Uma abordagem para predição da Retenção Escolar em cursos ministrados de forma híbrida pelo método de Sala de Aula Invertida

Adriana Mara Guimarães de Farias - PECIM - UNICAMP - a261661@dac.unicamp.br - https://orcid.org/0000-0002-3451-8946
Eduardo Galembeck - PECIM - UNICAMP - eg@unicamp.br - https://orcid.org/0000-0003-4238-5546
Marcos Evandro Cintra - Universidade Federal do Espírito Santo -mecintra@gmail.com https://orcid.org/0000-0001-7712-6865

Resumo: Este trabalho focou na questão da retenção escolar. Foi proposto um modelo de classificação para avaliar a situação dos alunos ao final do semestre letivo, se os mesmos seriam aprovados ou reprovados, com base nas notas das atividades avaliativas que os mesmos fazem ao longo do semestre. Foi usado o algoritmo *Naive Bayes* com dados de um conjunto de 16 atividades avaliativas aplicadas por meio do Ambiente Virtual de Aprendizagem MOODLE. Foram gerados 16 modelos distintos, cada um com a inclusão de mais uma atividade avaliativa, isto é, o primeiro modelo considerou apenas a avaliação 1, o segundo as duas primeiras avaliações, e assim por diante. Os resultados obtidos foram validados com uma turma de um semestre posterior. A principal contribuição deste trabalho é o fato de os modelos poderem ser incluídos a qualquer disciplina, usando dados de entrada de fácil obtenção, e resultando em uma acurácia alta. A aplicação do modelo permite que o docente identifique alunos que podem ser reprovados e realize ações para ajustar as atividades de ensino e reduzir a retenção escolar.
Palavras-chave: retenção escolar, *Naive Bayes*, sala de aula invertida, ensino híbrido.

A proposal to predict school retention in hybrid disciplines based on the flipped classroom methodology

Abstract: This work focuses on school retention. A model to classify students along the semester according to their probability of failing or being approved is proposed. The model generated by the *Naive Bayes* algorithm uses the students' grades as input variable.The grades are easily obtained from MOODLE. Sixteen models were generated, each one including all the grades from the tests taken during the semester. The first model used the grades of the first tests. The second model used the grades of the first and second test, and so on, incrementally. Results showed that the model with the information of the first test only was better than the other 15 models. The best model was validated using the data from a different class for the same subject. The main contribution of this work is that the model can be easily created for any discipline, as the data required is easily obtained, providing a highly accurate prediction regarding the students, which allows the professor to take actions in order to better adjust his classes with the challenges faced by students, thus, reducing school retention.
Key words: school retention, *Naive Bayes*, flipped classroom, blended learning.

DOI: https://doi.org/10.22456/1679-1916.129166

1. Introdução

O Sistema Educacional Superior Brasileiro enfrenta o problema crônico da retenção escolar. Com a expansão das Universidades Federais e Institutos Federais na última década, e, consequentemente, com o aumento da oferta de vagas, tanto no modelo tradicional e presencial quanto nos cursos à distância, esse problema tem se agravado. A retenção escolar é definida como sendo a dificuldade que o discente sente em concluir seu curso de graduação no período previsto no Projeto Pedagógico do curso (ARAÚJO; MARIANO; OLIVEIRA, 2021). Geralmente, a retenção escolar possui como consequência a evasão escolar, o qual é a situação em que o aluno abandona seu curso. A taxa de evasão escolar registrada no último Censo da Educação Superior chegou a 59% no ano de 2020 (BRASIL, 2022).

A evasão escolar causa imensos prejuízos pessoais ao aluno que abandona seus estudos, além de prejuízos ao erário, que investe na formação dos cidadãos. Além disso, um aluno que evade acaba impedindo um candidato de estudar em seu lugar. Esses, entre outros motivos, tornam o estudo de propostas para minimizar a evasão escolar um tópico de pesquisa de grande relevância.

As motivações que levam o aluno a se evadir são as mais diversas. Tais motivações podem ser classificadas como i) **Motivos Internos** e ligados à Instituição de Ensino, ou ii) **Motivos Externos** e pessoais. Apesar de serem muitas as razões que levam um aluno a abandonar seus estudos, uma delas está em primeiro lugar no relatório apresentado em UFPE (2016), especificamente, a *incompatibilidade entre a rotina de estudos e trabalho do aluno*.

