, independente da origem dos dados, sempre empregam algum tipo de aprendizagem de máquina, seja a aprendizagem supervisionada ou a não supervisionada (MITCHEL, 1997). Estes tipos

de aprendizagem são capazes de gerar modelos e identificar padrões sobre os dados fornecidos a ela. Na próxima seção, são apresentados conceitos relacionados com aprendizagem de máquina e os algoritmos utilizados para realizar a análise dos dados deste trabalho.

2.1.1 Aprendizagem de Máquina

A Aprendizagem de Máquina, traduzida do inglês “*machine learning*”, é uma subárea da Inteligência Artificial que explora a possibilidade que os computadores possuem de aprender utilizando grandes quantidades de dados.

Mitchel (1997) exemplifica a Aprendizagem de Máquina como “um programa de computador que aprende a partir da experiência E, no que diz respeito a alguma classe de tarefas T e medida de desempenho P, se o seu desempenho nas tarefas em T, conforme medido por P, melhora com a experiência E”, ou seja, um programa que “acerta” mais vezes à medida em que lida com novas experiências ou dados.

Algoritmos de aprendizagem de máquina são utilizados pela Análise de Dados por terem esta qualidade de aprender a partir de dados novos. Portanto, é importante que, ao resolver um problema utilizando um algoritmo desta classe, é necessário preparar e tratar a base de dados em questão antes de executar o algoritmo.

Alguns exemplos de aprendizagem de máquina podem ser observados no Quadro 1, adaptado a partir de casos mostrados por Gollapudi (2016).

Quadro 1 – Aplicações práticas de Aprendizagem de Máquina

Exemplo	Descrição
1. Detectar fraude de cartão de crédito	Baseando-se nos padrões de compra e utilização do cartão de um cliente, é possível identificar alguma transação que potencialmente não foi realizada por ele e tomar uma atitude sobre essa transação.
2. Processamento de Linguagem Natural	Utilizando aprendizagem de máquina, um sistema pode não só entender o que alguém está dizendo e realizar uma tarefa, como também criar textos escritos em linguagem natural.
3. Reconhecimento Facial	O reconhecimento facial está presente desde câmeras de celular que reconhecem pessoas em uma foto até em sistemas de segurança em lugares públicos com grande volume de pessoas.
4. Recomendação de produtos	A partir do histórico de compras e de várias interações de um usuário com sites e vídeos na Internet, a ideia é recomendar produtos que o usuário possa se interessar, tentando satisfazer seus gostos pessoais.

Fonte: Adaptado de Gollapudi (2016)

No que se refere aos tipos de aprendizagem de máquina pode-se citar a aprendizagem supervisionada, que recebe dados de entrada rotulados e pretende fazer previsões precisas para novos dados não vistos; e a aprendizagem não supervisionada, que recebe dados não rotulados e busca identificar padrões ou agrupamentos nos dados sem alguma orientação específica. Nas próximas seções, estes dois tipos de aprendizagem de máquina são detalhados.

2.1.1.1 Aprendizagem Supervisionada

Ludermir (2021) define a aprendizagem supervisionada como o tipo de aprendizagem na qual, para cada registro da base de dados, é apresentada ao algoritmo a resposta desejada. Cada registro possui uma série de atributos, geralmente organizados em colunas, que são analisados pelo algoritmo para atribuir o valor desejado ao rótulo, que seria a resposta correta esperada.

Para que um algoritmo de aprendizagem supervisionada seja executado de forma correta, deve-se separar a base de dados em duas partes: treinamento e teste (MITCHEL, 1997). A base de treinamento, que corresponde à maior porção da base de dados, serve para treinar o modelo gerado pelo algoritmo. Posteriormente, este modelo será aplicado na base de teste, na qual é possível perceber qual é a taxa de acerto do algoritmo, entre outros fatores de avaliação.

As duas principais tarefas utilizadas na aprendizagem supervisionada são classificação e regressão (LUDERMIR, 2021). Exemplos de algoritmos de aprendizagem supervisionada incluem árvores de decisão, algoritmos baseados em regras, regressão linear, algumas redes neurais artificiais e k-vizinhos mais próximos (K-NN) (MITCHEL, 1997).

Para Yang (2019), uma árvore de decisão é “um algoritmo que utiliza a estrutura de dados árvore, na qual cada caminho, começando no nó raiz, representa uma sequência de separação dos dados até que um resultado booleano seja alcançado em cada nó folha”. Na prática, a árvore de decisão forma um conjunto de regras, que podem ser facilmente traduzidas para linguagem natural ou para linguagens de programação.

Os algoritmos de árvore de decisão são capazes de gerar regras e criar uma hierarquia entre os atributos de uma base de dados. Utilizando métricas como ganho de informação e entropia (RUSSEL; NORVIG, 2021), estes algoritmos podem escolher o atributo que possui mais impacto na classificação de um registro.

A regressão linear (VETTER; SCHOBBER, 2018) é um algoritmo de aprendizagem supervisionada utilizado para resolver problemas de previsão e inferência estatística. Este método é empregado para modelar a relação entre uma variável dependente (também chamada de variável alvo) e uma ou mais variáveis independentes (ou preditoras). Desenvolvida para lidar com bases de dados tabulares (organizadas em linhas e colunas), a regressão linear assume que existe uma relação linear entre as variáveis preditoras e a variável alvo.

Inicialmente, o algoritmo de regressão linear ajusta uma linha reta (no caso de regressão linear simples) ou um hiperplano (no caso de regressão linear múltipla) aos dados de treinamento, minimizando a soma dos quadrados das diferenças entre os valores observados e os valores previstos pela linha ajustada. Após o ajuste inicial, o algoritmo calcula os coeficientes de regressão que melhor representam a relação linear entre as variáveis independentes e a variável dependente (VETTER; SCHOBBER, 2018).

No final da execução do algoritmo de regressão linear, obtém-se uma função matemática que modela a relação linear entre as variáveis independentes e a variável dependente, permitindo prever valores futuros da variável alvo com base em novos dados. Como esta função é linear, ela pode ser explorada visualmente, como uma linha reta. O modelo resultante pode então ser utilizado para prever valores da variável dependente com base em novos dados de entrada (RUSSEL; NORVIG, 2021).

2.1.1.2 Aprendizagem Não Supervisionada

Ghahramani (2003) define a aprendizagem não supervisionada como um programa que recebe como entrada os registros de uma base de dados, porém não recebe os rótulos, ou respostas, que são fornecidas na aprendizagem supervisionada. Portanto, em um algoritmo não supervisionado, não é possível, como na aprendizagem supervisionada, checar se a resposta que o algoritmo chegou é a correta: os registros são apenas separados em grupos (*clusters*), de acordo com padrões que o algoritmo detecta.

Em outras palavras, pode-se dizer que algoritmos de aprendizagem de máquina não supervisionada são algoritmos que não contam com a supervisão de um humano, ou seja, sem algum guia ou direcionamento para chegar à resposta correta. O próprio algoritmo procura encontrar padrões entre os dados recebidos como entrada e agrupar os registros de acordo com estes padrões. Os padrões encontrados são

fortemente dependentes dos parâmetros iniciais deste tipo de algoritmo, como número de *clusters* e ponto inicial de um *cluster*, que é gerado aleatoriamente.

O Quadro 2 apresenta as três estratégias mais utilizadas pelos algoritmos de aprendizagem não supervisionada. A seguir, é apresentado o algoritmo não supervisionado K-Means, que será utilizado na análise de dados deste trabalho.

Quadro 2 – Métodos utilizados por Algoritmos Não Supervisionados

Método	Descrição
<i>Clustering</i>	Armazenamento em <i>cluster</i> é uma técnica que agrupa dados não rotulados com base em suas semelhanças ou diferenças. Este tipo de estratégia é usado para transformar dados não classificados e brutos em grupos que representam algum padrão dentro dos dados.
Regras de Associação	É um método baseado em regras que visa encontrar relações entre variáveis de diferentes registros em uma base de dados. Um exemplo prático é o algoritmo “A Priori”, utilizado por sites de compras para sugerir produtos a um cliente.
Redução de Dimensionalidade	É uma técnica usada quando o volume de registros e atributos de uma base é muito extenso, o que dificulta a análise dos dados. Esta técnica reduz o número de entradas do algoritmo e, ao mesmo tempo, tenta preservar a integridade da base de dados.

Fonte: Autoria própria (2024)

O algoritmo K-Means, um dos algoritmos mais utilizados para análise e mineração de Dados (WU *et al.*, 2007), é um algoritmo simples e iterativo que separa uma base de dados em um número k de *clusters*, que serve como um parâmetro inicial deste algoritmo. Segundo Ahmed *et al.* (2020), a saída do algoritmo K-Means depende do número k de *clusters*, que sempre precisa ser especificado na entrada.

Quando o algoritmo K-Means é iniciado, cada *cluster* recebe um registro inicial (ou ponto inicial), com valores aleatórios para cada atributo a ser analisado pelo algoritmo. Como o K-Means é iterativo, cada registro da base é analisado e colocado em um dos *clusters* gerados. A cada iteração, o ponto inicial de cada *cluster* é recalculado a partir da média de seus registros. Após este cálculo, cada registro é analisado novamente e, se necessário, um registro pode ser realocado em outro *cluster*. A execução do K-Means é finalizada quando, após uma iteração, os *clusters* permanecem os mesmos, ou seja, quando não houve realocação de um registro para outro *cluster* (WU *et al.*, 2007).

Outro fator que influencia a saída do algoritmo K-Means é a escolha dos

pontos iniciais dos *clusters*, que é realizada de forma totalmente aleatória: dependendo dos valores do ponto inicial, alguns registros podem estar localizados em *clusters* diferentes. Por este motivo, é necessário testar o algoritmo K-Means várias vezes com a mesma base de dados, alterando os pontos iniciais e a quantidade de *clusters*.

2.1.2 Técnicas de visualização de dados

A visualização de dados, considerada uma subárea da Ciência de Dados, desempenha um papel importante ao permitir a exploração e interpretação dos dados por meio de representações visuais, auxiliando na identificação de padrões, tendências e *insights* importantes (AJIBADE; ADEDIRAN, 2016). A partir do uso de gráficos, tabelas, mapas e infográficos, é possível oferecer *insights* sobre os dados de uma ou mais bases de dados. Portanto, o objetivo da visualização é criar representações visuais interativas que explorem as capacidades cognitivas humanas na resolução de problemas, permitindo que o usuário compreenda e interprete facilmente conjuntos de informações extensos e complexos.

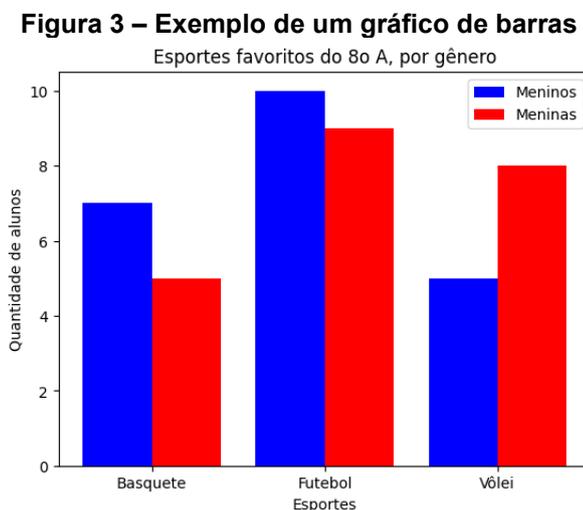
De acordo com Khan e Khan (2011), as técnicas de visualização de dados são muito importantes na análise de dados, pois elas geram modelos mentais de informação, fazendo com que grandes conjuntos de dados criem informações de fácil compreensão para o usuário. Além disso, a apresentação destes dados deve ocorrer de forma clara, descritiva e de fácil interpretação, para que o leitor possa entender rapidamente a informação que está sendo passada com a técnica utilizada.

Nas próximas subseções, algumas técnicas de visualização de dados, utilizadas posteriormente neste trabalho, são detalhadas.

2.1.2.1 Gráfico de Barras

Os gráficos de barras, também conhecidos como gráficos de colunas, são utilizados para comparar itens de diferentes grupos (AJIBADE; ADEDIRAN, 2016). É um gráfico de duas dimensões, na qual o eixo x representa o grupo a ser analisado e o eixo y representa a contagem de uma variável específica do conjunto de dados, quando se trabalha com barras verticais. O gráfico também pode apresentar barras horizontais: neste caso, é necessário inverter os eixos. A altura de cada barra, ou coluna, é proporcional à contagem desta variável.

A Figura 3 apresenta um exemplo de gráfico de barras. Este gráfico mostra a quantidade de alunos que têm como esporte favorito basquete, futebol ou vôlei, agrupados entre meninos e meninas.



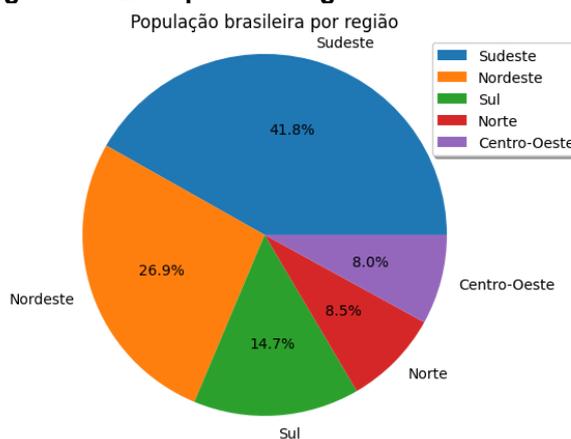
Fonte: Autoria própria (2024).

O gráfico de barras pode ser utilizado quando os valores apresentados mostram diferenças que podem ser vistas rapidamente. Neste exemplo, é possível verificar facilmente, por exemplo, que esta turma possui mais fãs de basquete e futebol entre os meninos, já que, nestes esportes, a barra azul é maior do que a barra vermelha.

Além disso, pode-se verificar qual esporte tem mais fãs nos dois grupos, verificando o tamanho das barras: também é possível fazer a contagem de cada barra, com o auxílio do eixo y, que mostra a quantidade de alunos. No gráfico de barras, é possível agrupar colunas relacionadas, como também pode ser visto no exemplo, no qual tem-se barras agrupadas por esporte.

2.1.2.2 Gráfico de Setores

O gráfico de setores, conhecido popularmente como “gráfico de pizza”, consiste em representar informações na forma de um setor circular. Este tipo de gráfico mostra valores que possuem maior frequência em um conjunto do que outros. Como este gráfico representa uma circunferência, é possível afirmar que um conjunto equivalente a 100% dos dados é um setor circular de 360 graus (CARVALHO; MENEZES; BONIDIA, 2024). Desta forma, pode-se representar qualquer porcentagem, de 0 a 100% como um setor deste círculo, representado de forma proporcional.

Figura 4 – Exemplo de um gráfico de setores

Fonte: Autoria própria (2024).

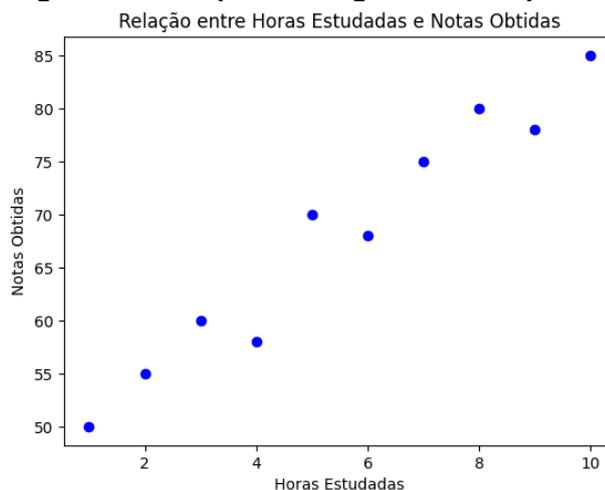
A Figura 4 exibe um exemplo de gráfico de setores, no qual a população brasileira está agrupada por região. Em um gráfico de setores, os valores absolutos podem ou não ser apresentados: em alguns casos, é mais interessante mostrar a porcentagem em relação ao número total de uma variável.

Assim como o gráfico de barras, o gráfico de setores traz de forma intuitiva qual dos grupos tem uma contagem maior de determinada variável: neste exemplo, a região Sudeste apresenta uma população maior do que todas as outras, não só em porcentagem, como visualmente: o setor desta região é maior, em comprimento, que as outras regiões.

2.1.2.3 Gráfico de Dispersão

Segundo Ajibade e Adediran (2016), um gráfico de dispersão é descrito como um gráfico de duas dimensões que exibe a variação de duas variáveis em um conjunto de dados. Este tipo de gráfico representa cada registro da base de dados como um ponto: este ponto é representado no plano cartesiano, no qual a coordenada de x representa o valor deste registro em relação a uma variável, enquanto a coordenada de y representa o valor em relação a outra variável.

Por este motivo, o gráfico de dispersão é utilizado quando se pretende descobrir uma relação entre duas variáveis de um conjunto. A Figura 5 apresenta um gráfico de dispersão que relaciona duas variáveis: o número de horas estudadas e a nota de uma prova. Cada ponto representa um aluno, que possui seu número de horas estudadas e a nota expressas como coordenadas no plano cartesiano.

Figura 5 – Exemplo de um gráfico de dispersão

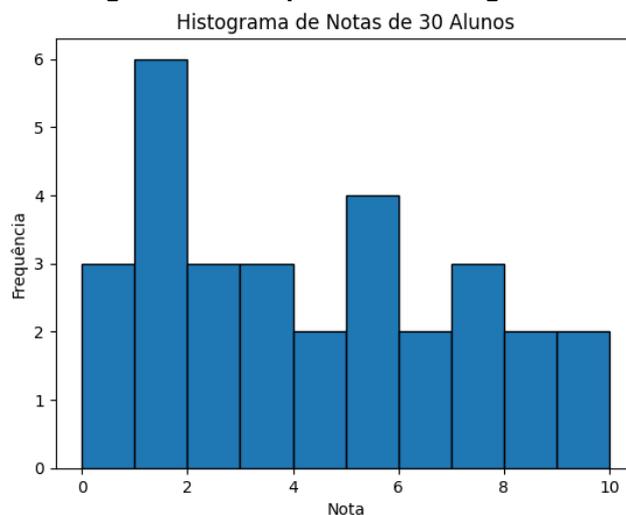
Fonte: Autoria própria (2024).

A informação que este tipo de gráfico mostra é se existe alguma relação proporcional entre duas variáveis e, se existe, como esta relação é apresentada. Além disso, é possível perceber se é possível formar grupos de elementos em um conjunto, ou identificar tendências e padrões nos dados. Neste exemplo, é possível inferir, visualmente, que os alunos que estudaram mais tiveram notas mais altas. Esta tendência se aproxima de uma função linear, representada no gráfico como uma linha reta.

