

Uma abordagem de Game Learning Analytics para identificação de habilidades de leitura e escrita no ensino infantil

José Rodrigues de Oliveira Neto¹, Rodrigo Lins Rodrigues², Americo Nobre Amorim¹

¹Escribo Inovação para o Aprendizado
Recife – PE – Brazil

²Departamento de Educação
Universidade Federal Rural de Pernambuco (UFRPE) – Recife, PE – Brazil

{jose,americo}@escribo.com, rodrigo.linsrodrigues@ufrpe.br

Abstract. *This work aims to develop an analysis of data obtained from the interaction of players in one game, out of 20, applied during a research that proved their positive impact on the development of reading and writing skills of 4-years-old children. Three classifiers algorithms were selected, which were trained with the data resulting from the interaction of these players with the game and demonstrated the hit rate of each of the classifiers. In addition, this work also makes an analysis of the interactions considered more relevant by one of the models, finding relationships between the words proposed as challenge in the test and those present in the game, raising reflections that can be taken into account during the development of a educational game that aims to improve children's reading and writing skills in early childhood education.*

Resumo. *Este trabalho se propõe a fazer uma análise de dados obtidos a partir da interação de jogadores com um jogo, dentre 20, aplicado durante uma pesquisa que comprovou o impacto positivo deles no desenvolvimento de habilidades de leitura e escrita de crianças de 4 anos de idade. Foram selecionados três algoritmos de classificação que foram treinados com os dados resultantes da interação desses jogadores com o jogo e demonstrada a taxa de acerto de cada um dos classificadores. Além disso, este trabalho também faz uma análise das interações consideradas mais relevantes na classificação de um dos modelos, encontrando relações entre as palavras propostas como desafio no teste e às presentes no jogo, levantando reflexões que podem ser levadas em consideração durante a produção de um jogo educacional que objetive aperfeiçoar habilidades de leitura e escrita de crianças no ensino infantil.*

1. Introdução

Apesar dos grandes avanços nos últimos anos, o Brasil ainda tem cerca de 11,5 milhões de analfabetos, o que corresponde a 7,0% da população de 15 anos ou mais, segundo [Agência de Notícias do IBGE 2018]. Além disso, ainda segundo [Agência de Notícias do IBGE 2018], 13 das 27 unidades da federação não conseguiram alcançar a meta do Plano Nacional de Educação (PNE) de 6.5% de analfabetismo até 2017. Todos os estados da região nordeste estão entre as 13 unidades da federação que não bateram a meta.

O fato deve ser visto como grave dado que, segundo [Equipe do Rádio pela Infância da UNICEF 2009], toda criança deve ler e escrever até os 8 anos de idade, porém o número de crianças e adolescentes de 7 a 14 anos que são analfabetos é superior a 2,3 milhões, o que corresponde a 8,3% do total. Muitos estudantes com mais de 8 anos frequentam as salas de aula, mas não sabem ler e escrever.

O número de professores no Brasil passa de 2.5 milhões, segundo [Assessoria de Comunicação Social do INEP 2018] referentes a 2017. Desse universo, 340 mil professores estavam atuando.

Paralelo a isso, diversas pesquisas têm sido feitas comprovando os efeitos positivos dos jogos no aprendizado. De acordo com [Granic et al. 2014], várias revisões já existem sobre os resultados da aprendizagem associados a jogos educativos, e uma meta-análise feita por [Vogel et al. 2006] concluiu que os jogos podem ajudar a promover avanços na reforma educacional necessários para lidar com os desafios de aprendizagem do próximo século.

Estes resultados positivos relacionados a aplicação de jogos na educação podem ser um indicio da possibilidade do uso de jogos educacionais como ferramenta em sala de aula, ajudando os professores brasileiros a lidar com a alta quantidade de alunos em sala de aula, mantendo ainda a qualidade do aprendizado de seus alunos, especialmente no que diz respeito a alfabetização.