Neste trabalho, foram estudados os motivos internos que levam à evasão escolar, em busca de dados que permitam a construção de um modelo que possa estimar o desempenho acadêmico dos estudantes. Com essa estimativa, é possível planejar ações para que o aluno não chegue a evadir. Essa análise pode permitir que docentes revejam seus planejamentos pedagógicos com base nos alunos que mostram maiores chances de evadir.

De fato, o desempenho acadêmico pode ser considerado um dos parâmetros mais adequados para avaliar a possibilidade de evasão de um aluno. Isso se deve ao fato de que há uma relação direta entre o desempenho acadêmico de um aluno e sua motivação para estudar e se manter no curso.

Esta percepção motivou os estudos realizados neste trabalho. Assim, foram realizados experimentos que buscavam responder à seguinte questão de pesquisa: ***É possível estimar o desempenho de um aluno em disciplinas de graduação ministradas no formato híbrido com a metodologia de Sala de Aula Invertida?***

Partindo dessa questão central, este trabalho tem como objetivo principal a proposta de estudar a aplicabilidade de técnicas de Aprendizado de Máquina, em especial o algoritmo *Naive Bayes*, para estimar a situação final de um discente. Para tal, foram usados dados registrados no Ambiente Virtual de Aprendizagem MOODLE (FURTADO; FERNANDES, 2018), que atua como ferramenta de suporte tecnológico educacional para a abordagem pedagógica desenvolvida.

Ao realizar os experimentos deste trabalho, foram obtidos 16 modelos preditivos com métricas bem satisfatórias em relação a erro e precisão. Em cada modelo é possível obter como informação importante se um aluno será considerado aprovado ou reprovado quando a disciplina encerrar. O modelo de melhor desempenho obtido ainda permite

DOI: https://doi.org/10.22456/1679-1916.129166

prever com uma margem bem pertinente de antecipação, assegurando tempo suficiente para estratégias pedagógicas de restauração do aprendizado do aluno.

Dessa forma, este artigo segue a seguinte estrutura: na Seção 2 é apresentada uma revisão sobre trabalhos relacionados à avaliação de desempenho de alunos. Na Seção 3, encontra-se a narrativa sobre a metodologia utilizada e os resultados alcançados. Por fim, na Seção 4 são apresentadas as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

No trabalho realizado em Hartatik, Kusrini e Prasetio (2020), os autores analisaram o desempenho dos alunos utilizando o algoritmo *Naive Bayes* para produzir um modelo para auxiliar no acompanhamento do progresso de alunos em cursos de graduação e estimar o período de suas integralizações de curso. Os autores utilizaram um indicador chamado Índice de Desempenho do Semestre que agrega o desempenho acadêmico vivenciado pelo discente em todas as disciplinas cursadas em um semestre letivo. Foram obtidos dois modelos de previsão considerando a análise de diferentes semestres. Também integram esses modelos variáveis como gênero e moradia. Utilizando o KNIME, os autores alcançaram com o modelo 2 um desempenho melhor por considerar o desempenho do discente em quatro diferentes semestres.

Em Haviluddin et al. (2018), o objetivo principal da pesquisa foi prever o tempo de estudo de cada aluno, buscando identificar a situação de evasão escolar. De fato, o tempo de permanência de um aluno reflete diretamente no desempenho dos departamentos acadêmicos. Para a construção de um modelo preditivo, os autores utilizaram o algoritmo *Naive Bayes* usando um conjunto de dados registrados entre os anos de 2014 a 2017 que incluem: idade, local de nascimento, sexo, se o ensino médio foi cursado em escola pública ou privada e duas taxas de progresso do curso. Como resultados obtidos, o algoritmo *Naive Bayes* teve precisão de 76,79% e 44,62% de Verdadeiros Positivos. Os autores declaram que o algoritmo *Naive Bayes* pode ser utilizado como um modelo alternativo na avaliação do desempenho acadêmico de alunos.