2.1.2.4 Histograma

Um histograma é um tipo de gráfico de barras que representa dados agrupados por intervalos de valores. No histograma, o eixo x apresenta cada intervalo de valores possíveis em um conjunto de dados, enquanto o eixo y representa a frequência, ou seja, a quantidade de registros no conjunto pertencente a cada intervalo (CARVALHO; MENEZES; BONIDIA, 2024).

É um tipo de gráfico utilizado em situações nas quais é interessante analisar a distribuição de frequências de um conjunto, para detectar se existe um padrão entre os dados, se a distribuição se concentra em um intervalo de valores similares ou não. A Figura 6, a seguir, mostra um histograma, ou gráfico de distribuição de frequências, de uma turma de 30 alunos e a nota que cada aluno obteve em uma prova. Este tipo de gráfico agrupa os registros em intervalos de comprimento 1, ou seja, de 0 a 1, por exemplo.

Figura 6 – Exemplo de um histograma

Fonte: Autoria própria (2024).

2.2 Trabalhos Correlatos

Para que fosse feita a seleção de trabalhos que utilizam análise de dados e aprendizagem de máquina em jogos sérios, foi realizado um mapeamento sistemático com o método de Kitchenham e Charters (2007). Este método foi baseado em conceitos de medicina baseada em evidências, a fim de tornar a condução de pesquisas na área da Computação mais eficiente.

O método de Kitchenham e Charters pode ser demonstrado em três fases principais: planejamento, condução e análise de resultados (relatório). Na primeira fase, foram definidas as questões de pesquisa e o objetivo do mapeamento sistemático. Na segunda fase, foram estruturadas as palavras-chave, *strings* de busca e os repositórios e bases de dados a serem utilizadas na pesquisa. A análise dos trabalhos restantes foi realizada na última fase, respondendo às questões de pesquisa definidas na fase de planejamento.

As *strings* de busca, definidas no Quadro 3, procuraram resgatar jogos educacionais com tema Datas Comemorativas e/ou jogos educacionais que utilizaram técnicas de análise de dados.

Quadro 3 – Strings de Busca (em Inglês, Português e Espanhol)

String	Descrição
S1	(game or educational game or serious game) and (commemorative date or historic event or holiday) / (jogo ou jogo educacional ou jogo sério) e (datas comemorativas ou eventos históricos ou feriados) / (juego o juego educacional o juego serio) y (datas comemorativas o eventos historicos o día festivo)
S2	(educational game or serious game) and (data mining or data analysis or machine learning) / (jogo educacional ou jogo sério) e (mineração de dados ou análise de dados ou aprendizagem de máquina) / (juego educacional o juego serio) y (minería de datos o análisis de datos o aprendizaje automático)

Fonte: Autoria própria (2024)

A *string* de busca 2 foi a que retornou mais resultados, como é apresentado na Tabela 1.

Tabela 1 – Trabalhos retornados no mapeamento sistemático

Base de busca	String 1	String 2
ACM	10	20
IEEE Xplore	5	6
Research Gate	8	6
ScienceDirect	4	25

Fonte: Autoria própria (2024)

Para filtrar os resultados dessa pesquisa, foram utilizados critérios de inclusão e exclusão, como exclusão de trabalhos duplicados, inclusão de apenas artigos publicados nos últimos 10 anos e exclusão de trabalhos que não eram artigos científicos. No final da filtragem, foram selecionados três trabalhos.

A partir deste mapeamento foi possível concluir que os jogos educacionais que utilizam técnicas de análise de dados possuem como principais contribuições a obtenção de padrões entre as jogadas, o perfil dos jogadores e a satisfação que eles tiveram ao jogar. Também foi possível explorar a variedade de temas que os jogos educacionais tratam, assim como os diferentes níveis educacionais que são contemplados por estes jogos.

Segundo Luz (2021), os trabalhos apresentados utilizam estratégias de gamificação para prolongar a interação do usuário com o jogo, apresentando missões e níveis mais complexos à medida que o jogador evolui. Após essas interações, os eventos de todas as partidas registradas podem ser salvos em um banco de dados, com atributos bem definidos, com a finalidade de aplicar algoritmos de aprendizagem de máquina e encontrar padrões entre as jogadas. A seguir, são apresentados os trabalhos obtidos pelos mapeamentos.

Daoudi *et al.* (2021) propõem um método para analisar dados de jogos sérios sobre “Crisis Management” (gerenciamento de crises) em situações de emergência,

como um incêndio, por exemplo. A partir da coleta de dados visuais (expressões faciais dos jogadores) e textuais (mensagens trocadas pelos alunos no jogo), foi possível criar uma base que traduziu estes dados para emoções humanas básicas e o grau de mudança e alternância entre essas emoções, como confusão, tédio, engajamento e frustração. O algoritmo de decisão J48 foi utilizado para traçar um perfil emocional dos jogadores, tanto individualmente quanto em grupo. A aplicação ocorreu em um jogo que simulava a evacuação de um prédio em chamas.

O trabalho de Kiguchi *et al.* (2022) mostra técnicas para prever *churn*, uma métrica utilizada para mostrar o nível de desistência de um serviço ou produto, em jogos sérios. Os autores utilizaram algoritmos de árvore de decisão, regressão logística e florestas de decisão aleatórias para prever o *churn* de um jogo educacional do gênero Role-Playing Game (RPG), com sete capítulos. Atributos como tempo total de jogada, tempo de *login*, tempo ativo e inativo no jogo, tempo por capítulo, número de moedas obtidas, nível máximo alcançado, gênero e idade foram escolhidos para compor a análise dos dados. No fim da análise, a porcentagem de churn obtida neste jogo foi de 56,77%. Os autores afirmam que os algoritmos e atributos utilizados em sua análise podem ser adaptados para outros jogos sérios, a fim de detectar possíveis problemas nestes jogos.

Fatima *et al.* (2019) apresentam um modelo de jogo que trata sobre privacidade dos dados do usuário na Internet. Uma revisão da literatura foi feita para perceber as maiores preocupações dos usuários com relação a esse tema. Em seguida, um jogo similar ao Monopoly foi modelado a partir de algumas teorias de jogos e aplicado com um grupo de pessoas. A análise dos dados obtidos nas jogadas foi feita com modelos estatísticos de correlação e regressão para relacionar conceitos sobre privacidade de dados trabalhados no jogo. Após a conclusão da análise, o jogo proposto foi avaliado e hipóteses levantadas pelos autores foram confirmadas a partir da relação entre diferentes atributos obtidos da análise. Tais hipóteses eram relacionadas com a segurança do usuário na Internet e seus efeitos a curto e longo prazo.

Nos artigos selecionados, foi possível constatar que cada trabalho possui uma metodologia de análise de dados similar à apresentada na Seção 2.1, a qual apresenta etapas de coleta, tratamento, uso dos algoritmos e criação de hipóteses. Entretanto, cada jogo possui sua particularidade: alguns jogos utilizam apenas dados textuais ou

discretos, enquanto outros coletam dados visuais, para avaliar a satisfação dos jogadores.

Os algoritmos de aprendizagem de máquina utilizados também variam, já que cada algoritmo possui uma funcionalidade e retorna resultados diferentes, dependendo dos parâmetros aplicados. Foi possível concluir que os trabalhos utilizam algoritmos de aprendizagem supervisionada, como árvores de decisão e algoritmos baseados em regras.

Todos os trabalhos resgatados explicitam a importância da análise de dados em jogos educacionais: a obtenção de padrões, a criação de perfis de jogadores e a análise do desempenho dos alunos, tanto individual quanto coletivamente, agregam não só ao jogo, mas ao processo de aprendizagem dos alunos, independentemente do nível acadêmico trabalhado.

2.3 Considerações Finais do Capítulo

Este capítulo tem como objetivo introduzir tópicos relacionados a Análise de Dados e Aprendizagem de Máquina, assim como a divisão entre Aprendizagem Supervisionada e Não Supervisionada, apresentando algoritmos que foram utilizados na análise deste trabalho.

Além disso, foram apresentados trabalhos correlatos a este, que também fizeram a análise de dados provenientes de jogos, com a finalidade de propor melhorias ao jogo e criar hipóteses sobre o processo de aprendizagem dos jogadores e características relevantes dos mesmos.

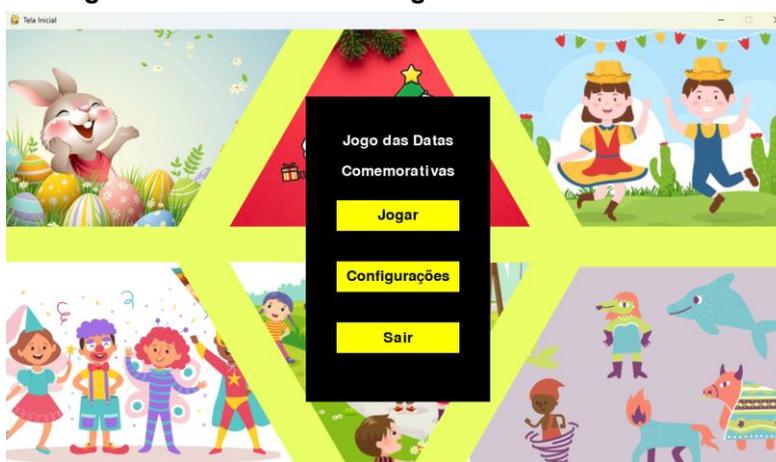
3 PROCESSO METODOLÓGICO PARA A ANÁLISE DE DADOS

Este Capítulo apresenta o processo metodológico utilizado neste trabalho, que é a execução de técnicas de análise de dados com base nos dados coletados a partir da aplicação de um jogo sério. A seção 3.1 apresenta algumas informações sobre o jogo sério utilizado neste trabalho, enquanto a seção 3.2 mostra os detalhes do processo metodológico, dividido em etapas. Por fim, a seção 3.3 apresenta as considerações finais deste capítulo.

3.1 Jogo Educacional sobre Datas Comemorativas

O Jogo das Datas Comemorativas (SALLES; MATOS; BORGES, 2023) é um jogo sério desenvolvido para pessoas com deficiência intelectual, criado com a finalidade de auxiliar no ensino de celebrações e datas comemorativas que ocorrem durante o ano letivo. A criação e o desenvolvimento deste jogo foram amparados pelo LESIC, que também auxiliou na criação de outros jogos para este público, como a Querida Floresta (SILVA *et al.*, 2022) e o PegAgente (TEIXEIRA *et al.*, 2021). A Figura 7 mostra a tela inicial do Jogo das Datas Comemorativas.

Figura 7 – Tela inicial do Jogo das Datas Comemorativas



Fonte: Autoria própria (2024)

O jogo se divide em três fases: em cada uma, o jogo trabalha uma data comemorativa, apresentando uma breve descrição, algumas imagens que ilustram essa descrição e símbolos da data comemorativa. Em todas as fases, o jogador possui como adversário o computador, chamado neste trabalho de robô, o qual implementa jogadas por meio de algoritmos de busca e de aprendizagem de máquina. As fases foram ordenadas por uma previsão do nível de dificuldade de cada uma, já as datas

comemorativas foram organizadas na ordem na qual elas ocorrem durante o ano. O Quadro 4 mostra a relação entre as fases do jogo e as datas comemorativas trabalhadas.

Quadro 4 – Relação entre fases e datas comemorativas no jogo

Nível	Fase	Data Comemorativa
1	Dominó	Páscoa
2	Jogo da Memória	Festa Junina
3	Labirinto	Natal

Fonte: Autoria própria (2024)

Além das telas de cada fase, o jogo apresenta três telas auxiliares: a tela inicial, que contém botões de iniciar o jogo, configurações e sair do jogo; a tela de seleção de fases, que mostra todas as fases do jogo; e a tela de configurações, na qual é possível ajustar o volume das músicas e dos efeitos sonoros utilizados no jogo. Informações adicionais do jogo, como as regras de cada fase, estão localizadas no Apêndice A.

3.2 Metodologia

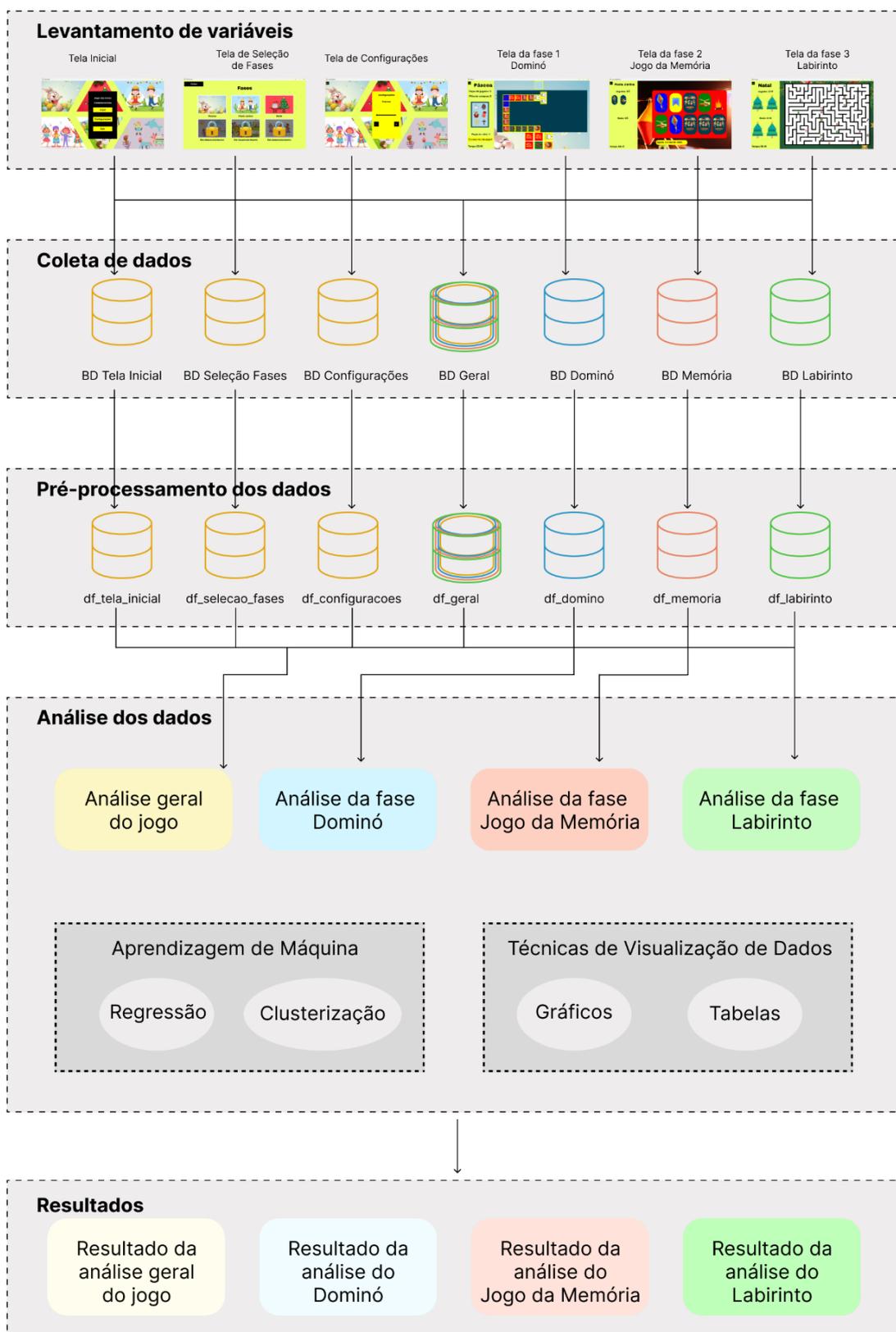
A Figura 8 apresenta o processo metodológico utilizado neste trabalho, o qual está dividida em 5 etapas: Levantamento de Variáveis, Coleta de Dados, Pré-Processamento dos Dados, Análise dos Dados e Resultados.

As etapas desta metodologia tiveram como base a análise dos dados obtidos de partidas do jogo Querida Floresta, realizada por Silva (2023) e que também utilizou algoritmos de aprendizagem de máquina, com a finalidade de comparar partidas realizadas pelos jogadores e propor melhorias ao jogo. A seguir, são detalhadas cada uma das etapas deste processo.

3.2.1 Levantamento de Variáveis

A primeira etapa da análise de dados, em qualquer contexto, é escolher quais atributos, ou variáveis, devem ser coletados. Para fazer a escolha das variáveis, cada uma das três fases do jogo sério foi analisada. Além disso, as telas auxiliares do jogo também foram analisadas nesta fase, com a finalidade de traçar um ciclo de jogada realizada pelo aluno, desde a iteração com a tela inicial até a última fase do jogo.

Figura 8 – Processo metodológico deste trabalho



Fonte: Autoria própria (2024).

Cada atributo foi definido com um nome, um tipo de dados (numérico, data, *string*, booleano, lista, etc.) e uma característica, ou seja, o que este dado traz de relevante para a análise. Posteriormente, cada atributo foi agrupado na tela que ele pertence. A escolha dos atributos mostrou que alguns deles eram comuns a todas as telas. Atributos relacionados ao tempo total, dia e hora que o jogador gastou em cada tela, assim como o número de *clicks* por tela, fazem parte desta relação, como pode ser observado no Quadro 5.

Quadro 5 – Atributos comuns a todas as telas do jogo sério

Atributo	Tipo	Descrição
<i>Token</i>	<i>String</i>	Cada ciclo de jogo possui um <i>token</i> formado por quatro letras do alfabeto, maiúsculas ou minúsculas, geradas de forma aleatória. Com este <i>token</i> , é possível mapear qual aluno realizou quais jogadas, sem precisar armazenar o nome de cada aluno.
Dia/hora	Data	No formato “DD-MM-YYYY hh:mm:ss”, cada tela registra a hora e o dia exatos que ela estava sendo utilizada pelo aluno. Este atributo é importante para organizar os dados de forma correta, garantindo que todas as jogadas ocorreram no horário de aula dos alunos.
<i>Clicks</i>	Inteiro	O número de <i>clicks</i> representa quantas vezes os alunos interagiram com cada tela via mouse.
Tempo	Inteiro	O tempo que o aluno passou em cada tela. No jogo, o tempo foi coletado no formato “mm:ss”, mas pode ser facilmente convertido para apenas segundos que o aluno passou em cada tela.