Assim, este trabalho objetiva analisar dados gerados a partir da interação das crianças com um jogo educacional e verificar a possibilidade de desenvolver um modelo com tais dados que consiga descobrir com alta taxa de acerto se uma criança possui habilidades de escrita e leitura superior ou inferior à média esperada para sua idade.

2. Game Learning Analytics

Em seu estudo, [Amorim et al. 2018] descobriu que grupos de estudantes que interagem com os jogos consegue resultados superiores em um teste de leitura e escrita comparado ao grupo que prosseguiu com as atividades comuns da escola, e que não interagiu com os jogos.

Diante do aumento de pesquisas que buscam estudar os benefícios dos jogos em contextos educacionais, surge uma área de pesquisa intitulada Game Learning Analytics. Esta área é uma combinação de duas áreas pré existentes, chamadas Game Analytics e Learning Analytics. Ter dados sobre o que acontece enquanto o jogador está jogando é a chave pra relacionar a interação com o jogo com o aprendizado. Com tais dados em mãos, se tem uma ferramenta para ajudar a entender melhor o aprendizado observado [Freire et al. 2016].

2.1. Tipos de dados gerados por usuários

Existem dois tipos de dados gerados em serious games: dados *in situ* e *ex situ*.

Dados *ex situ* são coletados fora do sistema onde o objeto ou evento em observação está. Um questionário, por exemplo, é um tipo de dado dessa categoria uma vez que ele é tipicamente feito no mundo real e não dentro do ambiente de jogo.

Diferentemente do *ex situ*, dados *in situ* são coletados "no habitat natural ou no sistema" onde o objeto interativo vive. Portanto, nessa abordagem o objeto digital é visto

como uma caixa branca, aberta para a manipulação de conteúdo e coleta dos dados. Um bom programador pode criar estruturas e usá-las também para automatizar o processo de coleta dos dados *in-situ* [Christian Sebastian Loh 2015].

2.2. Dados de Log

Alguns jogos oferecem oportunidades de customizar sua experiência dentro dele que mudam completamente o rumo da aventura vivida pelo jogador dentro do jogo. Isso leva a um certo problema na análise e avaliação de padrões de interação, uma vez que cada experiência pode ser bastante distinta. Recentemente, pesquisadores estão estudando esse cenário através dos logs (ou registros) que o sistema gera [Erica L. Snow and McNamara 2015].

Os dados de log possuem o potencial de capturar essas múltiplas faces das decisões de usuários dentro dos jogos, o que levou pesquisadores a intencionalmente fazer com que seus ambientes de jogo registrassem todas as interações dos jogadores com o jogo. Um dos benefícios de se desenvolver um jogo com que registra as ações dos jogadores é a possibilidade de coletar dados relevantes sem interromper a interação do usuário com os jogos [Erica L. Snow and McNamara 2015].

Os dados aqui analisados foram dados de log gerados durante a pesquisa de [Amorim et al. 2018]. Cada um dos jogos coletava e enviava para um servidor os dados referentes a interação das crianças com eles, formando um log de interação.

3. Trabalhos Relacionados

Já existem diversos trabalhos relacionados à efetividade dos jogos na aprendizagem, seja na alfabetização, no ensino superior ou para o desenvolvimento de habilidades relacionadas à inteligência emocional.

Em relação à game learning analytics aplicados a jogos usados no contexto escolar, [Fernández 2017] fez uma análise exploratória completa de dados obtidos antes, durante e depois da interação de estudantes, um jogo educacional direcionado ao ensino de técnicas de primeiros socorros, objetivando estabelecer qual o nível de conhecimento dos estudantes sobre o tema a partir de suas interações com jogos.