A preocupação com os motivos que levam um aluno a se evadir de seu curso foi o problema de pesquisa do trabalho desenvolvido em Hegde e Prageeth (2018). Para o desenvolvimento dessa pesquisa, foram realizadas duas coletas de dados por meio de formulário do Google, buscando distinguir o ponto de vista de um aluno que acaba de ingressar na faculdade das impressões de um aluno que ainda continua ativo no terceiro semestre letivo. O algoritmo *Naive Bayes* foi utilizado para encontrar um modelo que conseguisse prever a evasão de alunos e identificar o motivo de sua desistência. Para obtenção dos resultados, o algoritmo *Naive Bayes* foi implementado com linguagem R utilizando dados de 50 alunos e 24 atributos com informações acadêmicas, demográficas, de redes sociais, fatores psicológicos e de saúde. O modelo preditivo obtido alcançou uma acurácia de 72%, considerada pelos autores um desempenho bem relevante para o cenário da pesquisa.

O trabalho realizado em Páez (2022), teve em vista comparar o desempenho de modelos preditivos obtidos a partir da utilização dos seguintes algoritmos: *Naive Bayes*, *KMeans* e C4.5. Para a obtenção de cada um destes modelos, os autores utilizaram como variáveis de entrada a pontuação obtida em atividades acadêmicas realizadas com 260 estudantes. Com a obtenção de cada um desses modelos, estes foram aplicados em resultados de outros 112 alunos de uma turma subsequente. O algoritmo *Naive Bayes*

DOI: https://doi.org/10.22456/1679-1916.129166

obteve melhor desempenho com uma precisão de 70,5% considerando a duração de 21% do semestre. Como trabalhos futuros, os autores apontam que seria interessante aplicar a técnica do voto majoritário, além de incluir mais atributos como informações demográficas e socioeconômicas.

Na próxima seção é apresentada a metodologia de trabalho desta pesquisa e os resultados alcançados.

3. Metodologia do Trabalho e Resultados Alcançados

Para encontrar uma resposta para a questão de pesquisa norteadora deste trabalho, foi proposta a geração de um modelo baseado no algoritmo *Naive Bayes* com dados oriundos da participação de discentes em atividades avaliativas de uma disciplina de graduação ministrada sob o formato híbrido com a metodologia de Sala de Aula Invertida.

O Ensino Híbrido busca uma aprendizagem mais ativa e menos passiva dos discentes durante o tempo de sala de aula presencial. Ao invés do aluno apenas assimilar o conteúdo em sala de aula, busca-se planejar ações que permitam maior interação entre os alunos. Também é pensado no desenvolvimento das habilidades dos alunos por meio de problemas mais desafiadores, realizados em grupo. No Ensino Híbrido, é importante que o docente conheça o perfil de seus alunos. Nesse sentido, o planejamento do processo avaliativo deve ser feito cuidadosamente. De fato, a partir das avaliações propostas é possível a descoberta de falhas e potencialidades de cada aluno. De acordo com Bacich, Neto e Trevisani (2015), na abordagem do Ensino Híbrido a avaliação é voltada para o aluno e como ele aprende. Assim, a partir dos seus resultados pode-se pensar em como personalizar o ensino.

Um dos métodos mais adotados no ambiente escolar para reproduzir a abordagem do Ensino Híbrido é o da Sala de Aula Invertida, do inglês *Flipped Classroom*, (BERGMAN; SAMS, 2018). Sob esta abordagem, a disciplina é planejada pensando na inversão de momentos de aprendizado, tornando o roteiro mais crítico, participativo e colaborativo.

Para melhor compreensão da proposta abordada na componente curricular deste trabalho, o planejamento pedagógico implementado se estruturou da seguinte forma:

1. Distribuição do conteúdo programático em pequenos blocos discutidos e avaliados semanalmente. O objetivo almejado com esta abordagem é o de que os assuntos abordados na disciplina possam ser vistos e debatidos com os discentes em um curto espaço de tempo e o mesmo já possa ser avaliado.
2. Todo o conteúdo semanal a ser discutido na disciplina é disponibilizado previamente para que os alunos possam ter acesso antes da aula. O detalhe neste ponto é que o conteúdo deve ser preparado de forma sucinta e objetiva, permitindo que o discente consiga acessá-lo no tempo disponível antes do encontro na sala de aula.
3. A partir do acesso ao conteúdo semanal, é liberada uma avaliação formativa (HORN; STAKER, 2015). Uma avaliação formativa tem o objetivo de averiguar a compreensão formada sozinha pelo discente antes do momento na sala de aula.
4. Durante o momento da sala de aula é seguido o seguinte roteiro:
    a. Discussão sobre a atividade avaliativa da semana anterior;
    b. Revisão do conteúdo semanal postado;

DOI: https://doi.org/10.22456/1679-1916.129166

c. Esclarecimento das dúvidas;
d. Atividades práticas com resolução de problemas;
e. Suporte e mentoria individual para os discentes.