Fonte: Autoria própria (2024)

As três telas auxiliares do jogo (Tela Inicial, Seleção de Fases e Configurações) possuem apenas os atributos comuns a todas as fases, mostrados no Quadro 5. Porém, cada fase jogável (Dominó, Jogo da Memória e Labirinto) teve mais atributos escolhidos do que os comuns. O Quadro 6 mostra os atributos comuns a todas as fases, que estão relacionados a elementos que apenas as fases jogáveis possuem, excluindo os atributos já apresentados no Quadro 5.

Além disso, as fases tiveram atributos específicos escolhidos a partir de características específicas das regras de cada uma delas. A primeira fase (Dominó) foi a que contou com mais atributos entre as três fases, porque, na análise desta fase, verificou-se a importância de se registrar atributos como quantidade de peças para a esquerda e para a direita do tabuleiro, quantidade de compras do jogador e do robô, mão inicial e mão final de cada jogador, entre outros atributos específicos deste modo de jogo. O Quadro 7 mostra os atributos desta fase.

Quadro 6 – Atributos comuns às fases jogáveis

Atributo	Tipo	Descrição
Tempo_narracao	Inteiro	No início de cada fase, o jogador recebe uma descrição de cada data comemorativa, que é narrada para ele. Este atributo representa quanto tempo o aluno ouviu esta narração.
Tempo_jogo	Inteiro	Este atributo representa o tempo total que o aluno ficou realmente jogando cada fase, interagindo com o jogo via mouse/teclado.
Tempo_ganhou_perdeu	Inteiro	Após o término da fase, a tela mostra o <i>feedback</i> da jogada, ou seja, se o aluno ganhou ou perdeu. Este atributo representa quanto tempo o aluno ficou neste <i>feedback</i> .
Tempo_total	Inteiro	É a soma dos três atributos anteriores, mostra quanto tempo o jogador ficou em uma fase completa.
Qtd_jogadas_jogador	Inteiro	Quantidade de jogadas válidas realizadas pelo jogador em cada fase.
Qtd_jogadas_robô	Inteiro	Quantidade de jogadas válidas realizadas pelo robô, adversário do jogador, em cada fase.
Placar	Tupla	Placar final do jogo, representado por uma tupla, onde o primeiro elemento da tupla é a pontuação do jogador e o segundo elemento é a pontuação do robô. Exemplo: (2, 0)
Ganhou	Booleano	Retorna TRUE se o jogador ganhou aquela jogada ou FALSE se o robô ganhou ou se a jogada não foi finalizada.
Tabuleiro	Lista	Retorna uma lista de peças dispostas no tabuleiro, na fase do Dominó, a ordem das cartas na fase do Jogo da Memória, ou uma matriz que corresponde ao Labirinto.
Jogadas	Lista	Retorna a lista de jogadas que o jogador realizou em cada fase. Cada jogada possui um número (jogada 1, jogada 2) e a peça/carta/tecla escolhida naquela jogada.

Fonte: Autoria própria (2024)

Quadro 7 – Atributos específicos da fase 1 do jogo sério – Dominó

Atributo	Tipo	Descrição
Clicks_peca	Inteiro	Além de registrar a quantidade total de <i>clicks</i> , esta fase também registra os <i>clicks</i> válidos em peças do próprio jogador.
Qtd_pecas_esquerda	Inteiro	O tabuleiro do Dominó é representado por uma peça inicial, que fica no meio, peças à esquerda e peças à direita desta peça inicial. Este atributo reflete a quantidade de peças à esquerda da peça inicial.
Qtd_pecas_direita	Inteiro	O tabuleiro do Dominó é representado por uma peça inicial, que fica no meio, peças à esquerda e peças à direita desta peça inicial. Este atributo reflete a quantidade de peças à direita da peça inicial.
Qtd_pecas_tabuleiro	Inteiro	Representa quantas peças foram jogadas no tabuleiro, tanto do jogador quanto do robô.
Qtd_compras_jogador	Inteiro	A compra, nesta fase, é automática. Este atributo representa quantas vezes o jogador precisou comprar peças, por ter ficado sem possibilidades de jogada.
Qtd_compras_robô	Inteiro	A compra, nesta fase, é automática. Este atributo representa quantas vezes o robô precisou comprar cartas, por ter ficado sem possibilidades de jogada.

Qtd_pilha_compras	Inteiro	No início da fase, a pilha de compras possui 9 peças. Representa a quantidade final de cartas na pilha de compras.
Qtd_limpar_tabuleiro	Inteiro	Em algumas jogadas, o tabuleiro fica muito grande, então é necessário limpá-lo. Este atributo representa quantas vezes este processo foi feito.
Mao_inicial_jogador	Lista	Lista de 6 peças sorteadas para o jogador no início de cada fase.
Mao_inicial_robô	Lista	Lista de 6 peças sorteadas para o robô no início de cada fase.
Mao_final_jogador	Lista	Lista de peças do jogador no final de cada jogada. É uma lista vazia, caso o jogador tenha ganhado a jogada.
Mao_final_robô	Lista	Lista de peças do robô no final de cada jogada. É uma lista vazia, caso o robô tenha ganhado a jogada.

Fonte: Autoria própria (2024)

O jogo da Memória, também teve atributos específicos, relacionados aos pares que os jogadores encontraram no tabuleiro e à pontuação final de cada jogada. Estes atributos podem ser observados no Quadro 8.

Quadro 8 – Atributos específicos da fase 2 do jogo sério – Jogo da Memória

Atributo	Tipo	Descrição
Clicks_peca	Inteiro	Além de registrar a quantidade total de <i>clicks</i> , esta fase também registra os <i>clicks</i> válidos em peças do próprio jogador, semelhante ao Dominó.
Pontos_jogador	Inteiro	Reflete a pontuação do jogador no Jogo da Memória, ou seja, quantos pares de cartas foram encontrados pelo jogador no tabuleiro.
Pontos_robô	Inteiro	Reflete a pontuação do robô no Jogo da Memória, ou seja, quantos pares de cartas foram encontrados pelo robô no tabuleiro.
Cartas_jogador	Lista	Uma lista dos pares encontrados pelo jogador. Exemplo: ['fogueira', 'comidas', 'bandeira']
Cartas_robô	Lista	Uma lista dos pares encontrados pelo robô. Exemplo: ['fogueira', 'comidas', 'bandeira']

Fonte: Autoria própria (2024)

A última fase do jogo (Labirinto), por possuir interação via teclado, diferente das duas primeiras fases, precisou de alguns atributos específicos para representar este tipo de interação. Atributos como quantidade de teclas pressionadas e tipo de tecla, assim como os itens obtidos por jogador em cada jogada, foram escolhidos para esta fase, como mostra o Quadro 9.

Quadro 9 – Atributos específicos da fase 3 do jogo sério – Labirinto

Atributo	Tipo	Descrição
Qtd_teclado	Inteiro	Representa quantas vezes alguma tecla do teclado foi pressionada pelo jogador.
Teclas	Lista	Uma lista das teclas que o jogador pressionou, na ordem nas quais foram pressionadas.
Pontos_jogador	Inteiro	Este atributo mostra quantos itens do labirinto foram obtidos pelo jogador.
Pontos_robô	Inteiro	Este atributo mostra quantos itens do labirinto foram obtidos pelo robô.
Itens_jogador	Lista	Uma lista dos itens obtidos pelo jogador. Exemplo: ['tree', 'santa', 'santa', 'giftbox']
Itens_robô	Lista	Uma lista dos itens obtidos pelo robô. Exemplo: ['tree', 'santa', 'santa', 'giftbox']

Fonte: Autoria própria (2024)

Por fim, também foi percebida a necessidade de, no final de um ciclo completo de jogo, armazenar quantas vezes cada tela foi exibida ao jogador, pois, em alguns casos, o jogador poderia jogar cada fase mais de uma vez.

Após a etapa de levantamento, os atributos foram organizados em seis bases, cada uma representando uma tela diferente: as telas de Configurações, Tela Inicial e Seleção de Fases ficaram com 4 atributos cada, ou seja, os atributos comuns a todas as fases, mostrados no Quadro 5. A primeira fase (Dominó) ficou com 25 atributos, enquanto a segunda fase (Jogo da Memória) ficou com 18 atributos. A última fase (Labirinto) teve 19 atributos escolhidos.

Os atributos levantados têm como finalidade ajudar na obtenção de informações a respeito do jogo, de cada fase e da interação do jogador com cada elemento do jogo.

3.2.2 Coleta de Dados

Os dados coletados do jogo sério são armazenados em uma base de dados externa ao jogo. Foi escolhida a ferramenta *online* Google Planilhas, que permite organizar e armazenar dados em planilhas utilizando a nuvem.

Foi criada uma planilha no Google Planilhas com sete abas, seis delas representando as telas e uma aba chamada de “Geral”, que armazena um *token* de jogo e um registro de quantas vezes cada *token* utilizou cada tela do jogo.

A conexão entre os atributos armazenados no jogo e a planilha é possível com o uso de uma API (Application Programming Interface) fornecida pela própria ferramenta do Google Planilhas. Além disso, caso o jogo tivesse problemas com esta

conexão, o código-fonte do jogo é preparado para que seja possível salvar os dados das jogadas localmente.

Após a atualização do jogo sério, que envolveu a preparação do código-fonte para enviar os dados das jogadas para a planilha, o jogo foi disponibilizado para os alunos da instituição parceira do projeto de extensão Uso de Inteligência Artificial em Aplicações Computacionais. A instituição dispõe de várias máquinas, que atendem aos requisitos de *hardware* do jogo, possibilitando sua aplicação com mais de um aluno ao mesmo tempo e com várias turmas.

O jogo sério foi instalado nos computadores do laboratório e teve o início da sua aplicação no dia 02 de abril de 2024. A aplicação foi supervisionada pelo autor deste trabalho e pela professora de Informática da instituição em todas as turmas, presencialmente. A coleta dos dados foi finalizada no dia 12 de abril de 2024, envolvendo a participação de 52 alunos, divididos em 6 turmas, que continham alunos com grau leve e moderado de DI.

3.2.3 Pré-Processamento dos Dados

A base de dados, após a aplicação do jogo, passou por um processo de tratamento dos dados, com a finalidade de excluir registros que fossem repetidos ou que não representavam ciclos de jogo completos, como instâncias do jogo que foram abertas e não foram utilizadas pelos jogadores. Cada aba da planilha foi tratada com critérios diferentes. O Quadro 10 mostra o resultado do tratamento da base de dados, para cada aba.

Além deste tratamento, a aba “Geral” também recebeu mais atributos: a partir do *token* de cada jogada, foi possível obter os tempos totais e os *clicks* de cada jogador, em cada fase, e adicioná-los na base geral. Com estes dados, foi possível obter o número de *clicks* e o tempo total que cada aluno passou no jogo, por fase, por tela e a soma de todas as telas, já que cada aluno foi representado por um *token*.

Por fim, cada aba da base de dados original foi salva em um arquivo do tipo Comma-Separated Values (CSV), que possibilita o salvamento dos dados em uma estrutura de tabela. A Tabela 2 mostra a relação entre as abas da planilha, o número de atributos e o número de registros obtidos após o pré-processamento dos dados.

Quadro 10 – Critérios escolhidos para tratamento dos dados da base

Aba	Critérios	Resultados
Dominó	Exclusão de registros que representam jogos não concluídos, ou seja, placar sem zeros. Exemplo: (6,6)	3 registros removidos.
Jogo_Memória	Exclusão dos registros que representam placar zerado (0 a 0). Também foram considerados jogos que não foram concluídos e alterar a variável “Ganhou”.	9 registros removidos e 1 registro alterado.
Labirinto	Exclusão de registros que não apresentaram jogadas. Correção da variável “Ganhou” em jogos não concluídos, pois o labirinto é uma fase mais longa.	2 registros removidos e 12 registros alterados.
Configurações	Esta aba não precisou de tratamento, visto que todas as interações na tela de configurações foram realizadas com sucesso.	0 registro removido e/ou alterado.
Tela_Inicial	Exclusão dos registros que apresentaram 0 <i>clicks</i> . São situações, na coleta, nas quais o aluno abriu várias instâncias do jogo e só utilizou uma. Exclusão de registros que têm mais de 10 minutos no tempo total.	56 registros removidos.
Seleção_Fases	Exclusão dos registros que apresentaram 0 <i>clicks</i> . Registros com mais de 1 minuto na seleção de fases também foram eliminados, considerados como instâncias não utilizadas pelos alunos.	34 registros removidos.
Geral	Exclusão de registros que não tiveram fases jogadas e de registros que apenas a tela inicial foi mostrada.	66 registros removidos.

Fonte: Autoria própria (2024)

Tabela 2 – Resultado da coleta e do tratamento dos dados do jogo sério

Aba da planilha	Quantidade de atributos	Quantidade de registros
Dominó – Fase 1	25	152
Jogo da Memória – Fase 2	18	77
Labirinto – Fase 3	19	57
Configurações	4	11
Tela Inicial	4	116
Seleção de Fases	4	265
Geral	22	57

Fonte: Autoria própria (2024).

3.2.4 Análise dos Dados

A análise dos dados obtidos após o pré-processamento das bases de dados foi dividida em quatro partes, como mostra a Figura 7, sendo três análises exclusivas de cada fase do jogo sério (Dominó, Jogo da Memória, Labirinto) e uma análise descritiva de todas as bases de dados, chamada de Análise Geral.

A análise dos dados envolveu a aplicação de algoritmos de aprendizagem de máquina, com ênfase em métodos de regressão, como a regressão linear, e técnicas de clusterização, como o K-Means. Para a visualização dos resultados da análise, foram empregadas diversas representações gráficas, incluindo tabelas, gráficos de barras, gráficos de setores, gráficos de dispersão e histogramas.

Em relação aos algoritmos de aprendizagem de máquina, a regressão linear foi utilizada com o objetivo de encontrar qual foi a pior jogada de cada partida, nas fases do Dominó e do Jogo da Memória. O uso do K-Means teve como finalidade classificar as partidas de todas as fases do jogo, separando as partidas em três grupos: jogador não teve dificuldade para jogar, jogador teve pouca dificuldade ou jogador teve muita dificuldade. Os critérios utilizados para separar estes alunos foram atributos selecionados na coleta de dados, os quais serão apresentados nos resultados da aplicação do K-Means para cada fase.

Nesta etapa, foram utilizadas várias técnicas de organização de dados, com o uso da biblioteca Pandas (PANDAS, 2024), e técnicas de visualização de dados, como gráficos e tabelas, com o uso da biblioteca Matplotlib (MATPLOTLIB, 2024). Para os algoritmos de aprendizagem de máquina, foram utilizados modelos de regressão linear e de K-Means da biblioteca *Scikit-Learn* (SCIKIT-LEARN, 2024).

Os resultados destas análises são apresentados no Capítulo 4, no qual todos os resultados estão detalhados e agrupados por tipo de análise.

3.3 Considerações finais

Este capítulo teve como objetivo detalhar a metodologia deste projeto, dividida em cinco etapas: levantamento de atributos, aplicação do jogo sério e coleta de dados, tratamento dos dados, análise dos dados (apresentada no capítulo 4) e geração de resultados de cada análise. Esta divisão em etapas foi feita para garantir a qualidade dos dados coletados.

Além disso, foram mostrados quais atributos foram escolhidos para cada tela e para cada fase do jogo sério, assim como sua importância na posterior análise dos dados.

4 ANÁLISE DE DADOS

Neste Capítulo, são apresentados os resultados da análise dos dados, bem como a comparação e discussão desses resultados, obtidos por meio da aplicação de técnicas de organização e visualização de dados e algoritmos de aprendizagem de máquina, nas bases de dados geradas a partir de um jogo sério com o tema de datas comemorativas. A Seção 4.1 apresenta os resultados gerados a partir da análise das bases de dados, enquanto a Seção 4.2 mostra a discussão e a comparação dos resultados obtidos. Por fim, a Seção 4.3 apresenta as considerações finais deste capítulo.

4.1 Resultados

Nesta Seção, são apresentados os resultados da análise de dados realizada nas bases de dados obtidas com a aplicação do jogo sério. Os resultados foram agrupados em quatro categorias, como mostrou a Figura 7, apresentada no Capítulo 3.

Uma jogada é representada por dois valores: o primeiro é a pontuação esperada, ou seja, a pontuação da melhor peça a ser jogada naquele momento, e o segundo valor é a pontuação obtida, isto é, a pontuação que o jogador teve, a partir da peça que ele colocou no tabuleiro.

A partir deste conceito, também foi possível classificar uma jogada como a pior jogada de uma partida: é a jogada que mais se distancia do modelo de regressão linear. As jogadas que mais se aproximam do modelo são consideradas jogadas ótimas de uma partida.

4.1.1 Resultados da Análise no Dominó

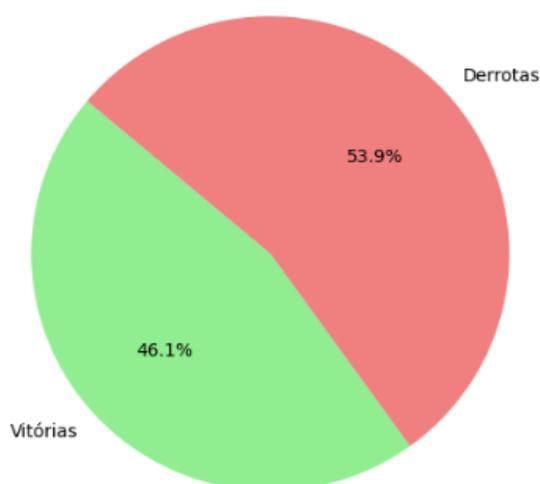
Esta Subseção apresenta os resultados da análise de dados na primeira fase do jogo sério, o Dominó. São mostrados o número de vitórias e derrotas ocorridas nessa fase, junto aos placares mais frequentes. Além disso, é analisada a taxa de acerto dos jogadores, medida pela razão de cliques. Por fim, são apresentados os resultados obtidos a partir da aplicação de algoritmos de aprendizagem de máquina, especificamente a regressão linear e o K-Means.

4.1.1.1 Técnicas de Visualização de Dados

A fase do Dominó foi jogada 152 vezes, com uma média de 2,44 partidas por jogador. No total, a fase foi jogada por 2 horas, 41 minutos e 53 segundos. Em média, cada partida teve uma duração de 2 minutos e 31 segundos. Além disso, a fase apresentou uma média de 12,99 cliques por partida.