E na área de educação infantil e alfabetização, [Amorim et al. 2018] desenvolveu 20 jogos, dentre eles o jogo analisado neste trabalho, aplicou-os em diversas escolas e demonstrou, através de testes *ex situ*, a relevância da interação dos estudantes com estes jogos, apontando que eles possuem impacto no desenvolvimento das habilidades de leitura e escrita dos estudantes. Além disso, fez uma análise inicial dos dados coletados *in situ* pelos jogos.

4. Metodologia

Aqui será apresentado qual dos jogos foi escolhido para ter seus dados analisados nesse trabalho, bem como a forma como os dados foram obtidos e processados, resultando em arquivos .csv que alimentaram os modelos de aprendizado supervisionado a serem testados neste trabalho.

4.1. Origem dos dados

Os dados estudados nesse trabalho foram dados de log coletados durante a pesquisa de [Amorim et al. 2018]. Seu estudo aplicou jogos educacionais em 17 escolas da região nordeste do Brasil, sendo 62 turmas e 749 estudantes. A pesquisa durou cerca de 19 semanas.

[Amorim et al. 2018] dividiu igualmente os 749 estudantes em dois grupos: um que não interagiu com os jogos (chamado de grupo de controle) e um que interagiu com os jogos digitais desenvolvidos (chamado de grupo de intervenção). Ao final da pesquisa, constatou que o grupo de intervenção desenvolveu mais as habilidades de leitura e escrita do que o grupo de controle.

A diferença de aprendizado foi aferida através da aplicação de um pré teste e de um pós teste com todos os alunos, para leitura e escrita. Nos testes de leitura, um dos pesquisadores mostrava para a criança uma palavra, que a criança deveria ler e dizer a sua maneira o que estava escrito. Nos testes de escrita, o pesquisador pronunciava uma palavra para a criança que, novamente, deveria escrever em uma folha como ela acreditava que a palavra era escrita.

Estes testes foram desenvolvidos por [Pazeto et al. 2012], e foram uma das principais formas de constatar a evolução das crianças no trabalho desenvolvido por [Amorim et al. 2018].

4.2. O Jogo Selecionado para o Experimento

Cada jogo desenvolvido pela pesquisa de [Amorim et al. 2018] é composto por uma sequência de telas. Cada tela apresentava alguma informação ou desafio ao jogador. Quando a necessidade da tela é satisfeita, ou seja, quando o desafio é vencido ou quando a informação é assimilada pelo jogador, a tela permite que o jogador avance para a próxima tela até chegar na última, que sinaliza ao jogador que o jogo foi finalizado.

O jogo selecionado para ser analisado neste estudo se chama "Gol da Aliteração", e ele possui cinco tipos de tela diferentes: abertura, instruções, mecânica, fim de nível e fim de jogo. Tem um total de 17 telas, dentre as quais 11 são de mecânica, ou seja, aquelas que apresentam algum desafio ao jogador.

As telas de mecânica são as telas que definem as perguntas aos jogadores. Nelas sempre estão definidos os seguintes elementos: A palavra alvo, as respostas, o personagem interativo, a bola e o botão de dica. Além do texto da pergunta, que é lido ao iniciar a fase. Essa mecânica foi desenhada para se assemelhar a uma cobrança de pênalti no futebol.

O texto da pergunta segue o seguinte padrão: "O que começa com o mesmo som da palavra..." e então é lida a palavra alvo. O jogador tem a possibilidade de pressionar um botão de dica, fazendo com que o texto seja lido novamente.

As respostas possíveis são sempre imagens, o jogador deve conseguir identificar que palavra representa a imagem e se essa palavra começa com o mesmo som da palavra alvo. Ao selecionar uma resposta, o jogador recebe um feedback animado de acerto ou erro.

O Gol da Aliteração foi escolhido para este trabalho por utilizar somente um tipo

de tela de mecânica, ou seja, só existe uma forma de interação entre o usuário e o jogo, mudando apenas o conteúdo desta tela de mecânica. Esse fato possibilita fazer mais facilmente a análise de palavras mais relevantes, uma vez que não existe diferença dentre um desafio e outro, somente as palavras que são propostas como desafio.

