

Aprendizado supervisionado e não-supervisionado aplicado ao ensino de lógica em ambiente escolar

Camilla Pozer de Matos
dept. de Ciência da Computação
Instituto Federal Catarinense
Videira, Brasil
camillapozerm@gmail.com

Resumo—O presente trabalho tem a finalidade de destacar a importância do desenvolvimento lógico na infância, bem como explorar a ideia da criação de uma plataforma direcionada ao ensino da lógica nas escolas. O objetivo da plataforma é oferecer questões lógicas subdivididas em fases e categorias, as quais serão apresentadas aos alunos e respondidas de forma objetiva. A partir dos hábitos dos alunos ao utilizarem a plataforma e com o auxílio de aprendizado supervisionado e não-supervisionado, será possível entender quais alunos possuem chance de finalizar todas as questões e quais não possuem, focando em diferentes cenários para cada grupo.

Palavras-chave—lógica, sistema, aprendizado, padrões.

I. INTRODUÇÃO

Todo ser humano possui a capacidade de raciocinar, sendo que essa habilidade é aperfeiçoada conforme recebe estímulos. Apesar disso acontecer com maior intensidade na fase adulta, é na infância que o cérebro possui uma maior capacidade de absorver e aprender com esses estímulos, visto que se encontra em fase de desenvolvimento [1].

Baseado na premissa de que é na infância que devemos iniciar uma rotina de pensar logicamente, diversos autores estudaram e afirmam que a melhor maneira de introduzir a lógica na vida de uma criança é através da escola, visto que é um local de aprendizado e o aluno já está acostumado a desenvolver o máximo de suas capacidades cognitivas [2][3][4].

Seguindo as propostas abordadas pelos autores citados, a forma mais eficaz de aperfeiçoar o raciocínio é aliar a lógica ao ensino de programação, sendo que este último traz inúmeras vantagens aos estudantes. Para isto, aulas de lógica de programação foram aplicadas em escolas e comprovaram sua eficácia com ótimos resultados [2].

Apesar dos bons resultados, existem muitos impedimentos para trazer essas aulas de lógica a todas as escolas, por isso o objetivo é criar uma plataforma exclusiva para o ambiente escolar, com o intuito de eliminar a necessidade de um professor especializado e ampliar a margem de horários que os alunos terão para as aulas. Além disso, baseado no perfil de cada aluno,

decisões poderão ser tomadas de forma mais assertiva para instigar o aluno a finalizar todas as questões.

No presente artigo, busca-se trazer os melhores métodos de implementar um sistema para identificar os hábitos do aluno e classificar entre alunos concluintes e não concluintes.

II. REFERENCIAL TEÓRICO

Atualmente, a nova geração de crianças e adolescentes que frequentam a escola vivem numa realidade chamada *cybercultura*, ou seja, estão imersos na tecnologia em um nível tão elevado, que os mesmos possuem dificuldades de direcionar a atenção ao ensino, assim como fixar um novo aprendizado. Por isso, [5] ressalta a importância de aliar a tecnologia ao ensino, visto que os dois fazem parte do cotidiano dos alunos e juntando-os pode-se conseguir o máximo potencial dos discentes.

Uma forma de trazer a tecnologia, e, principalmente, a lógica para o âmbito educacional é através da programação, visto que facilita o ensino e proporciona muitos benefícios. Além de melhorar a criatividade, organização, desempenho pessoal e escolar e facilitar o aprendizado, a programação simplifica o processo de ensino de lógica, visto que torna os problemas mais tangíveis e fáceis de solucionar. Dessa forma, o aluno possui ganhos bilaterais, seja no aperfeiçoamento de seu raciocínio, assim como o aprendizado de uma linguagem de programação, que pode acarretar em uma futura profissão.

Apesar de ser o ideal, aplicar aulas de lógica com professores especializados é uma tarefa complicada, visto que não há professores suficientes na área, além de ser necessário um tempo específico para essas aulas, acarretando na modificação da grade curricular [5]. Como isso não é possível atualmente, propõe-se a criação de uma plataforma web direcionada ao ensino de lógica, sendo que no laboratório de informática das escolas, os alunos poderão acessar a plataforma e responder questões lógicas, tendo a possibilidade de acompanhar sua evolução cognitiva. Ainda, o sistema será capaz de entender o perfil de cada aluno e classificá-lo, para uma posterior validação.