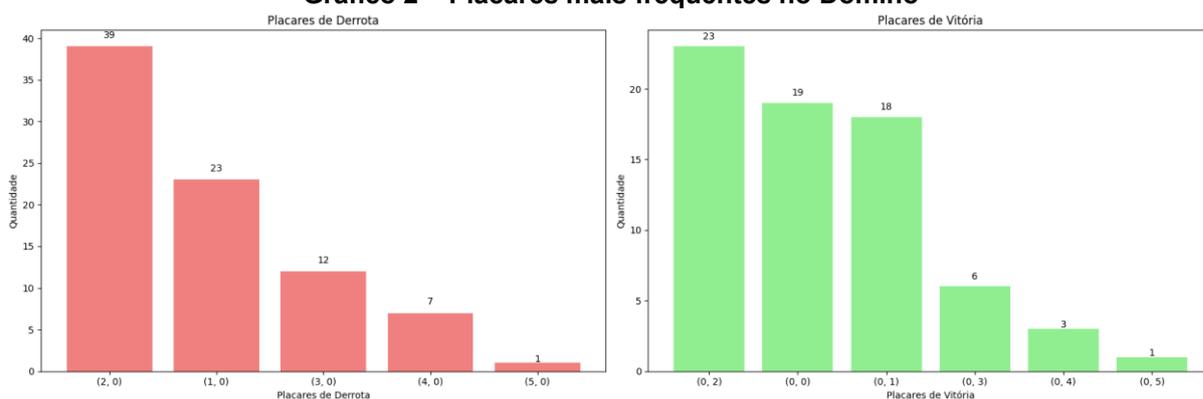
O Gráfico 1 mostra a relação de vitórias e derrotas na fase do Dominó. Das 152 partidas jogadas nesta fase, 70 partidas terminaram em vitória para o jogador, enquanto 82 partidas registraram uma derrota, ou seja, o robô saiu vitorioso. Nesta fase, a porcentagem de derrotas foi maior do que a porcentagem de vitórias.

Gráfico 1 – Distribuição de Vitórias e Derrotas no Dominó
Distribuição de Vitórias e Derrotas



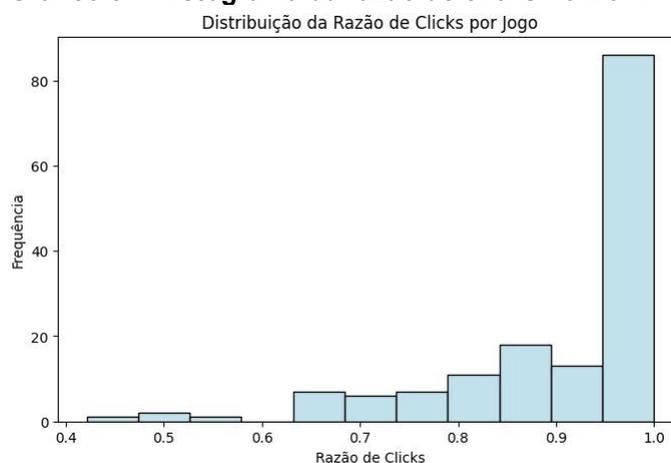
Fonte: Autoria própria (2024)

O Gráfico 2 apresenta os placares finais mais frequentes da fase do Dominó, para vitórias e derrotas. O placar mais frequente foi 2 x 0 (duas cartas restantes para o jogador e nenhuma para o robô) ou 0 x 2 (nenhuma carta restante para o jogador e duas cartas para o robô), no qual o jogador ou o robô conseguem ficar sem peças, enquanto seu adversário ficou com duas peças na mão.

Gráfico 2 – Placares mais frequentes no Dominó

Fonte: Autoria própria (2024)

Para esta fase, foi calculada a razão de *clicks* válidos, em relação à quantidade total de *clicks* que o jogador realizou em uma partida. Com esta razão, foi possível verificar se a interação via mouse ocorreu de forma satisfatória, ou seja, se o jogador entendeu que uma jogada ocorre clicando em uma das peças que estão na sua mão. A distribuição da razão de *clicks* por partida é apresentada no Gráfico 3. A frequência de cada intervalo da razão de *clicks* está representada pelas barras verticais.

Gráfico 3 – Histograma da razão de *clicks* no Dominó

Fonte: Autoria própria (2024)

4.1.1.2 Algoritmos de Aprendizagem de Máquina

Para aplicar o algoritmo de regressão linear nas jogadas do Dominó, primeiramente foi feita uma filtragem na base de dados, que retornou apenas os registros nos quais a variável “Qtd_limpar_tabuleiro” tinha valor igual a zero. Isso se deve ao fato de que, em algumas partidas, o tabuleiro ficava com muitas peças, o que atrapalhava os jogadores e, nestas partidas, foi necessário retirar todas as peças do

tabuleiro. Nestas partidas, não foi possível garantir a ordem correta das jogadas do jogador.

Algumas jogadas foram escolhidas, de forma aleatória, para servir de teste para a regressão. Uma destas jogadas é apresentada na Figura 8. As jogadas são representadas com um formato JSON (JavaScript Object Notation), no qual cada elemento possui um mapeamento chave-valor, semelhante a um dicionário. Uma jogada possui um número (primeira, segunda, terceira jogada da partida), uma marcação do tempo no qual aquela jogada ocorreu, qual foi a peça jogada, quais peças o jogador tinha em sua mão, quais eram as pontas do tabuleiro naquele momento e as peças que já tinham sido colocadas no tabuleiro, como mostra a Figura 9.

Figura 9 – Exemplo de uma jogada da base de dados Dominó

```
{'numero': 2,
'tempo': '00:39',
'peca': ('cruz', 'ovo'),
'mao_jogador': [('chocolate', 'ovo'), ('cruz', 'chocolate'), ('cruz', 'ovo')],
'pontas': ('cruz', 'ramos'),
'pecas_tabuleiro': [('cruz', 'coelho'),
('coelho', 'coelho'),
('coelho', 'ovo'),
('ovo', 'ovo'),
('ovo', 'cesta'),
('cesta', 'ramos')]}
```

Fonte: Autoria própria (2024).

Em cada jogada, o mesmo algoritmo de busca gulosa, utilizado pelo robô na fase do Dominó e apresentado no Apêndice A, foi utilizado com a finalidade de retornar qual seria a melhor peça a ser jogada naquele momento. Esse algoritmo pôde ser utilizado porque as jogadas possuem a mão do jogador, a peça jogada e as pontas do tabuleiro disponíveis naquele momento.

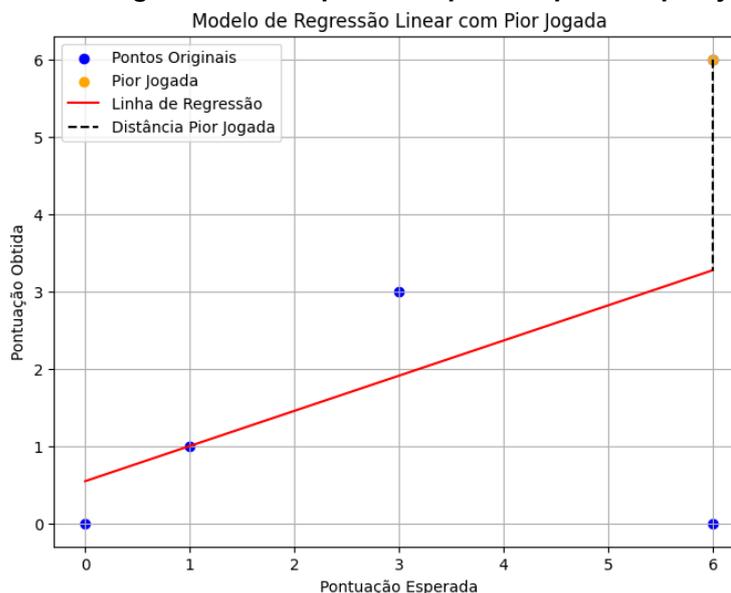
O algoritmo retorna à sequência de peças possíveis, cada uma com sua pontuação, baseada nas heurísticas que o robô utiliza: retorna o valor 6 se a peça for dupla, ou seja, se a peça tem duas imagens iguais. Se a peça for uma peça normal, com duas figuras diferentes, a busca gulosa retorna o número de peças com a mesma figura que ainda estão na mão do robô.

O Gráfico 4 mostra o resultado da aplicação do algoritmo de regressão linear em uma partida que apresentou uma jogada pior do que as outras, na fase do Dominó.

Os pontos em azul representam cada elemento da lista de jogadas, com a pontuação esperada representada no eixo x e a pontuação obtida representada no eixo y do gráfico. O resultado da aplicação do algoritmo de regressão linear é um modelo de regressão, representado no gráfico por uma linha vermelha: nessa linha,

estão contidos os pontos que mais se aproximam de todas as jogadas.

Gráfico 4 – Modelo de regressão linear para uma partida que teve pior jogada – Dominó

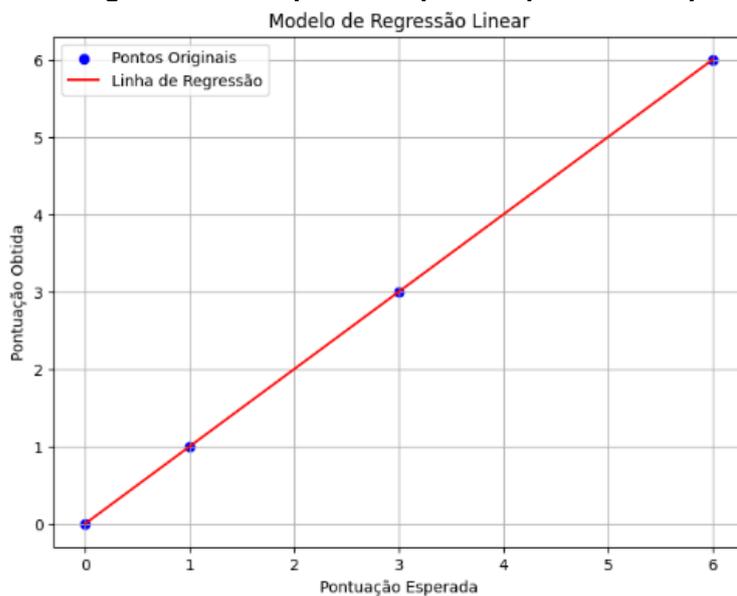


Fonte: Autoria própria (2024)

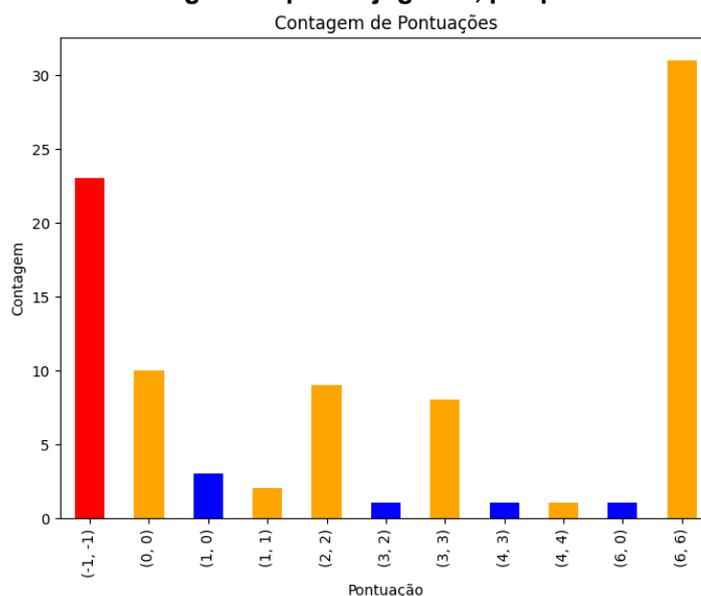
A pior jogada é representada pelo ponto amarelo: neste caso, a pior jogada foi o ponto (6,6). A linha pontilhada mostra a distância entre o ponto da pior jogada e o modelo de regressão: a jogada com a maior distância em relação ao modelo foi considerada a pior jogada, em cada partida.

O Gráfico 5 apresenta um exemplo de partida que não teve uma pior jogada. Neste caso, todos os pontos que representam jogadas estão contidos no modelo de regressão linear. Este modelo é uma função linear, na qual, para cada valor de x , o valor de y é o mesmo. Portanto, a distância de cada ponto em relação ao modelo é igual a 0.

O Gráfico 6 mostra o resultado da regressão linear para todas as jogadas. As barras em amarelo representam as partidas nas quais as piores jogadas foram aquelas com pontuação esperada e obtida iguais, enquanto as barras em azul mostram partidas nas quais as piores jogadas tiveram pontuação esperada e obtida diferentes. As barras em vermelho representam partidas que não tiveram pior jogada, ou seja, apresentaram apenas jogadas ótimas.

Gráfico 5 – Modelo de regressão linear para uma partida que não teve pior jogada – Dominó

Fonte: Autoria própria (2024)

Gráfico 6 – Contagem de piores jogadas, por partida – Dominó

Fonte: Autoria própria (2024)

Também foi identificada a frequência da ordem das piores jogadas, como pode ser observado na Tabela 3. A pior jogada mais frequente para esta fase foi a primeira jogada de uma partida, com 17 registros. As jogadas classificadas como as piores que apresentam uma frequência menor, com 1 registro, foram as sétimas e oitavas jogadas.

Tabela 3 – Ordem das piores jogadas no Dominó, por partida

Ordem da jogada	Frequência
1ª jogada	17
2ª jogada	12
3ª jogada	13
4ª jogada	7
5ª jogada	9
6ª jogada	7
7ª jogada	1
8ª jogada	1

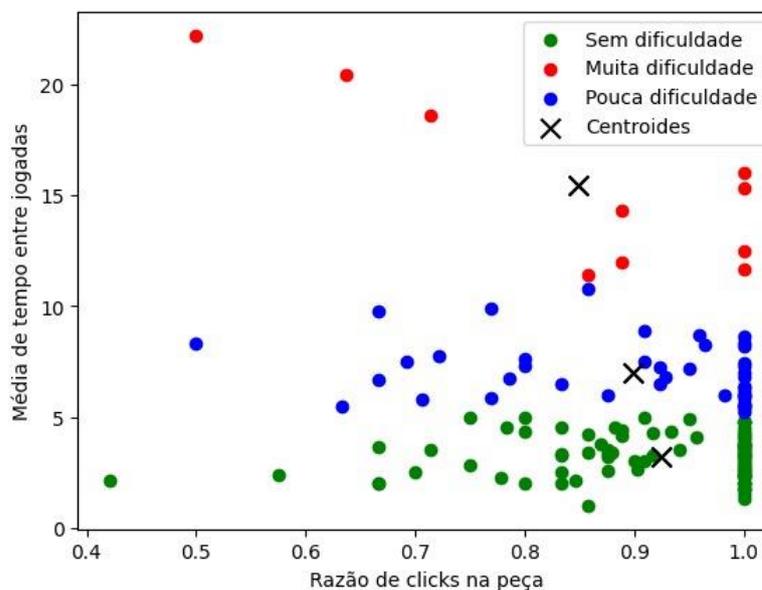
Fonte: Autoria própria (2024)

Por fim, foi calculada a média do Mean Squared Error (MSE), ou erro quadrático, uma das métricas utilizadas para medir a acurácia de um modelo de regressão linear para partidas que resultaram em vitória e partidas que terminaram em derrota para o jogador. O MSE é calculado utilizando a diferença média entre os valores previstos pelo modelo e os valores reais observados. Foi observado que a média do MSE para vitórias foi de aproximadamente 0,65, enquanto a média do MSE para derrotas foi de 1,78.

Na fase do Dominó, também foi utilizado o algoritmo de aprendizagem não-supervisionada K-Means, com a finalidade de separar as partidas em grupos, ou *clusters*, que representam o grau de dificuldade dos jogadores em cada fase. Nesta fase, as duas variáveis escolhidas para executar o K-Means foram a razão de *clicks* válidos, que corresponde à porcentagem de *clicks* válidos nas peças em relação ao número total de *clicks*, e o tempo médio entre cada jogada registrada pelo jogador. São variáveis que possibilitam separar as partidas com base no entendimento da interação usando o mouse e das regras de cada fase.

O resultado da aplicação do K-Means, na fase do Dominó, é apresentado no Gráfico 7, um gráfico de dispersão que mostra todas as partidas, cada uma representada por um ponto no plano cartesiano. O eixo x representa a razão de *clicks* válidos, enquanto o eixo y representa a média de tempo entre cada jogada. Os centroides de cada *cluster* são representados por um caracter 'x', na cor preta.

Gráfico 7 – Resultado da aplicação do K-Means para o Dominó



Fonte: Autoria própria (2024)

Cada *cluster* é representado por uma cor diferente: 98 partidas foram colocadas no *cluster* C0 (sem dificuldade), mostrado em verde, que apresenta mais valores com percentual alto de *clicks* válidos, 44 registros foram colocados no *cluster* C2 (pouca dificuldade), em azul, e 10 registros foram colocados no *cluster* C1, em vermelho, que representa o tempo médio entre jogadas maior, ou jogadores que tiveram mais dificuldade para realizar a fase.

4.1.2 Resultados da Análise no Jogo da Memória

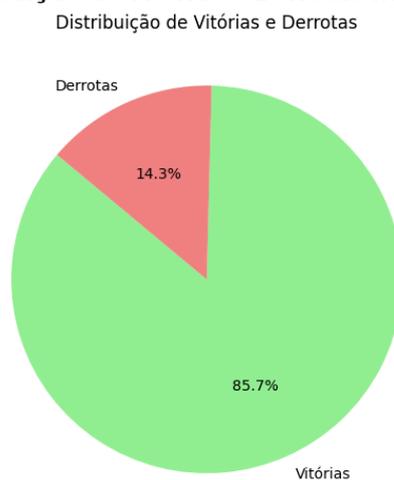
Esta subseção apresenta os resultados da análise de dados na segunda fase do jogo sério, o Jogo da Memória. Assim como na fase do Dominó, são apresentados o número de vitórias e derrotas ocorridas nesta fase, com os placares mais frequentes. Além disso, foi analisada a taxa de acerto dos jogadores, calculada pela razão de *clicks*. Por fim, são exibidos os resultados obtidos por meio da aplicação de algoritmos de aprendizagem de máquina (regressão linear e K-Means).

4.1.2.1 Técnicas de Visualização de Dados

A fase do Jogo da Memória foi jogada 77 vezes, com uma média de 1,19 partida por jogador. No total, a fase foi jogada por 2 horas e 58 segundos. Em média, cada partida teve uma duração de 1 minuto e 54 segundos. Além disso, a fase apresentou uma média de 24,12 *clicks* por partida.

O Gráfico 8 mostra a relação de vitórias e derrotas na fase do Jogo da Memória. Das 77 partidas jogadas nesta fase, 85,7% partidas terminaram em vitória para o jogador, enquanto 14,3% partidas registraram uma derrota, ou seja, o robô saiu vitorioso. Nesta fase, a porcentagem de derrotas foi maior do que a porcentagem de vitórias.