5. Encerrando o conteúdo semanal, é liberada uma atividade avaliativa somativa, (VASCONCELLOS, 2008), que compõe parte da sua nota final na disciplina, abordando os assuntos vistos e discutidos.

Com esta estratégia, o discente passa por cerca de 16 atividades formativas e 16 atividades somativas. Os dados relativos às 16 atividades somativas aplicadas a 92 alunos foram usados para gerar o modelo probabilístico pelo algoritmo *Naive Bayes*. O *Naive Bayes* é um algoritmo classificador probabilístico preditivo baseado no Teorema de Bayes, assumindo haver independência entre as variáveis para determinar as classes finais (CASTRO; FERRARI, 2016). Todos os experimentos foram realizados utilizando a plataforma do Google Colab (RESEARCH, 2022) e a linguagem de programação Python.

Os modelos foram gerados usando a variável Pontuação Final do aluno como entrada. Não foram usadas outras variáveis porque no momento da realização dos experimentos não estavam disponíveis outros dados.

As atividades incluídas neste trabalho tinham diferentes pesos: 0,25, 0,50 ou 1,00 ponto no total. Assim, foram definidas escalas conforme a pontuação máxima das atividades. Essas escalas relacionam a nota dos alunos com seu desempenho, como mostrado na Figura 1.

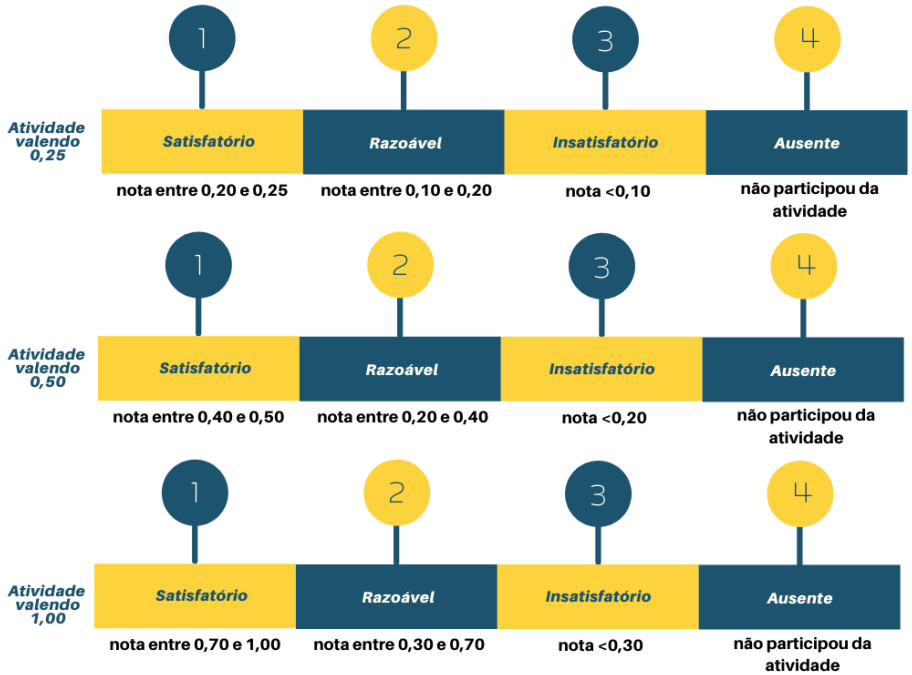


Figura 1 - Escalas de desempenho para cada peso de atividade. Fonte: Autoria própria.