4.3. Ações consideradas

Os jogos coletavam e armazenavam em log os dados das seguintes ações do jogador:

1. Lançar: Enviado sempre que o jogador inicia um determinado jogo.
2. Visualizar: Enviado sempre que o jogador visualiza qualquer elemento dentro do jogo.
3. Sair: Enviado sempre que o jogador realiza todas as ações necessárias para finalizar uma tela.
4. Passar: Enviado sempre que o jogador seleciona uma das respostas corretas de uma tela.
5. Falhar: Enviado sempre que o jogador seleciona uma das respostas incorretas de uma tela.
6. Pontuar: Enviado sempre que o jogador finaliza um determinado jogo.

Para a mineração feita nessa pesquisa, somente a ação de lançar não foi usado, uma vez que a quantidade de vezes que o jogador iniciou uma sessão de jogo não faz parte das métricas escolhidas para a análise proposta neste trabalho.

Além dessas ações, foram adicionadas outras métricas para as telas de mecânica: Porcentagem de acertos e Acertos líquidos. Acerto líquido é uma métrica descrita por [Amorim et al. 2018] como relevante, e equivale aos acertos do jogador subtraído dos erros, podendo assim ser negativa.

Portanto, os dados extraídos por tela foram os seguintes: quantidade de acertos, quantidade de erros, quantidade de visualizações, tempo médio por tela, quantidade de acertos líquidos (acertos - erros) e porcentagem de acertos.

A partir desses dados, foram gerados 2 arquivos .csv com 83 atributos iguais. A diferença é o rótulo deles, um contém a nota do pós teste de escrita e o outro a nota do pós teste de leitura. Cada um dos arquivos .csv possui um rótulo cujo valor é 0 ou 1. O rótulo será zero se a criança conseguiu uma média menor que a esperada para sua faixa de idade no pós teste, e será 1 em caso de nota acima da média. A média esperada para a faixa de idade das crianças foi determinada a partir dos resultados do grupo de controle da pesquisa de [Amorim et al. 2018].

Cada arquivo tem 318 amostras, e dada a quantidade de atributos foi decidido fazer uso da técnica de análise de componentes principais para verificar se ela afeta positivamente os resultados, como em outras pesquisas semelhantes encontradas na literatura.

4.4. Aplicação dos Algoritmos de Classificação

A técnica de classificação foi escolhida para analisar os dados deste trabalho pela sua capacidade de satisfazer todos os objetivos propostos.

O algoritmo de *Naive Bayes* foi escolhido por estar bastante presente em trabalhos relacionados, além de, segundo [Kotsiantis et al. 2007], baseado em trabalhos de

[Domingos and Pazzani 1997], mesmo assumindo inicialmente que as variáveis são completamente independentes, o algoritmo de *Naive Bayes* consegue resultados tão bons ou até melhores que algoritmos mais sofisticados em comparações de larga escala. Nesse estudo é usada a implementação *GaussianNB* do *Scikit Learn*.

O algoritmo de Regressão Logística foi selecionado devido a um dos objetivos da pesquisa, que é encontrar quais dentre as interações coletadas dos usuários são as mais relevantes ou impactantes no resultado do pós teste, para que sejam feitas discussões do ponto de vista educacional a partir disso. O algoritmo de Regressão Logística permite encontrar facilmente quais são as características mais impactantes na sua predição, tornando esse objetivo possível. No estudo é usada a classe *LogisticRegression* do *Scikit Learn*.

O algoritmo *SVM* foi escolhido também por pesquisas na literatura. Segundo [Kotsiantis et al. 2007], a complexidade do modelo de uma *SVM* não é afetada pela quantidade de características dos dados de treinamento, dado que o número de vetores suporte selecionados pelo algoritmo é usualmente pequeno. Por essa razão, as *SVMs* são consideradas adequadas para tarefas de treinamento onde o número de características é grande em relação ao número de instâncias de treinamento. O que acaba sendo o nosso caso uma vez que temos relativamente poucos exemplos e várias características deles. Para o experimento usamos a implementação *SVC (C-Support Vector Classification)* do *Scikit Learn*.