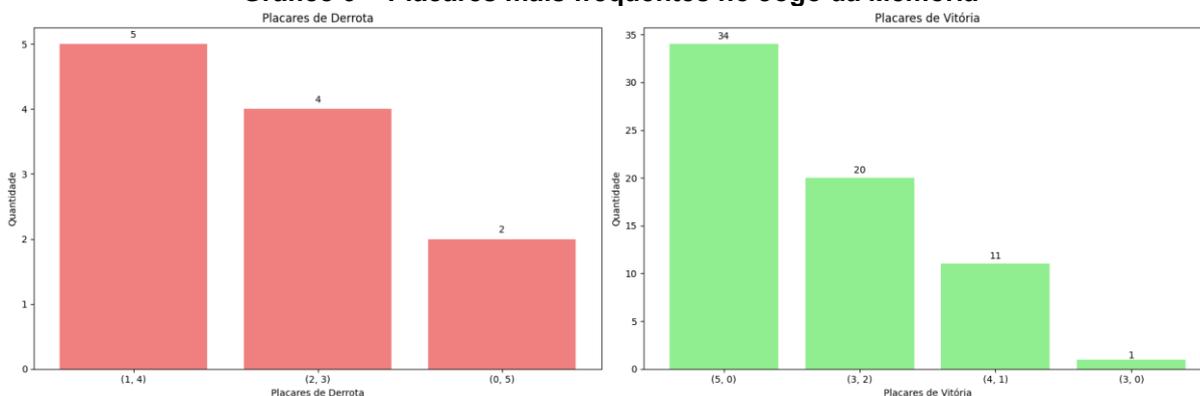
Gráfico 8 – Distribuição de Vitórias e Derrotas no Jogo da Memória



Fonte: Autoria própria (2024)

O Gráfico 9 apresenta os placares finais mais frequentes da fase do Jogo da Memória, para vitórias e derrotas. O placar mais frequente nas vitórias do jogador foi (5,0), ou seja, o jogador achou todos os pares de cartas possíveis no tabuleiro. Para as derrotas, o placar mais frequente foi (1,4): neste placar, o jogador achou apenas um par de cartas no tabuleiro, enquanto o robô encontrou quatro pares.

Gráfico 9 – Placares mais frequentes no Jogo da Memória

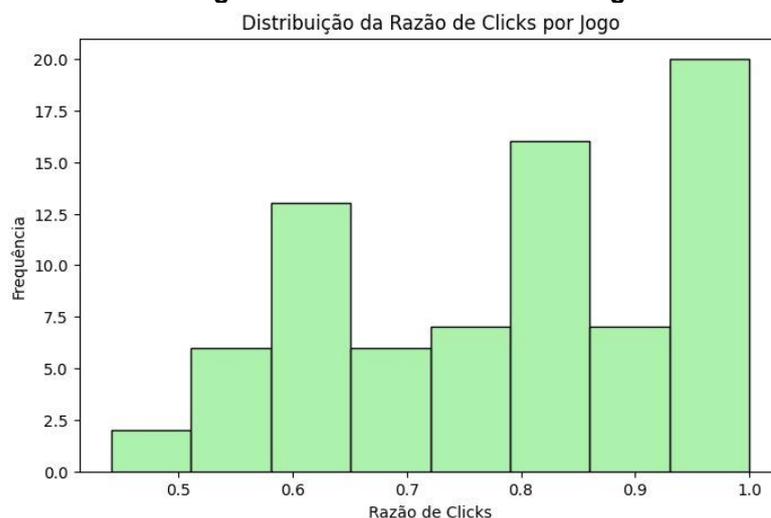


Fonte: Autoria própria (2024)

Para o Jogo da Memória, também foi calculada a razão de *clicks* válidos, já que a interação com esta fase ocorre via *mouse*, assim como o Dominó. O Gráfico 10

apresenta o histograma que contém a distribuição da razão de *clicks* no Jogo da Memória. As barras verticais representam a frequência de jogadas em cada intervalo de razão de *clicks* possível, entre 0 e 100%.

Gráfico 10 – Histograma da razão de *clicks* no Jogo da Memória.



Fonte: Autoria própria (2024)

4.1.2.2 Algoritmos de Aprendizagem de Máquina

A regressão linear também foi utilizada para analisar as jogadas da fase do Jogo da Memória. A metodologia utilizada foi semelhante à fase do Dominó: para cada partida, criar uma lista de jogadas com a pontuação esperada e a pontuação obtida, gerar um modelo de regressão linear com base nessas jogadas, analisar os resultados da regressão e medir a acurácia dos modelos. Nesta fase, não foi necessário realizar uma filtragem dos dados, já que todas as partidas do Jogo da Memória tinham os registros completos das partidas.

As jogadas registradas no Jogo da Memória também possuem um formato de mapeamento chave-valor, semelhante a um arquivo JSON. Cada jogada contém um número, que representa a ordem da jogada, as cartas viradas naquela jogada (sempre 2 cartas) e as posições das cartas no tabuleiro, numeradas de 0 a 9. A Figura 10 exemplifica o formato de uma jogada nesta fase.

Figura 10 – Exemplo de uma jogada da base de dados Jogo da Memória

```
{'numero': 0, 'cartas': ('balao', 'fogueira'), 'posicoes': (7, 8)}
```

Fonte: Autoria própria (2024).

Em primeiro lugar, foi elaborada uma tabela que atribui uma pontuação para cada tipo de jogada, com base na quantidade de cartas que ainda não foram exploradas no tabuleiro. A tabela 4 apresenta estas pontuações: a jogada que representa um par, ou seja, quando o jogador encontra duas cartas com a mesma imagem.

Tabela 4 – Pontuação das jogadas por tipo no Jogo da Memória

Tipo da jogada	Pontuação
Par – duas cartas iguais	10
2 cartas não vistas	7
1 carta não vista	4
2 cartas vistas anteriormente	1

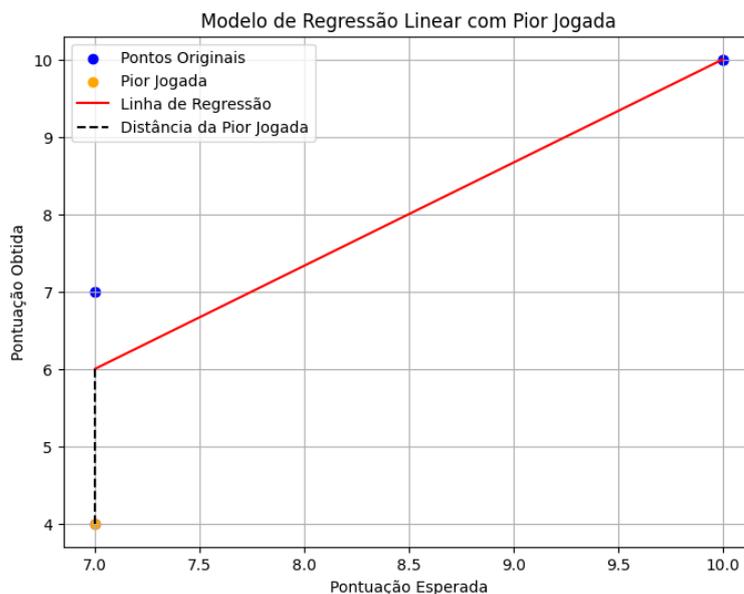
Fonte: Autoria própria (2024)

Os outros tipos de jogadas são aquelas que não representam um par: estas são separadas pela quantidade de cartas que ainda não foram exploradas no tabuleiro. Jogadas que possuem mais cartas inéditas, para o jogador, pontuam mais. Se a jogada tem menos cartas inéditas, quer dizer que alguma das cartas já foi vista, ou seja, o jogador não memorizou aquela carta como deveria.

O cálculo da pontuação esperada, em cada jogada, tem relação com o número de cartas exploradas no tabuleiro. No início de cada partida, nenhuma carta foi explorada. Se o número de cartas não exploradas for maior que 2, significa que ainda é possível ter uma jogada com 2 cartas não vistas. Nestes casos, a pontuação esperada sempre é 7. Em outros casos, como em uma jogada que resulta em um par, a pontuação esperada é o valor 10, que representa uma jogada de par. Esta correção é feita de modo que as jogadas ótimas sempre tenham valores de pontuação esperada e obtida iguais, assim como o Dominó.

O Gráfico 11 mostra o modelo de regressão linear para uma partida que apresentou uma jogada pior, em relação às outras. Os pontos em azul são as jogadas originais: nesta fase, muitas partidas apresentaram partidas com a mesma pontuação. Então, alguns pontos, como o ponto (10, 10), podem representar mais de uma jogada. A linha vermelha representa o modelo de regressão gerado a partir dos pontos das jogadas, e a pior jogada é mostrada com um ponto amarelo. A pior jogada é aquela que tem a maior distância em relação ao modelo de regressão. Esta distância é mostrada no gráfico com uma linha preta, pontilhada.

Gráfico 11 – Modelo de regressão linear para uma partida que teve pior jogada – Jogo da Memória



Fonte: Autoria própria (2024)

O Gráfico 12 mostra uma partida na qual onde nenhuma jogada foi classificada como a pior. Nesta partida, as jogadas têm duas possibilidades: ou o jogador achou um par (10, 10) ou ele fez uma jogada com duas cartas inéditas (7,7).

Gráfico 12 – Modelo de regressão linear para uma partida que não teve pior jogada – Jogo da Memória



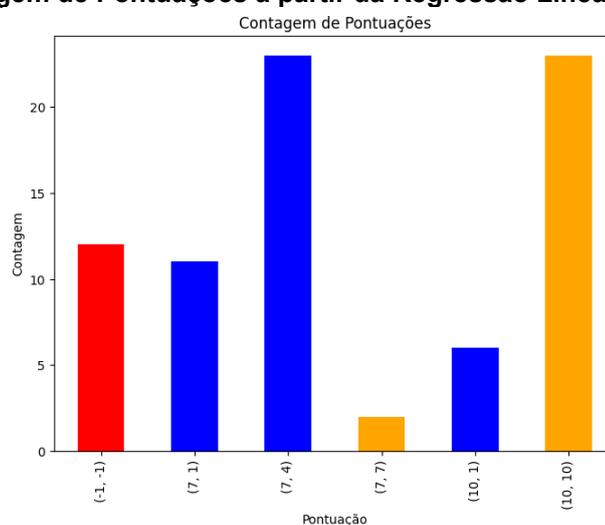
Fonte: Autoria própria (2024)

A análise resultante da aplicação da regressão linear em cada jogada foi registrada em uma nova base de dados. Esta base inclui variáveis como a identificação da pior jogada, bem como informações associadas a ela (como a peça

colocada, a mão do jogador e a ordem da jogada), a distância entre a pior jogada e o modelo, o erro quadrático do modelo e o resultado final da partida (vitória ou derrota).

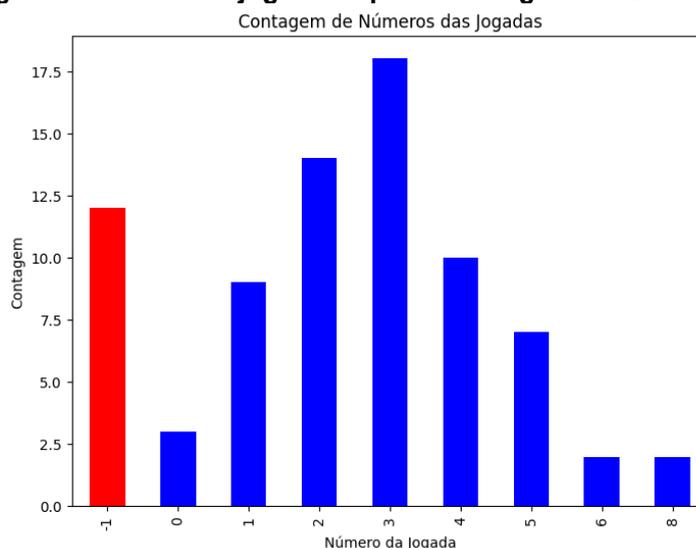
O Gráfico 13 apresenta visualmente os resultados da regressão linear para todas as jogadas. As barras amarelas indicam partidas em que as piores jogadas resultaram em pontuações esperadas e obtidas iguais. Por outro lado, as barras azuis representam partidas em que as pontuações esperadas e obtidas para as piores jogadas foram diferentes. Além disso, as barras vermelhas identificam partidas em que não houve uma pior jogada determinada.

Gráfico 13 – Contagem de Pontuações a partir da Regressão Linear – Jogo da Memória



Fonte: Autoria própria (2024)

Na fase do Jogo da Memória, foram levantadas 12 partidas que não apresentaram jogadas ruins: todas as jogadas, neste caso, seguem o padrão da partida apresentada no Gráfico 12, no qual o modelo é uma função na qual os valores de x e y são iguais.

Gráfico 14 – Contagem de ordem das jogadas a partir da Regressão Linear – Jogo da Memória

Fonte: Autoria própria (2024)

Além disso, foi feita uma análise de quais jogadas foram classificadas como as piores em relação à ordem (primeira, segunda jogada, entre outras). O Gráfico 14 mostra esta contagem. As piores jogadas nas partidas do Jogo da Memória mostraram uma frequência maior na quarta jogada, com 18 registros. As últimas jogadas, como a sétima e a oitava, apresentaram uma menor frequência na lista de piores jogadas.

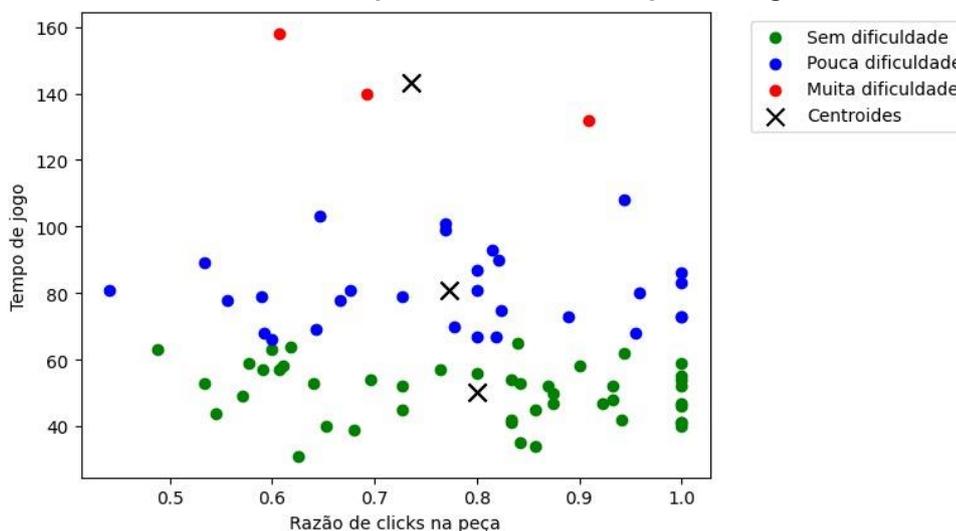
A última etapa da análise das jogadas do Jogo da Memória, utilizando a regressão linear, foi calcular a média do *Mean Squared Error* (MSE), ou erro quadrático, uma métrica frequentemente empregada para avaliar a precisão de um modelo de regressão linear. Esta métrica foi aplicada separadamente para partidas que culminaram em vitória e partidas que terminaram em derrota para o jogador. Foi observado que a média do MSE para vitórias foi de aproximadamente 7,86, enquanto a média do MSE para derrotas foi de 10,5.

Na fase do Jogo da Memória, também foi utilizado o algoritmo de aprendizagem não-supervisionada K-Means, com a finalidade de separar as partidas em grupos por grau de dificuldade dos jogadores. Nesta fase, não foi possível obter o tempo médio entre as jogadas, então foi utilizado o tempo total da partida como a segunda variável.

O Gráfico 15 mostra o resultado da aplicação do K-Means para as partidas do Jogo da Memória. Nesta fase, as partidas ficaram mais distribuídas entre os *clusters* C0 e C1, que representam nenhuma ou pouca dificuldade, mostrados em verde e azul, respectivamente. O *cluster* C0 possui 45 registros, enquanto o *cluster* C1 possui 29

registros. O *cluster* C2, que representa muita dificuldade no Jogo da Memória, em vermelho, conta com apenas 3 registros.

Gráfico 15 – Resultado da aplicação do K-Means para o Jogo da Memória



Fonte: Autoria própria (2024)

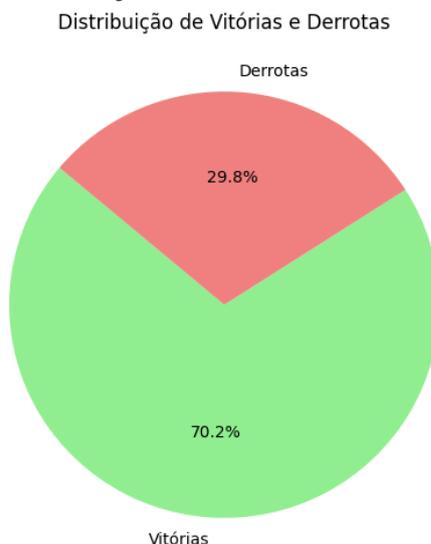
4.1.3 Resultados da Análise no Labirinto

Esta subseção apresenta os resultados da análise de dados na terceira fase do jogo sério, o Labirinto. São apresentados o número de vitórias e derrotas ocorridas nesta fase, além de mostrar a análise quantidade de jogadas por segundo em cada partida. Por fim, são exibidos os resultados obtidos por meio da aplicação de algoritmos de aprendizagem de máquina. Nesta fase, foi utilizado apenas o algoritmo K-Means, porque, nesta fase, não foi possível reconstruir as jogadas dos jogadores para utilizar a regressão linear.

4.1.3.1 Técnicas de Visualização de Dados

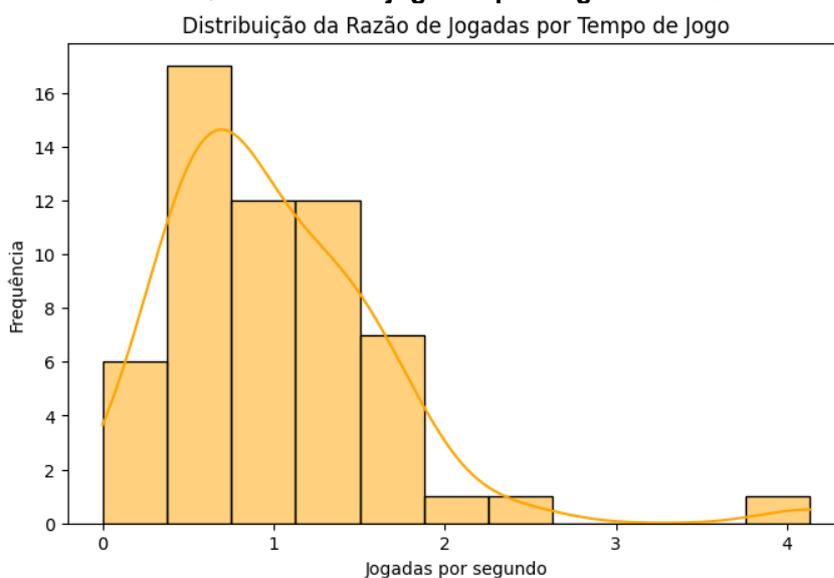
A fase do Labirinto foi jogada 57 vezes, com uma média de 1,06 partida por jogador. No total, a fase foi jogada por 5 horas, 52 minutos e 26 segundos. Em média, cada partida teve uma duração de 6 minutos e 41 segundos. Além disso, a fase apresentou uma média de 11 *clicks* por partida.