O grande diferencial da plataforma será essa classificação de alunos conforme os padrões que cada grupo segue, que serão dois: concluintes e não

concluintes. Os alunos concluintes são aqueles que respondem todas as questões, utilizam a plataforma diariamente e por fim, finalizam a ideia proposta. Diferente destes, os alunos não concluintes possuem mais dificuldade de utilizar o sistema frequentemente, não respondendo às questões propostas e necessitando de mais incentivo que os outros. Dessa forma, sabendo o perfil de cada aluno será mais fácil criar conteúdos específicos para cada um.

III. METODOLOGIA

A fim de criar um sistema eficaz para identificar o perfil de cada aluno, são utilizadas técnicas de inteligência artificial, focadas em reconhecimento de padrões. Assim, com os dados de um determinado aluno, será possível identificar seu padrão de comportamento e categorizá-lo. Para isso, são testados dois modelos de reconhecimento de padrões: aprendizado supervisionado e não-supervisionado.

Para entender e avaliar o perfil dos alunos, serão utilizados alguns atributos coletados através do uso da plataforma. Nesse sentido, ainda não há dados disponíveis, pois a plataforma se encontra em fase de desenvolvimento, sendo que a coleta de dados será feita após seu lançamento e utilização pelos alunos.

Com o intuito de saber de que maneira o aluno utiliza o sistema, os atributos coletados são referentes a frequência e forma de responder às questões, como por exemplo: tempo levado para responder cada questão; tempo de permanência na plataforma em cada acesso; quantidade de questões respondidas a cada acesso; quantidade de acessos por dia; quantidade de acessos por semana; entre outras métricas que serão abordadas quando o desenvolvimento da plataforma for concluído.

Antes de utilizar os métodos de aprendizado supervisionado e não-supervisionado, a base de dados será comprimida através do PCA, realizando testes com 50%, 75%, 90% e 99% de variância. Além disso, é utilizada uma base de dados pública fictícia, com a mesma estrutura já relatada, que após a conclusão da plataforma e obtenção dos dados reais, será alterada para a base de dados válida.

IV. APRENDIZADO SUPERVISIONADO

Quando são utilizadas técnicas de aprendizado supervisionado, é necessário que haja um supervisor, ou seja, entradas e saídas para cada dado, informando a qual classe esse dado pertence. Com essa informação, o sistema consegue identificar um padrão e classificar novas entradas, que geralmente é o tipo de aprendizado mais assertivo [6].

Para criar o classificador de plataforma de ensino de lógica, serão testados diferentes métodos e funções, a fim de oferecer diversas opções de escolha para focar na melhor obtida. Todos os classificadores podem ser encontrados no *scikit-learn*, que oferece ferramentas de desenvolvimento de aprendizado de máquina em Python.

Os classificadores utilizados para testes e posterior avaliação e uso são: SVM Linear, SVM RBF, C4.5, SGDC e KNeighbors.

V. APRENDIZADO NÃO-SUPERVISIONADO

Se tratando de aprendizado não-supervisionado, os dados recebidos são referentes apenas aos atributos de cada dado, não contendo a qual classe ele pertence. Sem saber a classe, o sistema agrupa os dados com base em algum padrão encontrado, e conforme novos dados são adicionados, é calculado em qual grupo ele provavelmente deve pertencer. Com isso, deve ser feita uma análise posterior para entender o que cada agrupamento significa, visto que não há conhecimento da classe real de cada dado [6].

Igualmente a forma de avaliação citada no aprendizado supervisionado, neste caso também são testados diferentes métodos, a fim de comparar e avaliar seus resultados.

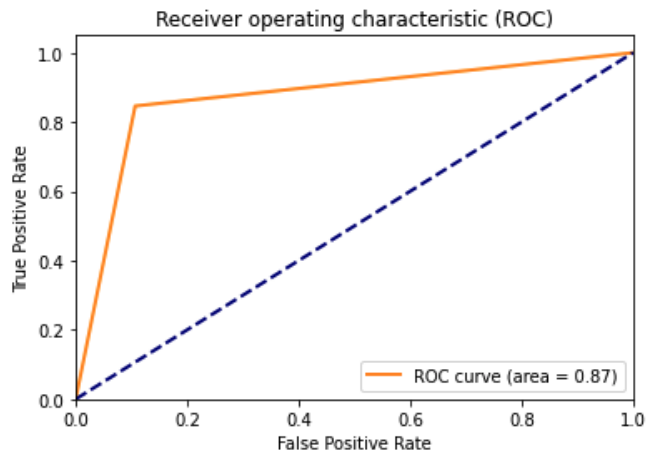
Para realizar os testes, são utilizados os métodos: K-Means e Fuzzy C-Means.