5. Resultados

5.1. Quantidade de Componentes Principais

A quantidade de componentes foi definida depois de elaborado o gráfico de quantidade de variância dos dados por quantidade de componentes principais.

Escolhemos 40 componentes principais pela possibilidade de manter mais de 95% da variância dos dados. Os resultados para os testes de leitura e escrita que serão mostrados a seguir usam essa quantidade de componentes.

5.2. Detalhes do experimento

Na implementação *SVC* e *LogisticRegression* era possível escolher dois parâmetros de grande impacto no desempenho dos classificadores: o *kernel* para o *SVC* e o *solver* para o *LogisticRegression*. Dentre os valores disponíveis na biblioteca, usamos aqui os *kernels linear, rbf e sigmoid* e usamos os *solvers libfgs, liblinear, sag e saga*. Para o modelo *GaussianNB* foram escolhidas as configurações padrão da biblioteca.

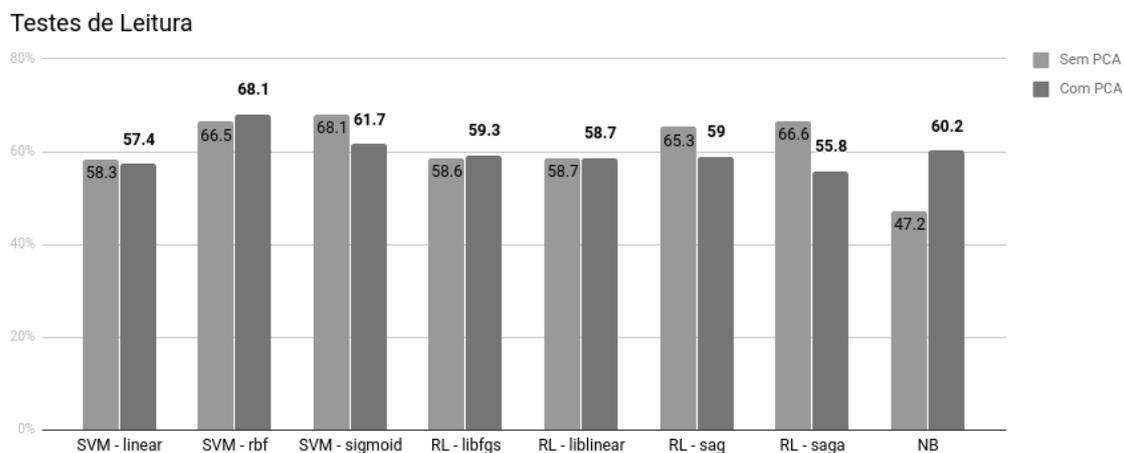
O experimento foi feito usando a técnica de validação cruzada *k-folds*, com $k = 10$ e a métrica de acurácia. Tínhamos uma amostra de 318 entradas, das quais 166 eram de rótulo 1 e 152 eram de rótulo 0. Sendo uma amostra razoavelmente balanceada, podemos usar a acurácia unicamente para medir o desempenho dos classificadores.

5.3. Resultados para os testes de leitura

Foi feita a média de acurácia para cada classificador, com e sem a técnica de Análise de Componentes Principais. Foram encontrados os seguintes resultados.