O Gráfico 16 mostra a relação de vitórias e derrotas na fase do Jogo da Memória. Das 57 partidas jogadas nesta fase, 70,2% partidas terminaram em vitória para o jogador, enquanto 29,8% partidas registraram uma derrota, ou seja, o robô saiu vitorioso.

Gráfico 16 – Distribuição de Vitórias e Derrotas no Labirinto

Fonte: Autoria própria (2024)

Para esta fase, foi calculada a relação entre o número de jogadas e o tempo utilizado pelos jogadores para concluir uma partida. Esta relação traz como resultado a quantidade de jogadas por segundo de cada partida. O Gráfico 17 apresenta a distribuição desta variável, para todas as partidas. Assim como os outros histogramas apresentados neste trabalho, as barras verticais representam a frequência de partidas com a razão de jogadas em cada intervalo apresentado no eixo x, enquanto a linha representa a distribuição de partidas de cada frequência.

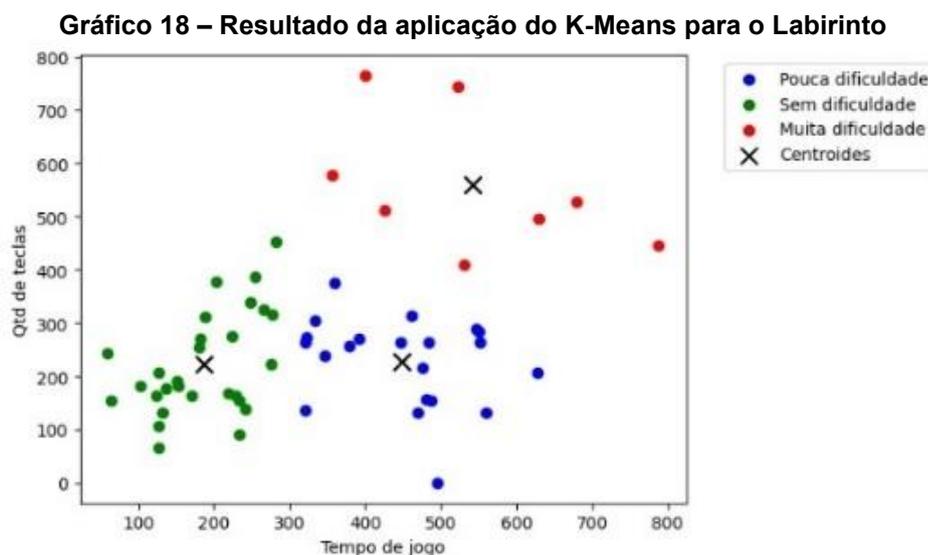
Gráfico 17 – Quantidade de jogadas por segundo no Labirinto

Fonte: Autoria própria (2024)

4.1.3.2 Algoritmos de Aprendizagem de Máquina

Na fase do Labirinto, foi utilizado apenas o algoritmo de aprendizagem não-supervisionada K-Means, com a finalidade de separar as partidas em grupos por grau de dificuldade dos jogadores. Nesta fase, a coleta não retornou a razão de teclas válidas ou o tempo entre as jogadas. Portanto, foram escolhidas outras variáveis: a quantidade total de jogadas e o tempo total de cada partida. Estas variáveis também permitem explorar o entendimento não só das regras da fase, como da agilidade que os jogadores tiveram para percorrer o labirinto.

O resultado da aplicação do K-Means, na fase do Labirinto, é apresentado no Gráfico 18, um gráfico de dispersão que mostra todas as partidas, cada uma representada por um ponto no plano cartesiano. O eixo *x* representa a razão de *clicks* válidos, enquanto o eixo *y* representa a média de tempo entre cada jogada. Os centroides de cada *cluster* são representados por um caracter 'x', na cor preta.



Fonte: Autoria própria (2024)

Cada *cluster* é representado por uma cor diferente: 28 partidas foram colocadas no *cluster* C0 (sem dificuldade), mostrado em verde, que apresenta mais valores com percentual alto de *clicks* válidos, 21 registros foram colocados no *cluster* C1 (pouca dificuldade), em azul, e 8 registros foram colocados no *cluster* C2, em vermelho, jogadores que tiveram mais dificuldade para realizar a fase.

4.1.4 Resultados da Análise Geral do Jogo

Esta subseção apresenta os resultados da análise geral do jogo sério, que envolveu todas as bases de dados obtidas na aplicação do jogo. Esta análise apresenta resultados relacionados a variáveis comuns a todas as fases e telas do jogo, como número de *clicks*, quantidade de acessos a cada tela, tempo médio do jogador no jogo, entre outras.

As perguntas propostas nesta etapa da análise de dados foram selecionadas e agrupadas no Quadro 11.

Quadro 11 – Perguntas propostas na análise geral dos dados

Pergunta proposta na análise	Variáveis da base envolvidas
Qual foi a fase mais jogada?	Número de partidas, tempo médio por jogador, número médio de partidas por jogador, quantidade de tempo absoluto de cada fase.
Quais foram as telas mais e menos acessadas pelos jogadores?	Número absoluto de acessos, tempo médio de cada tela, média de acessos por jogador.
Quais telas tiveram mais interações via mouse/teclado?	Número de <i>clicks</i> , número de teclas pressionadas.
Qual fase teve mais vitórias? Qual teve mais derrotas? Por quê?	Número de vitórias/derrotas absoluto, tempo empregado em cada fase.
Qual fase teve mais tempo na interface de narração? Alguma data comemorativa chamou mais atenção dos alunos?	Tempo médio na interface de narração de cada fase.
Qual foi a taxa de acerto dos jogadores em cada fase? Eles entenderam como funcionava a interação via mouse/teclado?	Número de <i>clicks</i> e teclas válidos, em cada fase.
Existe alguma relação entre o número de jogadas e o tempo empregado pelos jogadores em cada fase?	Número de jogadas, tempo jogando cada fase.
Qual foi a interface mais utilizada pelos jogadores em cada fase: a interface de narração, a de jogo ou a interface de <i>feedback</i> da partida?	Tempo médio e absoluto em cada interface do jogo.
Quem realizou mais jogadas em cada fase: o jogador ou o robô?	Número de jogadas do jogador e do robô em cada fase.

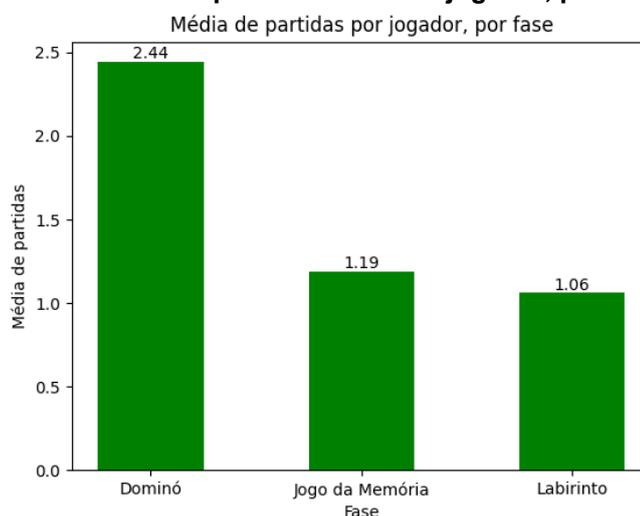
Fonte: Autoria própria (2024).

4.1.4.1 Técnicas de Visualização de Dados

Observou-se que a fase mais acessada pelos jogadores foi a do Dominó, que também era a primeira fase do jogo, com 152 partidas registradas: em segundo lugar, a fase do Jogo da Memória teve 77 partidas e, por último, a fase do Labirinto, com 57

partidas. Esta sequência também se reflete quando se analisa a fase mais jogada em média, por jogador. O Gráfico 19 mostra a ordem das fases mais jogadas em média.

Gráfico 19 – Tempo médio de cada jogador, por fase



Fonte: Autoria própria (2024)

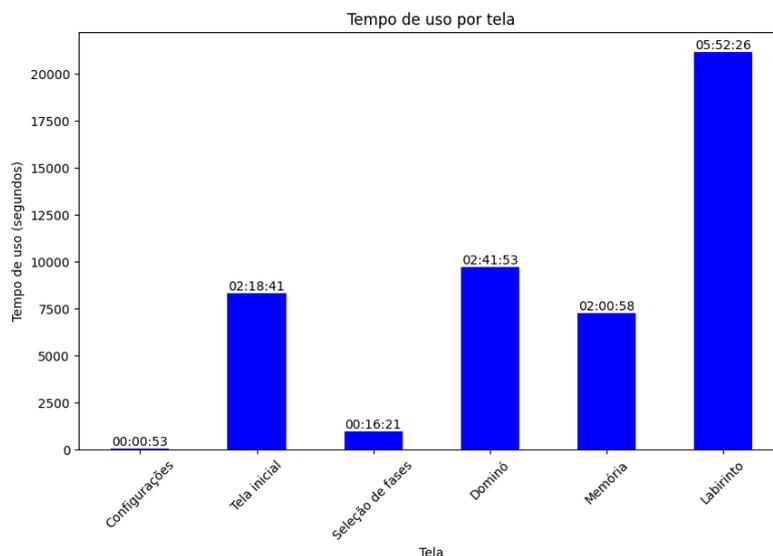
A Tabela 5 mostra o número absoluto de acessos em cada uma das telas do jogo, não só as fases, como também telas auxiliares, como as telas de seleção de fases, tela inicial e configurações.

Tabela 5 – Resultado da coleta e do tratamento dos dados do jogo sério

Aba da planilha	Tipo de tela	Quantidade de acessos
Seleção de Fases	Auxiliar	265
Dominó	Fase do jogo	152
Tela Inicial	Auxiliar	116
Jogo da Memória	Fase do jogo	77
Labirinto	Fase do jogo	57
Tela de Configurações	Auxiliar	11

Fonte: Autoria própria (2024)

O Gráfico 20 mostra a quantidade de tempo total em cada tela. A tela mais acessada, quando se fala de tempo, é a fase do Labirinto, com um total de 5 horas e 52 minutos. A tela com menor tempo total continua sendo a tela de Configurações, com apenas 53 segundos.

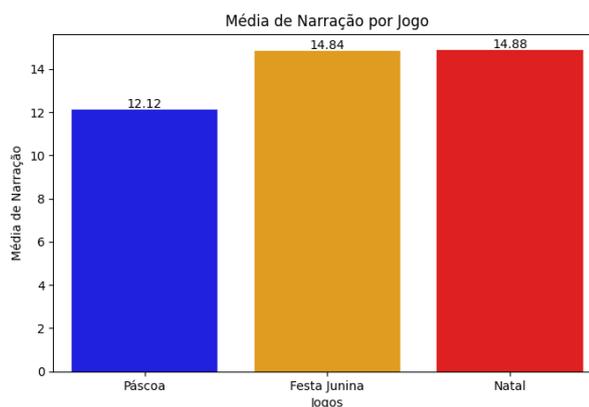
Gráfico 20 – Tempo total de uso em cada tela do jogo

Fonte: Autoria própria (2024)

Para responder à pergunta “Qual tela teve mais interações via mouse/teclado?”, foi levantada a quantidade média de *clicks* em cada tela. Em relação à interação apenas via teclado, a fase do Labirinto se mostrou a com mais interações. No que se refere à interação via mouse, a tela com mais interações deste tipo, em média, foi a tela do Jogo da Memória, com 24,12 interações. Em segundo e terceiro lugar, ficaram as outras fases do jogo: Dominó e Labirinto, com 12,99 e 11,6 *clicks*, respectivamente. Em relação às telas auxiliares, a tela de Configurações teve 3,73 *clicks* médios, enquanto as telas de Seleção de Fases e Inicial mostraram 1,35 e 1,05 *clicks*.

Em relação às fases, também pretendeu-se explorar quais fases apresentaram mais vitórias ou derrotas, em quais fases os jogadores passaram mais tempo na interface de narração, entre outras perguntas apresentadas anteriormente, no Quadro 11. O número total de vitórias dos jogadores nas fases foi de 176 partidas, contra 110 partidas que os jogadores perderam.

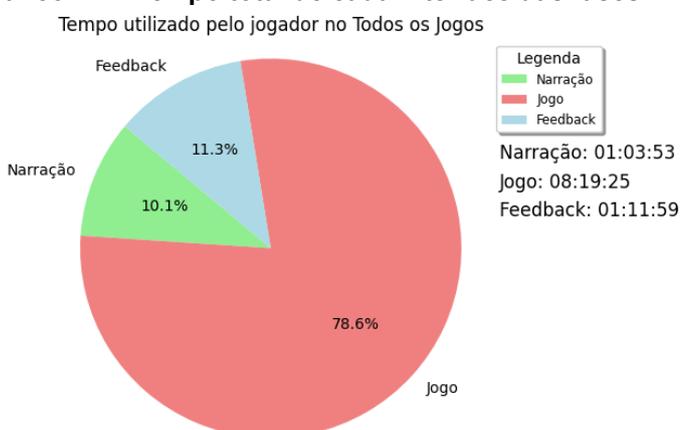
O Gráfico 21 mostra o tempo médio de cada jogador na interface de narração para cada fase: o Dominó é representado pela Páscoa, o Jogo da Memória é representado pela Festa Junina e o Labirinto é representado pelo Natal.

Gráfico 21 – Quantidade de vitórias e derrotas por fase

Fonte: Autoria própria (2024)

Com a finalidade de responder à pergunta “Qual foi a taxa de acerto dos jogadores em cada fase? Eles entenderam como funcionava a interação via mouse/teclado?”, as fases com interação via mouse (Dominó e Jogo da Memória) foram analisadas. Os Gráficos 3 e 10, apresentados anteriormente, ilustram esta relação.

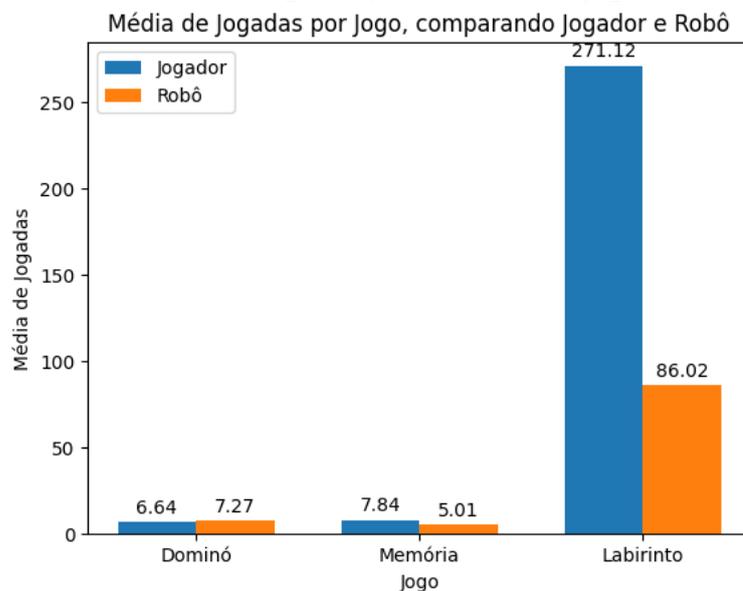
O jogo apresenta três tipos de interface para cada fase: a interface de narração, a interface principal e a interface de *feedback* da partida, que poderia ser de vitória ou de derrota. Foi possível perceber que a interface que ficou mais tempo apresentada para o jogador foi a interface principal, na qual o aluno realiza as jogadas, como mostra o Gráfico 22. Em segundo, a interface mais apresentada foi a de *feedback*. Por último, a interface de narração foi a menos apresentada.

Gráfico 22 – Tempo total de cada interface das fases

Fonte: Autoria própria (2024)

Por fim, responde-se à pergunta “Quem realizou mais jogadas em cada fase: o jogador ou o robô?”. O Gráfico 23 mostra a relação entre o número de jogadas médio dos jogadores e do robô em cada fase.

Gráfico 23 – Média de Jogadas por fase, entre o jogador e o robô

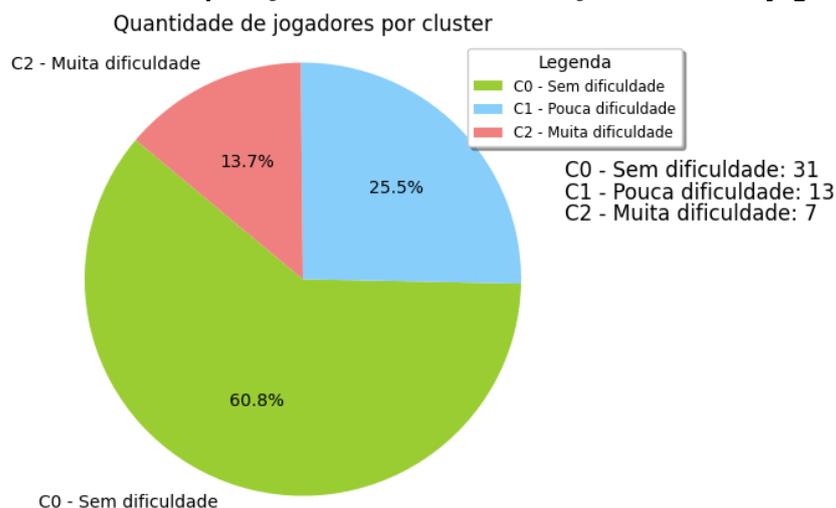


Fonte: Autoria própria (2024)

4.1.4.2 Algoritmos de Aprendizagem de Máquina

Nesta análise, foi utilizado apenas o algoritmo K-Means, para separar não só as partidas, como também os alunos em três grupos, por grau de dificuldade em cada partida. Este agrupamento dos alunos em *clusters* foi possível, visto que cada partida possui um *token*, e cada *token* representa um aluno, na base de dados geral. Para cada jogador, foram resgatados os *clusters* das partidas que ele jogou, utilizando seu *token* único. Estes *clusters* foram agrupados em uma lista de *clusters* e a moda deste conjunto (elemento que mais aparece na lista) foi armazenada junto ao *token* do jogador.