VI. RESULTADOS

Após a execução de todos os métodos para testes, a fim de escolher o mais apropriado, houveram ótimos resultados em diferentes funções, os quais serão mostrados conforme a variância do PCA.

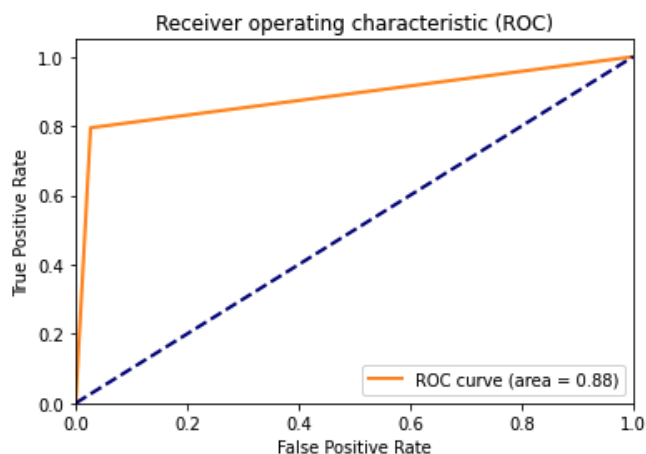
Quando a variância foi de 50% em testes de aprendizado supervisionado, apenas 5 atributos foram utilizados. Na execução de todos os métodos, os melhores foram SVM linear, que obteve uma acurácia de 92%, porém quando se analisa o TPR (taxa de verdadeiro positivo) e TNR (taxa de verdadeiro negativo), eles chegaram, respectivamente, a 79% e 98%, o que pode ser considerada uma diferença alta. No caso do método C4.5, o TPR e TNR foram praticamente semelhantes, obtendo 84% e 89%, gerando uma acurácia de 87%, o que foi o melhor resultado na variância atual. Na figura 1 pode ser visto o gráfico da curva ROC.

Figura 1. C4.5 com 50% de variância



Nos testes realizados com a variância de 75%, a base de dados foi reduzida para 11 atributos, onde o melhor método, novamente, foi o C4.5, obtendo 89% de acurácia. Em relação ao TPR e TNR, o resultado foi de, respectivamente, 87% e 90% (figura 2).

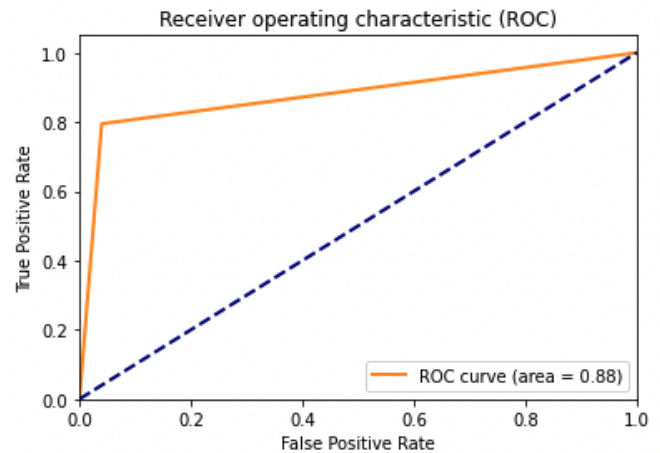
Figura 2. C4.5 com 75% de variância



Quando a variância começa a ter um grande aumento, com valor de 90% e um número de atributos de 17, o melhor método passa a ser o SGDC, com 90%, 84% e 93%, para acurácia, TPR e TNR. Porém, quando usa-se 99% de variância com 23 componentes, o C4.5 se sobressai, obtendo uma variância de 89%, um TPR de 84% e TNR de 92%.

Baseado em todos os resultados obtidos no aprendizado supervisionado, o melhor foi o SGDC com variância de 90% (figura 3), visto que obteve os melhores valores de acurácia, TPR e TNR, além de não precisar de muitos componentes, sendo que 17 é um valor razoavelmente baixo.