Os melhores desempenhos foram de 68.1%, para o modelo de *SVM* com o *kernel rbf* usando *PCA* e para o mesmo modelo em configurações diferentes, usando o *kernel*

Figura 1. Resultados para os testes de leitura



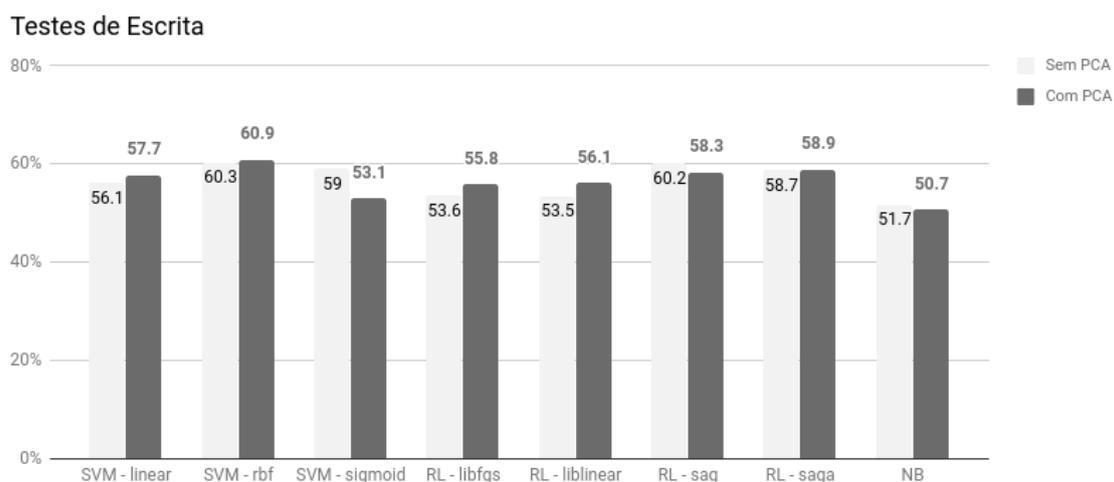
elaborada pelos autores do trabalho.

sigmoid e sem o uso dos componentes principais. Para a Regressão Logística, o melhor desempenho foi a do solver *saga* sem o uso de componentes principais, tendo 66.6% de acurácia. O modelo de *Naive Bayes* (NB) teve o pior desempenho observada sem componentes principais, mas foi o que apresentou uma maior diferença quando comparado ao desempenho sem o uso de *PCA*.

5.4. Resultados para os testes de escrita

Foi feita a média de acurácia para cada classificador, com e sem a técnica de Análise de Componentes Principais. Foram encontrados os seguintes resultados.

Figura 2. Resultados para os testes de leitura



Elaborada pelo autor do trabalho.

O melhor desempenho observado foi a da *SVM* com o *kernel rbf* usando componentes principais, com 60.9%. Para a Regressão Logística, o solver *sag* sem componentes

principais teve um desempenho relativamente próximo com 60.2% de acurácia. O modelo de *Naive Bayes* novamente apresentou o pior desempenho observado, e neste caso acabou sendo prejudicado pelo uso de componentes principais.

5.5. Atributos relevantes para a classificação positiva

Para determinar quais foram os componentes mais relevantes, foram usados os resultados obtidos pelo modelo de Regressão Logística, com os solvers *sag* para escrita e *saga* para leitura por terem sido os que obtiveram os melhores desempenhos em cada teste. Os resultados foram obtidos através dos coeficientes fornecidos pelo modelo para cada um dos atributos. Os atributos que mais fizeram o modelo classificar os exemplos como acima da média em cada teste foram os seguintes.