O Gráfico 24 mostra o resultado final da análise do K-Means para todos os jogadores. Por meio desta análise foi possível confirmar que a maioria dos alunos (31) teve dificuldade para jogar as três fases. Além disso, 13 alunos apresentaram um pouco de dificuldade nas fases, enquanto 7 alunos, ou 13,7%, mostraram muita dificuldade nas fases do jogo.

Gráfico 24 – Resultado da aplicação do K-Means em relação a todos os jogadores

Fonte: Autoria própria (2024)

Após a aplicação do algoritmo K-Means, o *cluster* atribuído a cada partida foi integrado e registrado nas bases de dados originais de cada fase. Essa análise proporcionou uma compreensão sobre a dificuldade relativa das partidas, tanto em relação às fases individuais quanto aos jogadores específicos.

4.2 Discussão e Comparação dos Resultados

Esta seção apresenta discussões e comparações sobre os resultados obtidos, gerando hipóteses em relação ao jogo sério e às partidas realizadas pelos alunos.

Em relação às três fases do jogo sério, observou-se que a fase mais acessada pelos jogadores foi a do Dominó, que também era a primeira fase do jogo, com 152 partidas registradas: em segundo lugar, a fase do Jogo da Memória teve 77 partidas e, por último, a fase do Labirinto, com 57 partidas. Esta sequência corresponde à ordem na qual a seleção de fases foi implementada: o jogo permite que o jogador jogue a próxima fase apenas se ele vence as fases anteriores. Por isso, a fase mais jogada foi a primeira (Dominó), que tem um número de partidas maior do que as duas últimas fases somadas. Esta sequência também se reflete quando se analisa a fase mais jogada em média, por jogador.

Entretanto, quando se leva em consideração o tempo empregado pelos jogadores em cada fase, a sequência se altera: a fase mais jogada, em relação ao tempo, foi o Labirinto, que contou com uma quantidade menor de partidas, porém partidas mais longas. A fase do Labirinto foi jogada por 5 horas e 52 minutos, somando os tempos de todos os jogadores. A fase do Dominó foi a segunda mais jogada, com

2 horas e 42 minutos. Por último, o Jogo da Memória foi jogado por 2 horas e 1 minuto.

Esta tendência também se mantém quando se obtém o tempo médio que cada jogador passou em cada fase. A fase do Labirinto apresentou um tempo médio de 401 segundos por jogador (6 minutos e 41 segundos), enquanto as fases do Dominó e do Jogo da Memória apresentaram tempo médio de 151 e 114 segundos, respectivamente.

Em relação ao número total de acessos, como apresentado na Tabela 5, pode-se concluir que a fase de configurações, responsável por controlar o volume dos sons e das músicas do jogo, não foi tão acessada, já que a maioria dos alunos se sentiu confortável com o volume, ou ajustaram o volume pelo sistema operacional do próprio computador.

Em contrapartida, as telas de Seleção de Fases e Tela Inicial tiveram mais acessos que as fases do Labirinto e do Jogo da Memória, por serem mais importantes: a Tela Inicial é a primeira interface apresentada ao jogador, e a tela de Seleção de Fases é mostrada logo depois do fim de uma partida, em todas as fases. Por isso, esta tela teve mais acessos.

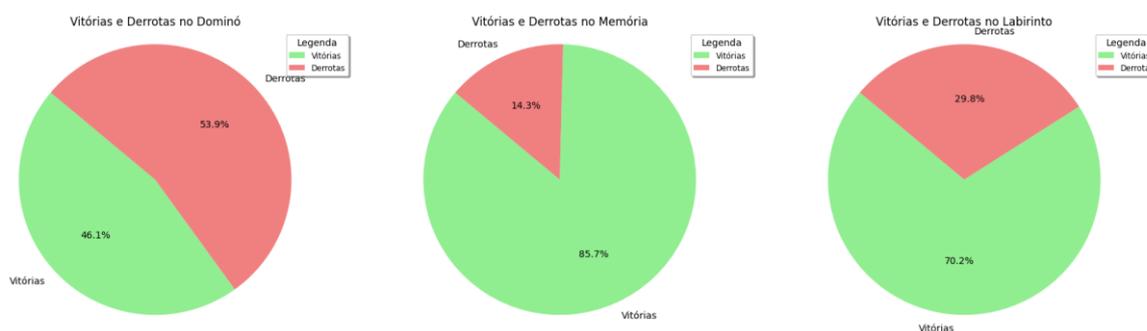
Assim como a contagem das fases, a ordem de acesso entre as telas se altera quando a variável estudada é o tempo total que os jogadores passaram em cada fase, como foi apresentado no Gráfico 20. Mesmo que a tela de Seleção de Fases tenha tido mais acessos, o tempo de cada acesso correspondia apenas a alguns segundos, para que o jogador pudesse clicar na próxima fase. A tela mais acessada, quando se fala de tempo, é a fase do Labirinto, com um total de 5 horas e 52 minutos. A tela com menor tempo total continua sendo a de Configurações, com apenas 53 segundos.

Para responder à pergunta “Qual tela teve mais interações via mouse/teclado?”, foi levantada a quantidade média de *clicks* em cada tela. Em relação à interação apenas via teclado, a fase do Labirinto se mostrou a com mais interações, já que era a única tela do jogo que era capaz de receber este tipo de interação.

Portanto, em relação à interação via *mouse*, a tela com mais interações deste tipo, em média, foi a tela do Jogo da Memória, com 24,12 interações. Em segundo e terceiro lugar, ficaram as outras fases do jogo: Dominó e Labirinto, com 12,99 e 11,6 *clicks*, respectivamente. As telas auxiliares apresentaram uma quantidade menor de *clicks* médios, por serem telas com menos possibilidade de interação do que as telas de fases. A tela de Configurações teve 3,73 *clicks* médios, enquanto as telas de Seleção de Fases e Inicial mostraram 1,35 e 1,05 *clicks*.

Em relação às fases, também pretendeu-se explorar quais fases apresentaram mais vitórias ou derrotas. O número total de vitórias dos jogadores nas fases foi de 176 partidas, contra 110 partidas que os jogadores perderam. A relação de vitórias e derrotas em cada fase é apresentada no Gráfico 25.

Gráfico 25 – Quantidade de vitórias e derrotas por fase



Fonte: Autoria própria (2024)

É possível observar que a fase que apresentou mais vitórias foi o Jogo da Memória. Em contrapartida, a fase que apresentou mais derrotas foi o Dominó, que também foi a fase mais jogada, em relação ao número de acessos. A fase do Labirinto foi a mais jogada em questão de tempo, mas apresentou um número de vitórias maior do que o de derrotas, assim como o Jogo da Memória.

Além da quantidade de vitórias e derrotas, também foram levantados os placares mais frequentes, tanto nas vitórias quanto nas derrotas de cada fase. O resultado e a discussão destes placares se encontram na Tabela 6.

Tabela 6 – Placares mais frequentes em cada fase

Fase	Placar mais frequente em vitórias	Placar mais frequente em derrotas	Explicação
Dominó	0 x 2	2 x 0	O placar do Dominó representa o número de peças na mão do jogador e do robô no final da partida.
Jogo da Memória	5 x 0	1 x 4	O placar do Jogo da Memória representa o número de pares obtidos pelo jogador e pelo robô no tabuleiro.
Labirinto	10 x 3	5 x 10	O placar do Labirinto representa o número de itens na árvore do jogador e do robô, respectivamente.

Fonte: Autoria própria (2024).

Na fase do Dominó, os placares mais comuns foram aqueles que o jogador ou o robô ficaram sem cartas, deixando o adversário com 2 cartas na mão. Na fase

do Jogo da Memória, o placar mais comum nas vitórias foi quando o jogador obteve todos os pares do tabuleiro, enquanto o placar mais comum na derrota teve pelo menos um par obtido pelo jogador. O Labirinto apresentou uma variedade maior de placares, entre 10 x 0 até 10 x 8, ou seja, 10 itens obtidos pelo jogador e de 0 a 8 itens obtidos pelo robô, sem algum tipo de padrão a ser observado.

Com a finalidade de responder à pergunta “Qual foi a taxa de acerto dos jogadores em cada fase? Eles entenderam como funcionava a interação via mouse/teclado?”, as fases com interação via mouse (Dominó e Jogo da Memória) foram analisadas. A variável estudada nesta pergunta foi a razão entre *clicks* válidos (*clicks* em peças) e número de *clicks* total.

No Dominó, foi possível perceber que, mesmo sendo a fase mais jogada, o que poderia resultar em partidas com mais *clicks* não válidos, o Dominó teve, em sua grande maioria, partidas com um alto percentual de *clicks* válidos. Logo, os jogadores entenderam a interação via *mouse* desta fase.

O Jogo da Memória apresentou menos partidas com percentual de *clicks* válidos próximos a 100%, porém o resultado ainda foi considerado como ótimo: a distribuição da razão de *clicks* ficou mais concentrada entre 80% e 100% de *clicks* válidos.

Por fim, responde-se à pergunta “Quem realizou mais jogadas em cada fase: o jogador ou o robô?”. Para esta pergunta, foi levantada a média do número de jogadas do jogador e do robô em cada fase, apresentada anteriormente no Gráfico 23. A única fase que apresentou um número de jogadas maior por parte do robô foi o Dominó, por ser uma fase que apresenta mais oportunidades de jogada para o robô.

As fases do Jogo da Memória e do Labirinto mostraram mais oportunidades para o jogador realizar suas jogadas: no Jogo da Memória, que teve mais vitórias, os jogadores obtiveram mais pares e, por isso, ganhavam uma jogada extra; no Labirinto, os jogadores tinham um número infinito de jogadas, desde que pegassem os itens no tabuleiro.

A aplicação dos algoritmos de aprendizagem de máquina também trouxe algumas comparações e *insights* para esta análise. Em primeiro lugar, o uso da regressão linear possibilitou detectar alguns padrões de quais são as piores jogadas que aconteceram em cada partida, assim como a ordem nas quais elas aconteceram. Em segundo lugar, o uso do K-Means possibilitou agrupar jogadores com grau de dificuldade parecido em *clusters* iguais, o que permite mostrar quantos jogadores

tiveram mais dificuldade com o jogo.

Em relação à regressão linear aplicada no Dominó, como mostram o Gráfico 6 e a Tabela 3, foi possível observar que o tipo de pior jogada mais frequente foi a jogada (6,6), com 31 registros, na qual a melhor peça a ser descartada é uma peça dupla, que também foi a peça descartada pelo jogador. Apresentaram apenas jogadas ótimas 23 partidas, ou seja, não foi possível encontrar uma jogada ruim nessas partidas.

A maioria das piores jogadas foram aquelas que tiveram pontuação esperada e obtida iguais, como (0,0), (1,1), (2,2), (4,4) e (6,6). Mesmo sendo as melhores jogadas naquele momento, estas jogadas foram classificadas como piores, em comparação com as outras, possivelmente pela ordem nas quais elas aconteceram: se a peça descartada na pior jogada fosse colocada em uma jogada anterior, ou em uma jogada futura, as chances de vencer aquela partida poderiam aumentar.

Sobre a ordem das piores jogadas no Dominó, foi possível perceber que as jogadas classificadas como piores ocorrem, em sua maioria, no início do jogo, nas primeiras, segundas ou terceiras jogadas. Isso se deve ao fato de que geralmente, no início da partida, o jogador é apresentado com mais possibilidades de jogada do que no final, o que abre uma margem maior para jogadas não ótimas.

Em relação ao cálculo do MSE, o qual apresentou a média de 0,65 para vitórias e 1,78 para derrotas no Dominó, o fato de a média do MSE em vitórias ter sido consideravelmente menor que a média do MSE em derrotas mostra que jogadas mais próximas ao modelo de regressão linear, em média, resultaram em vitórias para o jogador. Em contrapartida, as derrotas apresentaram jogadas com uma distância maior do modelo.

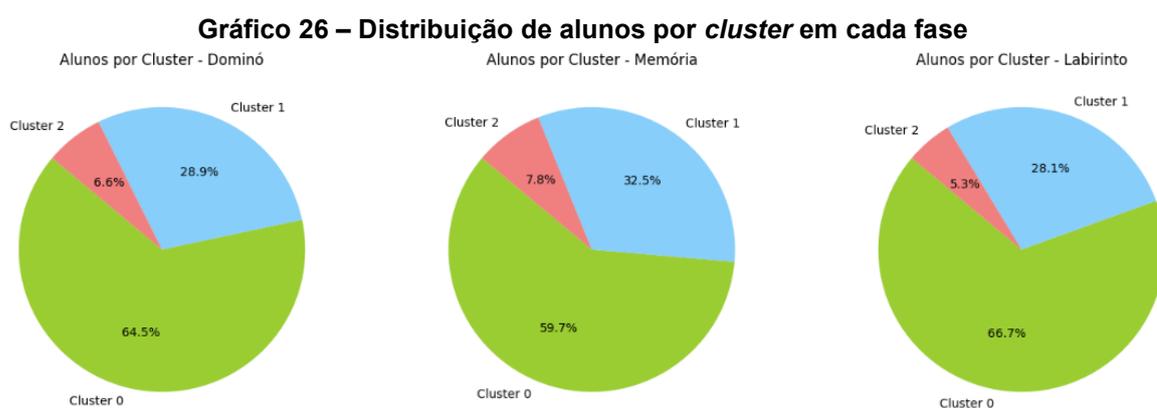
Sobre a regressão linear aplicada às partidas do Jogo da Memória, o tipo de jogada que apresentou maior frequência como a pior jogada, nas partidas, foram aquelas jogadas que a pontuação esperada foi diferente da obtida, com 23 registros no ponto (7,4), 11 registros no ponto (7,1) e 6 registros no ponto (10, 1). O ponto (10, 10), que representa uma jogada de par, também apresentou um grande número de registros, possivelmente por partidas que apresentaram muitas jogadas erradas e poucas jogadas de par. Este tipo de partida altera o modelo de regressão para se adequar às jogadas ruins, deixando a melhor jogada, que ocorreu poucas vezes, mais distante do modelo.

Sobre a ordem das piores jogadas no Jogo da Memória, apresentada no

Gráfico 14, a distribuição das piores jogadas acontece aproximadamente na metade do jogo, entre a terceira e a quarta jogada. São jogadas nas quais o jogador provavelmente já explorou metade das cartas do tabuleiro e acabou não memorizando alguma delas. Por este motivo, são poucas as piores jogadas que acontecem no início ou no final do jogo. O início do jogo apresenta poucas jogadas ruins porque, na primeira jogada, todas as cartas são inéditas, ou seja, a probabilidade de ter uma jogada ruim aqui é menor. No final do jogo, são poucas as cartas que não foram vistas, o que também diminui a probabilidade de se ter uma jogada ruim neste momento do jogo.

Em relação ao cálculo do MSE nesta fase, a tendência da média do MSE ser maior nas derrotas do que nas vitórias se manteve, assim como foi observado no Dominó. As partidas que resultaram em vitória foram capazes de gerar, em média, um modelo que se adequou melhor às jogadas, enquanto as derrotas apresentaram um erro quadrático um pouco maior.

Sobre a aplicação do K-Means nas partidas das três fases do jogo sério, foi possível perceber que, em todas as fases, como mostra o Gráfico 26, o *cluster* com mais registros sempre foi o *cluster* em verde: são os *clusters* que possuem menor tempo entre jogadas, menor tempo total e maior percentual de *clicks* válidos. Além disso, os *clusters* com menos registros foram os *clusters* vermelhos, que representam maior quantidade de teclas, tempo maior para concluir as fases e menor percentual de *clicks*.



Fonte: Autoria própria (2024)

A partir desta análise, também foi possível perceber que a maioria dos jogadores não teve muita dificuldade para jogar todas as fases, como mostrado no Gráfico 24. Foi possível confirmar que a maioria dos alunos (31) não teve dificuldade para jogar as três fases. Além disso, 13 alunos apresentaram um pouco de dificuldade

nas fases, enquanto 7 alunos, ou 13,7%, mostraram muita dificuldade nas fases do jogo.

4.3 Considerações finais

Este capítulo apresentou, na Seção 4.1, os resultados obtidos na análise dos dados coletados a partir da aplicação do Jogo das Datas Comemorativas. Esta análise foi dividida em Análise Geral, que apresenta resultados relacionados à execução do jogo do início ao fim, além de resultados relacionados às telas auxiliares do jogo e comparação entre variáveis das três fases: Dominó, Jogo da Memória e Labirinto.

A apresentação dos resultados utilizou gráficos e tabelas para representar as informações obtidas de forma visual e clara. Além disso, foram utilizados algoritmos de aprendizagem de máquina para gerar hipóteses relacionadas ao jogo.

Após a aplicação do jogo, foram percebidos alguns atributos que não foram escolhidos na etapa de levantamento de variáveis e que eram importantes para a análise, como, por exemplo, as jogadas do robô.

Na Seção 4.2, os resultados apresentados foram discutidos e comparados, com a finalidade de responder a algumas perguntas apresentadas no Quadro 11, sobre o jogo sério, suas fases e as jogadas realizadas pelos jogadores.

Após a análise, foi percebido que algumas mudanças podem ser feitas no jogo, como deixar fases que são mais fáceis no início do jogo. Neste trabalho, uma mudança seria alterar a temática do jogo do Dominó para a Festa Junina, pois, como observou-se na análise, os alunos tiveram mais dificuldades no Dominó do que no Jogo da Memória, que tem como temática a Festa Junina (segunda fase).

5 CONCLUSÃO

Este trabalho aplicou técnicas de análise de dados e algoritmos de aprendizagem de máquina, como a regressão linear e o K-Means nos dados de partidas de um jogo sério com tema datas comemorativas, voltado para pessoas com deficiência intelectual. Esta análise foi feita com o intuito de descobrir características sobre cada fase do jogo e encontrar padrões de partidas entre os jogadores, como qual foi a fase que os alunos tiveram mais vitórias, qual foi a estratégia utilizada em cada fase, entre outros aspectos.

Foi mostrado que, para este jogo, a análise de dados se mostrou como uma alternativa aos métodos tradicionais de avaliação de um jogo educacional, como pesquisas de opinião e questionários.

Após a etapa de desenvolvimento do jogo sério, foram levantados os atributos de cada fase que foram importantes na análise. Estes atributos foram identificados ao observar a estrutura de cada fase e a finalidade de cada atributo.