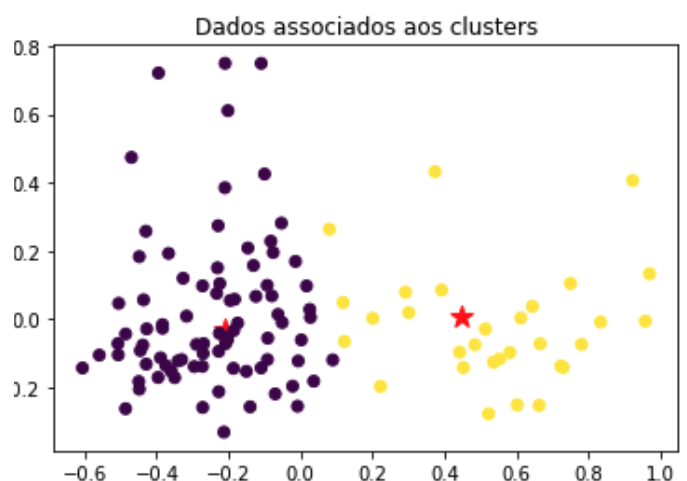
Figura 3. SGDC com 90% de variância



A fim de realizar testes com o aprendizado não-supervisionado, os resultados são demonstrados seguindo o mesmo padrão do aprendizado supervisionado. Para validar se os resultados foram bons, duas técnicas são utilizadas: critério interno, que é baseado em média silhouette e critério de Fisher; e critério externo, onde são utilizadas as classes válidas de cada dado.

Com 50% de variância e 5 atributos, o método Fuzzy C-Means foi muito melhor em comparação ao K-Means, sendo que o primeiro obteve 90% de acurácia com o método externo e o segundo apenas 13%. Se tratando do método interno de validação, o critério de Fisher foi muito baixo nos dois casos, e a média silhouette foi de 41%, o que ainda não é o ideal. Na figura 1 pode ser conferido o gráfico do método Fuzzy C-Means:

Figura 1. Fuzzy C-Means com 50% de variância



Com 75% e 90% de variância, os dados foram irrelevantes, obtendo acurácias muito baixas, como 13% e 8%. A média silhouette e critério de Fisher também não foram bons em todos os casos, mesmo

que o gráfico da figura 2 e 3 mostram que as classes estejam bem separadas:

Figura 2. K-Means com 75% de variância

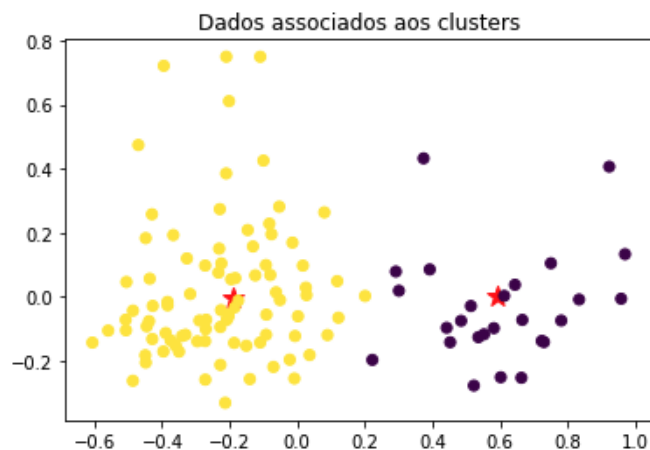
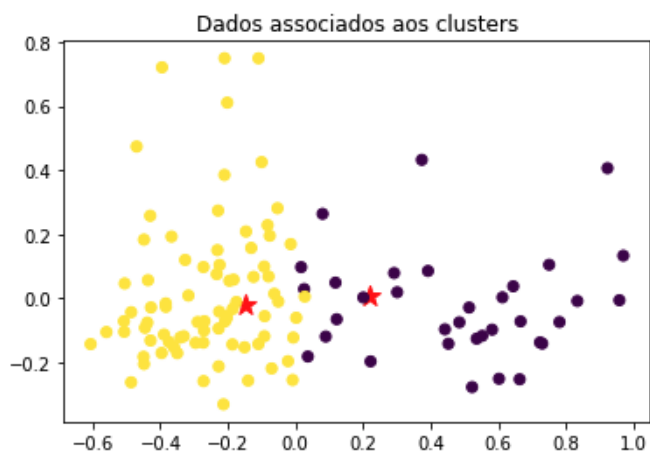
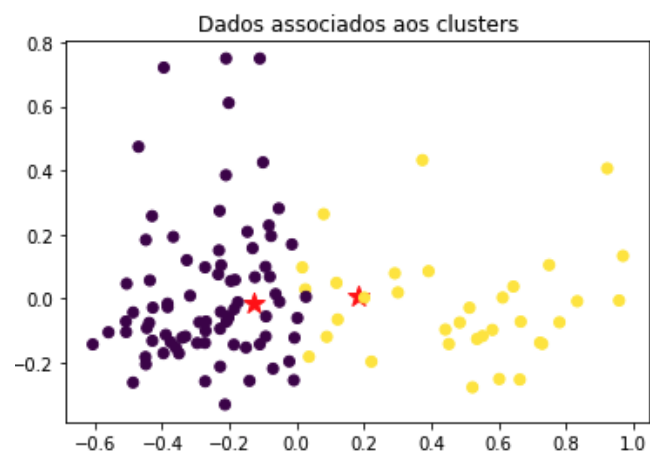