Tabela 1. Atributos mais relevantes para classificação acima da média

Teste de Leitura	Teste de escrita
Tempo médio de jogo na tela #7 (RAMO)	Quantidade de (acertos - erros) na tela #3 (FACE)
Tempo médio de jogo na tela #14 (PAÇO)	Quantidade de (acertos - erros) na tela #6 (CAVALO)
Tempo médio de jogo na tela #10 (BONÉ)	Tempo médio de jogo na tela #7 (RAMO)
Tempo médio de jogo na tela #4 (PIÃO)	Tempo médio de jogo na tela #5 (PIRULITO)
Tempo médio de jogo na tela #9 (FORMA)	Tempo médio de jogo na tela #2 de Instruções
Quantidade de (acertos - erros) na tela #3 (FACE)	Tempo médio de jogo na tela #14 (PAÇO)
Quantidade de (acertos - erros) na tela #6 (CAVALO)	Quantidade de (acertos - erros) na tela #9 (FORMA)
Tempo médio de jogo na tela #12 (FACA)	Quantidade de (acertos - erros) na tela #12 (FACA)
Quantidade de respostas incorretas na tela #5 (PIRULITO)	Quantidade de (acertos - erros) na tela #5 (PIRULITO)
Quantidade de (acertos - erros) na tela #15 (TESOURO)	Quantidade de (acertos - erros) na tela #15 (BONÉ)

Para os testes de leitura, 6 das 10 métricas são tempo médio em telas de mecânica. Isso pode indicar que o classificador considerou como relevante o esforço que a criança faz em focar e compreender o desafio apresentado antes de respondê-lo. Pode significar uma tentativa por parte do jogador de ler as palavras e refletir a respeito do som inicial delas e das respostas possíveis.

Para os testes de escrita, 6 entre as 10 melhores métricas serem de (acertos - erros). Esse resultado pode indicar que o classificador tende a considerar como acima da média os jogadores que já tem uma habilidade de leitura bem desenvolvida.

5.6. Atributos relevantes para a classificação negativa

Usando a mesma metodologia aplicada aos componentes mais relevantes para a classificação positiva, analisando os mesmos coeficientes desta vez para determinar os componentes que mais fazem o classificador definir o jogador como abaixo da média, temos os seguintes resultados.

Para os testes de leitura, metade das métricas mais relevantes são as de tempo médio em telas que não apresentam nenhum desafio para o jogador e que outras duas são métricas de erros em tela de mecânica. Estes resultados podem representar que o jogador pode estar perdendo o foco nos desafios ao passar por essas telas. Uma análise mais profunda dessas telas identificou que elas não possuem seus textos lidos pelo leitor embutido no jogo, o que reforça essa teoria. Essas telas só exibem um determinado quantitativo de estrelas de acordo com o desempenho do jogador, variando de 1 a 3 estrelas, e exibindo a palavra "Parabéns!" ao centro, com um pequeno botão de *play* para que o usuário prossiga para próximas telas de desafio. Já os erros nas telas de desafio podem indicar que o usuário está apenas respondendo aleatoriamente para passar as telas de desafio.

Tabela 2. Atributos mais relevantes para classificação abaixo da média

Teste de Leitura	Teste de escrita
Tempo médio de jogo na tela #17 fim de jogo	Tempo médio de jogo na tela #16 fim de nível
Tempo médio de jogo na tela #16 fim de nível	Quantidade de respostas incorretas na tela #6 (CAVALO)
Tempo médio de jogo na tela #3 (FACE)	Tempo médio de jogo na tela #8 fim de nível
Tempo médio de jogo na tela #1 de abertura	Tempo médio de jogo na tela #17 fim de jogo
Tempo médio de jogo na tela #15 (TESOURO)	Quantidade de visualizações para a tela #1
Tempo médio de jogo na tela de fim de jogo	Quantidade de respostas incorretas na tela #5 (PIRULITO)
Quantidade de respostas incorretas na tela #6 (CAVALO)	Tempo médio de jogo na tela fim de jogo
Quantidade de respostas incorretas na tela #10 (BONÉ)	Quantidade de respostas incorretas na tela #3 (FACE)
Quantidade de visualizações para a tela #1 de abertura	Quantidade de respostas incorretas na tela #4 (PIÃO)
Quantidade de (acertos - erros) na tela #4 (PIÃO)	Quantidade de visualizações para a tela #2

Para os testes de escrita, 6 das 10 métricas são de telas que não apresentam desafio ao jogador, enquanto as outras 4 métricas são de erros em telas de desafio. Isso reforça a teoria da resposta aleatória por parte do jogador levantada anteriormente. E de que as telas que não apresentam desafio podem estar fazendo o jogador perder a atenção.