Além disso, o jogo sério foi aplicado em turmas de uma instituição parceira do Projeto de Extensão intitulado Uso da Inteligência Artificial em Aplicações Computacionais da Universidade Tecnológica Federal do Paraná – Campus Ponta Grossa. Os alunos jogaram todas as fases mais de uma vez, o que permitiu a coleta dos dados necessários para a análise. Estes dados foram salvos em uma planilha disponível na nuvem.

A análise ocorreu primeiramente de forma descritiva, para responder algumas perguntas relacionadas à quantidade de acessos em cada tela, tempo gasto em cada fase, número de interações via mouse/teclado, entre outras.

Além disso, algoritmos de aprendizagem de máquina foram aplicados utilizando os dados obtidos na coleta. A regressão linear foi aplicada com a finalidade de encontrar as piores jogadas de cada partida, nas fases do Dominó e do Jogo da Memória. Para todas as fases, o K-Means, um algoritmo de aprendizagem não supervisionada, foi aplicado para classificar as partidas de todas as fases em grupos, ou *clusters*, que indicam se o jogador teve algum grau de dificuldade para realizar suas partidas.

A aplicação de técnicas de organização, visualização de dados e algoritmos de aprendizagem de máquina demonstrou ser efetiva na descoberta de padrões e criação de hipóteses, com base nos dados coletados na aplicação do jogo sério.

5.1 Trabalhos futuros

Como trabalho futuro, pode-se inserir outros minijogos e temáticas no jogo. Assim, será possível realizar novas análises, levando em consideração outras fases. A partir da análise descritiva, foi possível descobrir que a primeira fase do jogo (Dominó) foi a fase que teve menos vitórias. Com essa informação, pode-se alterar a ordem das fases do jogo, deixando a fase que apresentou mais vitórias (Jogo da Memória) como a primeira.

A análise de dados também pode continuar de forma cíclica: alterações no código-fonte do jogo podem ser feitas, o jogo pode ser disponibilizado com essas alterações e outra análise pode ser realizada. Outros atributos, como a ordem de jogadas do robô e a marcação de tempo nas fases do Jogo da Memória e do Labirinto, que não foram obtidos na coleta, podem ser coletados em uma futura análise.

Outra melhoria proposta para este jogo sério é um botão de dicas, no qual o aluno pode receber um *feedback* na tela que informa para ele qual carta representa a melhor jogada, naquele momento. Por exemplo, na fase do Labirinto, a interface poderia fornecer a rota mais eficiente para um item próximo do jogador. Para jogadores que estão tendo muita dificuldade, o que pode ser obtido na análise com o algoritmo K-Means, as dicas poderiam aparecer automaticamente para o aluno.

REFERÊNCIAS

- AHMED, M. *et al.* The K-Means Algorithm: A Comprehensive Survey and Performance Evaluation. **Electronics**, v. 8, n. 9, p. 1295, 2020.
- AJIBADE, S. S.; ADEDIRAN, A. An Overview of Big Data Visualization Techniques in Data Mining. **International Journal of Computer Science and Information Technology Research**, v.4, n. 3, p. 105-113, jul./nov. 2016.
- ALONSO-FERNÁNDEZ, C. *et al.* Improving evidence-based assessment of players using serious games. **Telematics and Informatics**, v. 60, p. 101583, jul/2021.
- BNCC, 2018. Disponível em:
http://basenacionalcomum.mec.gov.br/images/BNCC_EI_EF_110518_versaofinal_sit e.pdf. Acesso em 15 jun 2023.
- CARVALHO, André C. P. L. F de; MENEZES, Angelo G.; BONIDIA, Robson P. **Ciência de Dados: Fundamentos e Aplicações**. Rio de Janeiro: Grupo GEN, 2024.
- CROMPTON, H.; BERNACKI, M.; GREENE, J. A. Psychological foundations of emerging Technologies for teaching and learning in higher education. **Current Opinion in Psychology**, v. 36, p. 101-105, dez/2020.
- DAOUDI, I. *et al.* Improving Learners' Assessment and Evaluation in Crisis Management Serious Games: An Emotion-based Educational Data Mining Approach. **Entertainment Computing**, v. 38, p. 100428, mai/2021.
- FATIMA, R. *et al.* Sharing information online rationally: An observation of user privacy concerns and awareness using serious game. **Journal of Information Security and Applications**, v. 48, p. 102351, out/2019.
- GHAHRAMANI, Z. Unsupervised Learning. *In*: BOUSQUET, O.; LUXBURG, U; RÄTSCH, G. **Advanced Lectures on Machine Learning**. Berlin: Springer Berlin, 2004. p. 72-112.
- GOLLAPUDI, S. **Practical Machine Learning**. Packt Publishing: Birmingham, 2021.
- IBM. What is Data Science? Disponível em: <https://www.ibm.com/topics/data-science>. Acesso em: 21 nov. 2023.
- KHAN, M; KHAN, S. S. Data and Information Visualization Methods, and Interactive Mechanisms: A Survey. **International Journal of Computer Applications**, v.34, n. 1, p.1-14, nov/2011.
- KIGUCHI, M.; SAEED, W.; MEDI, I. Churn prediction in digital game-based learning using data mining techniques: Logistic regression, decision tree, and random forest. **Applied Soft Computing**, v. 118, p. 108491, mar/2020.
- KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing systematic literature reviews in software engineering**. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report (2007).

KOTU, V.; DESHPANDE, B. **Data Science: Concepts and Practice**. Morgan Kaufmann: Cambridge, 2019.

LUDERMIR, T. B. Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências. **Estudos Avançados**, v. 35, n. 101, p. 85-94, 2021.

LUZ, V. S. G. da. **Método para ajuste de nível de dificuldade em jogos educacionais fundamentado em aprendizagem de máquina**. 112 p. 2021. Dissertação – Mestrado em Ciência da Computação, Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2021.

MATPLOTLIB, 2024. Disponível em <https://matplotlib.org>. Acesso em 14 mai. 2024.

MITCHELL, T. M. **Machine Learning**. McGraw-Hill Science/Engineering/Math: New York, 1997.

MORATORI, P. **Por que utilizar jogos educativos no processo de ensino e aprendizagem?**. 2003. Trabalho de Conclusão de Curso – Núcleo de Computação Eletrônica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2003.

NUMPY, 2024. Disponível em: <https://numpy.org>. Acesso em 13 mai. 2024.

PANDAS, 2024. Disponível em: <https://pandas.pydata.org>. Acesso em 14 mai. 2024.

PRIETO, L. *et al.* Uso das Tecnologias Digitais em Atividades Didáticas nas Séries Iniciais. **Renote: revista novas tecnologias na educação**, Porto Alegre, v. 3, n. 1, p.1-11, mai. 2005.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**, Pearson Education Limited: London, 2021.

SÁ, E. DE J. V.; TEIXEIRA, J. S. F.; FERNANDES, C. T. Design de Atividades de Aprendizagem que usam Jogos como Princípio para Cooperação. *In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO*. 18, 2007, São Paulo. **Anais [...]** São Paulo: USP, 2007. p. 539-549.

SALLES, G. D. *et al.* Ensino da Educação Financeira para Pessoas com Deficiência Intelectual usando o Jogo Dominó Monetário. *In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO*. 12., 2023, Passo Fundo. **Anais [...]** Passo Fundo: UPF, 2023. p. 1030-1039

SALLES, G.D.; MATOS, S. N.; BORGES, H.B. Processo de criação de um jogo sério sobre datas comemorativas para pessoas com deficiência intelectual. *In: XIII SEMINÁRIO DE EXTENSÃO E INOVAÇÃO DA UTFPR*. 13, 2023, Ponta Grossa. **Anais [...]** Goiânia: Zoe, 2023. p. 1-6.

SAVI, R.; UBRICHT, V. R. Jogos digitais educacionais: benefícios e desafios. **Revista Novas Tecnologias na Educação**, Porto Alegre, v.6, f. 2, 2008.

SCIKIT-LEARN. Scikit-Learn: machine learning in Python, 2024. Disponível em: <https://scikit-learn.org/stable/>. Acesso em 13 mai. 2024.

SILVA, Luan L. **Aplicação de Algoritmos de Aprendizagem de Máquina nos Dados de Partidas do Jogo SÉrio Querida Floresta**. 2023. Trabalho de Conclusão de Curso, Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2023.

SILVA, Luan L. *et al.* Aplicação de um Modelo Híbrido na Criação do Jogo Querida Floresta para Pessoas com Deficiência Intelectual. *In: SIMPÓSIO BRASILEIRO DE JOGOS E ENTRETENIMENTO DIGITAL (SBGAMES)*. **Anais [...]** Natal: UFRN, 2022. p. 653-662.

SOUZA, E. D. Um relato de experiência do uso de jogos educativos com um aluno com deficiência intelectual. *In: CONGRESSO INTERNACIONAL DE EDUCAÇÃO E TECNOLOGIAS; ENCONTRO DE PESQUISADORES DE EDUCAÇÃO À DISTÂNCIA*, 7, 2018, São Carlos. **Anais [...]** São Carlos: UFSCar, 2018.

TEIXEIRA, T. da S. *et al.* PegAgente: Modelagem de Agentes por Aprendizado de Reforço em Jogos Educacionais. **Revista Novas Tecnologias na Educação**, Porto Alegre, v. 18, n. 2, p. 225–234, 2021.

VETTER, Thomas R.; SCHOBBER, Patrick. Regression: The Apple Does Not Fall Far From the Tree. **Anesthesia & Analgesia**, v. 127, n. 1, p. 277-283, jul. 2018.

WU, X. *et al.* Top 10 algorithms in data mining. **Knowledge and Information Systems**, v. 14, p. 1-37, 2008.

YANG, F. An Extended Idea About Decision Trees. *In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL SCIENCE AND COMPUTATIONAL INTELLIGENCE (CSCI)*. **Anais [...]** Las Vegas, 2019. p. 349-354.

ZADEH, A. Y.; SHAHBAZY, M. A Review Into Data Science and Its Approaches in **Mechanical Engineering**. *In: INTERNATIONAL CONFERENCE ON APPLIED RESEARCHES IN SCIENCE AND ENGINEERING*. 5. **Anais [...]** Amsterdam, 2020.

APÊNDICE A - Jogo S3rio das Datas Comemorativas

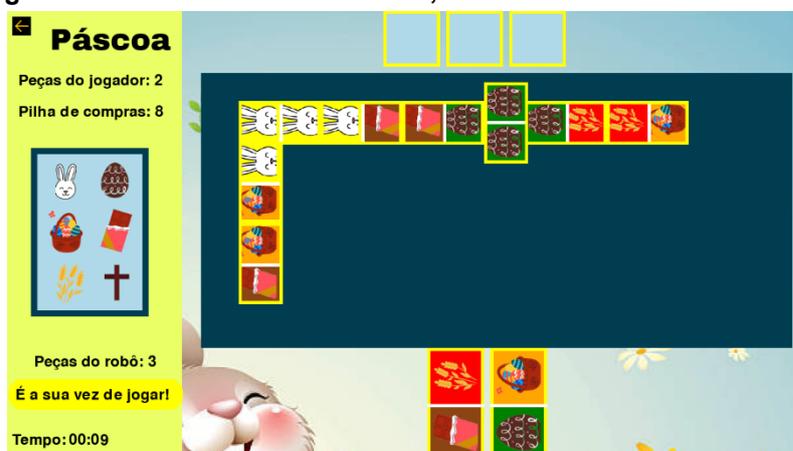
Jogo Sérió sobre Datas Comemorativas

O jogo sério Jogo das Datas Comemorativas tem como foco auxiliar no ensino das várias datas comemorativas celebradas durante o ano. O jogo conta atualmente com três fases, cada uma representando uma data comemorativa diferente: o Dominó, que tem como data comemorativa a Páscoa; o Jogo da Memória, que representa a Festa Junina.

A primeira versão deste jogo, utilizada neste trabalho, possui três fases, organizadas pela ordem na qual as datas comemorativas acontecem durante o ano. Futuramente, planeja-se construir mais fases, que tenham como tema outras datas comemorativas. Além disso, cada fase contém imagens e uma descrição da sua data comemorativa. Cada fase traz um adversário para o jogador, chamado neste trabalho de “robô”. Cada robô utiliza um algoritmo de inteligência artificial para realizar suas jogadas: estes algoritmos serão descritos a seguir.

A primeira fase traz o minijogo do Dominó, que segue as mesmas regras do dominó tradicional: tanto o aluno quanto o robô recebem um número de peças e ambos têm como objetivo ficar sem peças mais rápido que o seu oponente. Nesse jogo, as peças mostram símbolos que representam a data comemorativa desta fase, que é a Páscoa. Os símbolos escolhidos para colocar nas peças do tabuleiro do Dominó foram: a cruz, os ramos, o coelho, o ovo, a cesta de ovos e o chocolate. A Figura 11 mostra a tela da fase do Dominó.

Figura 11 – Tela da fase do Dominó, com data comemorativa Páscoa



Fonte: Autoria própria (2024)

O tabuleiro possui 21 peças, contendo todas as combinações entre os 6 símbolos. No início de cada rodada, o jogo sorteia 6 peças para o aluno e 6 peças

para o robô, deixando 9 peças na pilha de compras. Quando um dos jogadores fica sem possibilidade de jogadas na sua mão, o jogo realiza a compra de forma automática: compra uma peça para um deles e passa a vez para o outro.

Além disso, antes de começar a rodada, o jogo escolhe uma peça “dupla” (que contém dois símbolos iguais) entre o aluno e robô, entre quem tem mais peças duplas. O jogo tira uma dessas peças e passa a vez para o outro jogador, ou seja, se o robô tem mais peças duplas, o jogo escolhe uma dessas duplas para ser a peça inicial e passa a vez para o aluno. Se ninguém tem peças duplas, o jogo escolhe uma peça dupla da pilha de compras e passa a vez de forma aleatória para o aluno ou para o robô.

Para escolher sua jogada, o robô utiliza um algoritmo de busca gulosa, que recebe como parâmetros iniciais as peças que estão no tabuleiro e as peças da sua mão. O algoritmo percorre todas as jogadas possíveis e retorna aquela que possui menor custo para o robô. O custo de cada peça é representado por uma heurística, que retorna quantas peças com a mesma figura já estão na mão do robô.

Quando um dos jogadores fica sem peças na mão, o jogo acaba e o *feedback* de vitória/derrota é apresentado ao jogador: se ele ganhou, a mensagem apresentada na Figura 12 é mostrada, apresentando a possibilidade de voltar à tela de seleção de fases.

Figura 12 – Interface de vitória nas três fases



Fonte: Autoria própria (2024)

Se o robô ganhou, a tela de derrota é apresentada e o jogador pode jogar aquela fase novamente, como mostrado na Figura 13. Essa interface foi reaproveitada para as três fases.

Figura 13 – Interface de derrota nas três fases



Fonte: A autoria própria (2024)

A segunda fase traz o minijogo do Jogo da Memória, que, assim como o Dominó, não se distancia das regras tradicionais. Os símbolos trabalhados nesta fase, que tem como tema a Festa Junina, foram as comidas típicas, a fogueira, o balão de São João, as bandeirinhas e o chapéu.

O tabuleiro é composto por 10 cartas, que totalizam 5 pares de cartas, cada um contendo um dos símbolos mencionados anteriormente. O aluno sempre começa cada rodada, com o objetivo de conseguir mais pares de cartas que o seu adversário, o robô. A tela do jogo da memória é mostrada a seguir, na Figura 14.

Figura 14 – Tela da fase do Jogo da Memória, com data comemorativa Festa Junina



Fonte: A autoria própria (2024)

Cada jogada, tanto do aluno quanto do robô, é exibida na tela por 2 dois segundos, para que o aluno possa lembrar a posição de cada carta que já foi virada.

Cada rodada começa com a vez do aluno, que vira duas cartas no tabuleiro. Se as cartas viradas não correspondem a um par, o aluno passa a vez para o robô. Caso um dos jogadores ache um par de cartas iguais, ele tem a chance de jogar mais uma vez.

O algoritmo utilizado pelo robô nesta fase é uma rede neural do tipo *Perceptron* de Múltiplas Camadas, que é treinada com um conjunto de possibilidades de jogada no tabuleiro. A primeira carta na jogada do robô é escolhida de forma aleatória, enquanto a segunda carta da jogada é aquela que apresenta a melhor saída em relação à primeira carta. Para descobrir a melhor saída, os parâmetros resultantes do treino da rede neural são utilizados.

O jogo termina quando todas as cartas já foram viradas, ou seja, quando todos os pares de cartas já foram encontrados. Vence a rodada o jogador que encontrou mais pares. As interfaces de vitória e derrota são mostradas da mesma forma nesta fase.

A terceira e última fase é um labirinto, que tem como tema a data comemorativa Natal. Esta fase é um pouco diferente em relação às duas primeiras fases, já que a interação com o jogo ocorre via teclado, e não via mouse, como as fases do Dominó e do Jogo da Memória.

Nessa fase, o aluno deve percorrer um labirinto e competir com o robô para pegar o maior número possível de itens, que são enfeites para a árvore de Natal. Os itens escolhidos para esta fase, que representam o Natal, são: o Papai Noel, a caixa de presentes, a árvore, os sinos, as bolas de Natal e a bengala doce. A interface da fase do Labirinto, com tema do Natal, é apresentada abaixo, na Figura 15.

Figura 15 – Tela da fase do Labirinto, com data comemorativa Natal



Fonte: Autoria própria (2024)

O labirinto tem 22 quadrados de largura por 16 quadrados de altura, comportando 19 itens dentro dele. No início de cada rodada, o aluno é posicionado no canto superior esquerdo do tabuleiro e o robô é posicionado no canto inferior direito. A posição dos itens e das paredes do labirinto é aleatória, ou seja, cada rodada possui um labirinto diferente, para não dar vantagem a nenhum jogador.

O aluno percorre o labirinto utilizando as teclas direcionais do teclado, podendo se deslocar para cima, para baixo, para a esquerda e para a direita no labirinto: movimentos em diagonal não são permitidos. Além disso, o jogo não permite que o aluno vá para uma posição caso tenha uma parede no meio do caminho, representada na interface pela cor preta. Os quadrados em branco representam os caminhos possíveis.

O robô desloca uma posição no tabuleiro a cada 5 segundos. Este robô também utiliza busca gulosa para escolher sua próxima posição no tabuleiro: a cada jogada, a busca retorna o menor caminho até um dos itens espalhados no tabuleiro. Quando o robô obtém um item, este caminho é recalculado.

O jogo acaba quando um dos jogadores completa suas duas árvores de Natal representadas na interface, ou seja, quando um deles coleta 10 itens do tabuleiro. As interfaces de vitória e derrota são apresentadas da mesma forma que nas outras fases.