Como foram feitas avaliações semanais, a nota final representa a soma de todas as pontuações obtidas em cada atividade somativa, podendo variar de 0,25 a 1,00 ponto. A Figura 1 mostra como cada nota é relacionada ao desempenho do aluno. Por exemplo, um discente que obtenha uma pontuação 0,20 em uma atividade avaliativa valendo 1,00

DOI: https://doi.org/10.22456/1679-1916.129166

ponto tem seu desempenho considerado *Insatisfatório*. No entanto, a mesma pontuação em uma atividade que vale 0,50, tem o desempenho considerado *Razoável* e em uma atividade valendo 0,25 seu desempenho é considerado *Satisfatório*.

Assim, ao aplicar o algoritmo *Naive Bayes*, foi realizado um pré-processamento nas pontuações de cada atividade de forma que os valores obtidos por cada aluno em cada atividade refletissem o seu desempenho, conforme mostrado na Figura 1. O objetivo ao aplicar o algoritmo *Naive Bayes* foi o de criar um modelo para auxiliar na estimativa da situação final do discente na disciplina: Aprovado ou Reprovado. A Tabela 1 apresenta o desempenho de cada um dos 10 alunos em cada uma das 16 atividades.

Tabela 1 - Desempenho dos alunos por atividade avaliativa e a situação final na disciplina.

Aluno	Situação Final	Atividade 1	Atividade 2	Atividade 3	...	Atividade 16
Aluno 1	Reprovado	Ausente	Ausente	Razoável		Ausente
Aluno 2	Aprovado	Satisfatório	Ausente	Razoável		Satisfatório
Aluno 3	Aprovado	Razoável	Ausente	Satisfatório		Satisfatório
Aluno 4	Reprovado	Satisfatório	Ausente	Satisfatório		Ausente
Aluno 5	Reprovado	Satisfatório	Ausente	Razoável		Ausente
Aluno 6	Aprovado	Satisfatório	Ausente	Razoável		Satisfatório
Aluno 7	Reprovado	Ausente	Ausente	Ausente		Ausente
Aluno 8	Aprovado	Satisfatório	Ausente	Razoável		Satisfatório
Aluno 9	Reprovado	Ausente	Satisfatório	Ausente		Ausente
Aluno 10	Aprovado	Satisfatório	Ausente	Satisfatório		Satisfatório

A segunda coluna da Tabela 1 apresenta a situação final do aluno, se ele foi aprovado ou reprovado. As demais mostram o desempenho deles em cada atividade.

Ao final do semestre letivo, sabendo-se a situação final de cada aluno, ou seja, se foram Aprovados ou Reprovados, foram criados 16 modelos distintos: o primeiro considerou apenas a nota da primeira atividade; o segundo considerou as notas das atividades 1 e 2; o terceiro considerou as notas das atividades 1, 2 e 3; até o último modelo que usou as notas de todas as 16 atividades.

A Tabela 2 mostra os resultados obtidos com os 16 experimentos, incluindo o erro, a precisão para Aprovados (A) e Reprovados (R), o *recall* para Aprovados (A) e para Reprovados (R) e o valor de F1 para Aprovados (A) e Reprovados (R). Além disso, a Tabela 2 apresenta dados de validação do modelo de melhor desempenho aplicado em uma turma subsequente da mesma componente curricular.

Tabela 2 - Métricas de desempenho obtidas pelo *Naive Bayes* para os 16 experimentos e para a turma de validação.

Atividade	Erro	Precisão A	Precisão R	*Recall* A	*Recall* R	F1 A	F1 R
1	0,09	0,91	1	1	0,33	0,95	0,5
1 e 2	0,17	0,9	0,33	0,9	0,33	0,9	0,33
1 a 3	0,17	0,9	0,33	0,9	0,33	0,9	0,33
1 a 4	0,17	0,9	0,33	0,9	0,33	0,9	0,33