Figura 3. Fuzzy C-Means com 90% de variância



Por fim, o último teste realizado foi com 99% de variância, o que trouxe os melhores resultados no aprendizado não-supervisionado. O método K-Means foi desconsiderado pois obteve uma taxa de acertos muito baixa, porém o Fuzzy C-Means foi diferente, chegando a uma taxa de 92% de acurácia, com TPR de 82% e TNR de 97%. Mesmo com os bons resultados no critério externo, o critério interno não alcançou o que era esperado: 20% de silhouette e 2% no critério de Fisher. O gráfico do teste realizado pode ser conferido na imagem 4:

Figura 4. Fuzzy C-Means com 99% de variância



VII. CONCLUSÃO

Existem vários métodos de aprendizado, tanto supervisionado quanto não-supervisionado, que atendem os mais diversos problemas e devem ser usados conforme a necessidade. No caso da plataforma de lógica abordada no trabalho, esperava-se melhores resultados com o aprendizado supervisionado, o que foi comprovado com os testes de cinco métodos deste tipo de classificação. Da mesma maneira, os testes com aprendizado não-supervisionado foram feitos para demonstrar seu funcionamento, mesmo que não atenda as necessidades do projeto em questão.

Após a validação de todos os métodos com diferentes variâncias, chega-se à decisão que os melhores resultados obtidos são com o uso de aprendizado supervisionado, utilizando o método de SGDC e uma variância de 90%.

Apesar de ainda não haver a base de dados real, foi possível fazer uma simulação com ótimos resultados, os quais são esperados quando a plataforma for usada e dados dos alunos forem coletados, para assim poderem passar pelo treinamento supervisionado com SGDC.

REFERÊNCIAS

- [1] PONTES, Edel Alexandre Silva et al. Raciocínio lógico matemático no desenvolvimento do intelecto de crianças através das operações adição e subtração. *Diversitas Journal*, v. 2, n. 3, p. 469, 2017.
- [2] ZANATTA, Andrei C. et al. Desenvolvimento de Jogos para Estimular a Programação/Lógica em Crianças de 9 a 12 anos. In: *I SIMPOSIO IBERO-AMERICANO DE TECNOLOGIAS EDUCACIONAIS*. Araranguá, 2017.
- [3] ZANETTI, Humberto A. P. et al. Proposta de ensino de programação para crianças com Scratch e Pensamento Computacional. *Tecnologias Sociedade e Conhecimento*, Campinas, v. 4, n. 1, 2017.
- [4] KROHL, Diego R., DUTRA, Taynara C., MATOS, Camilla P. Programação para alunos do Ensino Fundamental II: os benefícios dos jogos lógicos digitais no aprendizado. *Revista Tecnologias na Educação*, v. 28, n. 10, 2018.
- [5] NICOLETE, Priscila. et al. Integração tecnológica na educação básica pública brasileira: uma análise sobre a evolução temporal e a interdisciplinaridade do tema. *RIAEÉ*, v. 11, n. 4, p. 2064-2086, 2016.
- [6] LUDERMIR, Teresa. *Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências*. Estudos Avançados: RJ, 2021.