6. Conclusões

Este trabalho conseguiu estabelecer uma metodologia capaz de ser aplicada para os outros 19 jogos que fizeram parte da pesquisa de [Amorim et al. 2018]. Fica confirmada a possibilidade de extração e análise dos dados de qualquer um desses jogos a partir dos arquivos de log armazenados.

Este estudo pode ser considerado como uma extensão do trabalho iniciado por [Amorim et al. 2018], uma vez que buscamos explorar com mais detalhes quais ações, ou quais palavras apresentadas como desafio para os jogadores, são mais importantes para detectar o desenvolvimento das habilidades de leitura e escrita.

O desempenho dos algoritmos escolhidos foi considerada insatisfatória. Os melhores resultados foram 60,9% para os testes de escrita e 68,1%. Tais valores não são suficientes para fazer uma substituição do pós teste.

Foram encontrados mais traços que reforçam as suspeitas levantadas no trabalho de [Amorim et al. 2018], como a relevância positiva da métrica de acertos líquidos, resultante da subtração dos erros pelos acertos. Além de uma discussão inicial relativa à quantidade de métricas predominantes dentre as que impactam positivamente o resultado do modelo em cada um dos testes: Acertos líquidos para escrita e Tempo médio para leitura.

Para as métricas que afetam negativamente a classificação, acreditamos que o jogador pode ter demorado por não ter compreendido bem o que deveria fazer quando estava em telas que não apresentavam desafio, o que pode ter tirado sua atenção da atividade. Tais telas não têm nenhuma frase lida, o que reforça essa hipótese.

Para as palavras em específico, foi encontrada uma relação relevante nos testes de escrita. Duas palavras chamaram atenção pela quantidade de métricas relacionadas a elas, e foram encontradas palavras no teste de escrita que são correspondentes por terem as primeiras sílabas iguais.

Referências

- Agência de Notícias do IBGE (2018). Analfabetismo cai em 2017, mas segue acima da meta para 2015. Disponível em: <http://bit.ly/2Zrg5lc>.
- Amorim, A. N. G. F. et al. (2018). *THE USE OF DIGITAL GAMES BY KINDERGARTEN STUDENTS TO ENHANCE EARLY LITERACY SKILLS*. PhD thesis, Johns Hopkins University.
- Assessoria de Comunicação Social do INEP (2018). Censos educacionais do inep revelam mais de 2,5 milhões de professores no brasil. Disponível em: <http://bit.ly/2U9tSvI>.
- Christian Sebastian Loh, Y. S. (2015). Measuring expert performance for serious games analytics: From data to insights.
- Domingos, P. and Pazzani, M. (1997). On the optimality of the simple bayesian classifier under zero-one loss. *Machine learning*, 29(2-3):103–130.
- Equipe do Rádio pela Infância da UNICEF (2009). Rádio pela infância - boletim de setembro de 2009. Disponível em: https://www.unicef.org/brazil/pt/rpi_setembro2009.pdf.
- Erica L. Snow, L. K. A. and McNamara, D. S. (2015). The dynamical analysis of log data within educational games.
- Fernández, C. A. (2017). *Applying data mining techniques to Game Learning Analytics*.
- Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., and Fernández-Manjón, B. (2016). Game learning analytics: learning analytics for serious games. *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, pages 1–29.
- Granic, I., Lobel, A., and Engels, R. C. (2014). The benefits of playing video games. *American psychologist*, 69(1):66.
- Kotsiantis, S. B., Zaharakis, I., and Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160:3–24.
- Pazeto, T. d. C. B. et al. (2012). Avaliação de funções executivas, linguagem oral e escrita em pré-escolares.
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., and Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3):229–243.