DOI: https://doi.org/10.22456/1679-1916.129166

1 a 5	0,22	0,94	0,33	0,8	0,67	0,86	0,44
1 a 6	0,22	0,94	0,33	0,8	0,67	0,86	0,44
1 a 7	0,22	0,94	0,33	0,8	0,67	0,86	0,44
1 a 8	0,22	0,94	0,33	0,8	0,67	0,86	0,44
1 a 9	0,22	0,94	0,33	0,8	0,67	0,86	0,44
1 a 10	0,17	1	0,43	0,8	1	0,89	0,6
1 a 11	0,17	1	0,43	0,8	1	0,89	0,6
1 a 12	0,17	1	0,43	0,8	1	0,89	0,6
1 a 13	0,17	1	0,43	0,8	1	0,89	0,6
1 a 14	0,17	1	0,43	0,8	1	0,89	0,6
1 a 15	0,17	1	0,43	0,8	1	0,89	0,6
1 a 16	0,17	1	0,43	0,8	1	0,89	0,6
Validação	**0,08**	**0,95**	**0,33**	**0,95**	**0,33**	**0,95**	**0,33**

Dentre os desempenhos obtidos nas 16 simulações realizadas, a que obteve melhor conjunto de resultados foi a simulação utilizando apenas a pontuação obtida na 1ª atividade avaliativa regular da disciplina, obtendo o menor erro, 9% (coluna Erro), e precisão da situação final de aprovado em 91% (coluna Precisão A) e de reprovados em 100% (coluna Precisão R).

É importante mencionar que quanto antes o docente tiver uma previsão assertiva da situação final dos discentes de suas disciplinas, mais tempo ele irá dispor para realinhar o seu planejamento pedagógico e evitar situações de abandono ou reprovação em massa. No entanto, mesmo o modelo apontando situações de aprovação e reprovação dos discentes, o docente deve se manter atento ao desempenho global dos participantes da turma, considerando não apenas os resultados das avaliações, mas também a participação em sala de aula, a interação e colaboração com demais colegas, entre tantos pontos relevantes para a avaliação dos alunos. Já a pontuação obtida em uma atividade, quando usada para geração de um modelo de previsão, permite o diagnóstico de aprendizagem de um aluno de forma quantitativa. Porém, qualquer resultado preditivo com estes valores deve ser acompanhado de impressões a partir do cotidiano da disciplina.

Em relação à validação dos modelos, o modelo preditivo obtido com a Atividade 1, que obteve desempenho mais satisfatório em relação aos demais, foi selecionado para realizar uma previsão em dados de outra turma da mesma componente curricular do semestre subsequente aos dos dados aplicados na obtenção deste experimento. Foram utilizados dados de 48 alunos considerando a pontuação obtida na 1ª atividade avaliativa e, como resultado dessa previsão, foi obtido um erro de 8% com precisão de 95% para os aprovados (A) e 33% para os reprovados (R). Para as métricas obtidas com a previsão a partir dos dados da turma subsequente percebe-se que o erro diminui e a precisão para prever os aprovados aumentou. No entanto, outra reflexão que merece destaque em relação à precisão de previsão de alunos reprovados é que, apesar de ela ter sido menor para a turma de validação, isso ainda pode ser considerado algo positivo especificamente para este cenário: uma vez que um discente é apontado como reprovado na situação final da disciplina, a ação por parte do docente da turma é reforçar a atenção e o acompanhamento em relação a este aluno justamente para que se possa evitar o

DOI: https://doi.org/10.22456/1679-1916.129166

abandono ou reprovação. Nesse contexto, até os falsos negativos para a previsão de reprovados não causa prejuízo aos alunos, visto que quaisquer ações tomadas pelo docente para adequar suas aulas, podem ser consideradas positivas de uma forma geral para o acompanhamento da turma. De fato, para a turma de validação com 48 alunos, o modelo previu que 45 deles seriam aprovados e 3 reprovados ao final do semestre. Quando o semestre finalizou, 46 alunos foram aprovados, sendo 1 aluno apontado como reprovado na previsão.

4. Conclusões e Trabalhos Futuros

A proposta deste trabalho foi definida com o objetivo estudar a aplicabilidade de técnicas de Aprendizado de Máquina, em especial a abordagem probabilística do *Naive Bayes*, para estimar a situação final de um discente usando dados de atividades avaliativas registrados no MOODLE.

Como resultados, foram obtidos 16 modelos preditivos com métricas satisfatórias em relação a erro e precisão de cada classe (Aprovados e Reprovados). Cada modelo foi obtido considerando as pontuações registradas pelos discentes em cada atividade avaliativa, sendo incrementado o número de atividades a cada simulação. O modelo preditivo com as melhores métricas em relação aos demais foi obtido a partir da Atividade 1. Este modelo obteve erro de 9%, precisão de previsão de alunos aprovados em 91% e precisão de previsão de reprovados em 100%.

Como forma de validar o desempenho deste modelo, este foi aplicado em dados de uma turma subsequente da mesma disciplina e alcançou métricas bem convenientes para o cenário da proposta deste trabalho, apresentando um erro de 8%, precisão da previsão de alunos aprovados em 95% e precisão da previsão dos reprovados em 33%.

As conclusões que podem ser observadas a partir do experimento realizado e com os resultados obtidos é que o modelo de previsão se mostrou eficiente para analisar a situação de alunos aprovados ao final de um semestre letivo, permitindo que essa identificação fosse realizada na primeira participação avaliativa do discente.

Outra consideração bem pertinente é que a principal contribuição deste trabalho é a simplicidade do modelo e facilidade de aplicação em qualquer outra disciplina, uma vez que os dados utilizados como entrada do modelo foram as pontuações de atividades avaliativas, informação facilmente registrada em qualquer contexto educacional.

Por fim, como trabalhos futuros, pretende-se ampliar o número de variáveis de entrada, entre elas o tempo de resolução de cada questão avaliativa, a quantidade de acertos e erros e a complexidade de cada avaliação. Também pretende-se utilizar outros algoritmos de Aprendizado de Máquina para possibilitar comparações entre os modelos.

5. Referência Bibliográficas

ARAÚJO, A. C. P. L.; MARIANO, F. Z.; OLIVEIRA, C. S. Determinantes acadêmicos da retenção no Ensino Superior. **Ensaio: Avaliação e Políticas Públicas em Educação**, v. 29, p. 1045–1066, mar. 2021.

BACICH, L.; NETO, A. T.; TREVISANI, F. M. **Ensino Híbrido:** Personalização e Tecnologia na Educação. Porto Alegre: Penso Editora, 2015.

BERGMAN, J.; SAMS, A. **Sala de Aula Invertida:** uma metodologia ativa de aprendizagem. Rio de Janeiro: LTC, 2018.

DOI: https://doi.org/10.22456/1679-1916.129166

BRASIL. **Censo da Educação Superior 2020:** Notas Estatísticas. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), 2022. Disponível em: <https://download.inep.gov.br/publicacoes/institucionais/estatisticas_e_indicadores/notas_estatisticas_censo_da_educacao_superior_2020.pdf>.

CASTRO, L. N.; FERRARI, D. G. **Introdução a Mineração de Dados:** conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva Editora, 2016.

FURTADO, U. M.; FERNANDES, J. O. **Introdução à EAD:** moodle para estudantes. Mossoró: EdUFERSA Editora, 2018.

HARTATIK; KUSRINI, K.; PRASETIO, A. B. Prediction of Student Graduation with Naive Bayes Algorithm. **In: Anais do Fifth International Conference on Informatics and Computing (ICIC).** nov. 2020.

HAVILUDDIN et al. Student Academic Evaluation using Naïve Bayes Classifier Algorithm. In: **Anais da 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT).** nov. 2018.

HEGDE, V.; PRAGEETH, P. P. Higher education student dropout prediction and analysis through educational data mining. In: **Anais da 2nd International Conference on Inventive Systems and Control (ICISC).** jan. 2018.

HORN, M. B.; STAKER, H. **Blended:** usando a inovação disruptiva para aprimorar a educação. Porto Alegre: Penso Editora, 2015.

RESEARCH, G. **Welcome to Colaboratory!** Disponível em: <https://colab.research.google.com/>.

PÁEZ, A. R. Modelos predictivos progresivos del rendimiento académico de estudiantes universitarios. **RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo**, v. 12, n. 24, maio 2022.

UFPE. **Causas da evasão de alunos nos cursos de graduação presencial da UFPE**. Universidade Federal de Pernambuco, 2016. Disponível em: <https://www.ufpe.br/documents/38954/371376/r_evaso_16.pdf/53642e52-41fb-4b43-b098-98db6a470176>.

VASCONCELLOS, C. S. **Avaliação:** concepção dialética-libertadora do processo de avaliação escolar. São Paulo: Libertad, 2008.

DOI: https://doi.org/10.22456/1679-1916.129166