

Um Modelo para Análise do Impacto da Retenção e Evasão no Ensino Superior Utilizando Cadeias de Markov Absorventes

Juliana F. dos Santos¹, José Diogo A. de Sousa¹,
Rafael Ferreira Mello¹, Cláudio T. Cristino², Gabriel Alves¹

¹Programa de Pós-Graduação em Informática Aplicada (PPGIA)
Universidade Federal Rural de Pernambuco (UFRPE)

²Programa de Pós-Graduação em Biometria e Estatística Aplicada (PPGBEA)
Universidade Federal Rural de Pernambuco (UFRPE)

{julifs.13, diogosousa36}@gmail.com,

{rafael.mello, claudio.cristino, gabriel.alves}@ufrpe.br

Abstract. *Dropout and retention are recurrent problems in undergraduate courses. Survival analysis allows verifying the influence of events, such as dropout, graduation, and student bonding, according to time and probability of occurrence. It allows finding solutions to solve these problems in advance and avoids losses to universities. This work aims to propose a Markov Chain model to carry out the survival analysis of students throughout undergraduate courses. This work evaluated two undergraduate computing courses of a Brazilian public university, observing a different dropout behavior between them. Analyzes show that controlling retention rates in early periods is the most significant feature to reduce the overall dropout rate.*

Resumo. *A evasão e a retenção são problemas recorrentes em cursos de graduação. A análise de sobrevivência permite verificar a influência de eventos, como a evasão, conclusão e vínculo estudantil, de acordo com o tempo e a probabilidade de ocorrência. Ela permite buscar soluções para resolver esses problemas com antecedência e evitar prejuízos às universidades. Este trabalho tem como objetivo propor um modelo de Cadeia de Markov para realizar a análise de sobrevivência dos estudantes ao longo de cursos de graduação. Neste trabalho, foram analisados dois cursos de graduação da área de computação de uma universidade pública brasileira, onde foi identificado que a evasão entre os dois cursos possui um comportamento diferente. As análises realizadas mostraram que o controle da retenção nos primeiros períodos é a ação mais importante para a redução da evasão dos cursos.*

1. Introdução

A evasão estudantil é um fenômeno estudado há décadas e pode ser definida como a interrupção do progresso acadêmico do estudante ao longo dos semestres [Xavier and Meneses 2020]. Este fenômeno é considerado um problema de âmbito internacional, tanto para as Instituições de Ensino Superior (IES) públicas, devido à ociosidade de recursos, como para as IES privadas por implicar em perda de receita [Da Silva and Santos 2017]. A evasão pode ser observada com relação a um único curso,

a uma instituição ou mesmo ao sistema de ensino. Assim, um(a) estudante troca de curso em uma mesma instituição, pode ser visto como um “evadido” para o curso, mas não para a instituição e muito menos para o sistema de ensino como um todo. Neste trabalho, observamos a evasão sob a ótica do curso.

Trabalhos quantitativos e qualitativos realizados anteriormente, mostram que a evasão é um problema multifatorial [Da Silva and Santos 2017, Saccaro et al. 2019, Campos 2016], que possui causas emocionais, financeiras ou mesmo problemas relacionados ao curso, como a sua estrutura curricular. Do ponto de vista quantitativo, muitos trabalhos mostram que a retenção e a evasão ocorrem de forma mais intensa nos primeiros períodos dos cursos [Xavier and Meneses 2020, Boumi and Vela 2020]. Este trabalho foca a análise quantitativa do problema de evasão, não sendo seu objetivo investigar as suas causas.

A análise de sobrevivência de Kaplan-Meier, a mais comumente encontrada em diversos estudos, é uma estatística não paramétrica usada para estimar a função de sobrevivência a partir de dados de eventos. Essa análise permite avaliar o tempo até a ocorrência de eventos, como a evasão e a conclusão do estudante, observando a probabilidade e duração média até o evento [Kleinbaum and Mitchel 2010]. Cadeias de Markov são utilizadas para representar o sistema através de um modelo estocástico a partir do qual podem ser realizadas análises e aferir métricas probabilísticas [Sheskin 2016]. No caso de Cadeias de Markov absorventes, é possível aferir métricas como o tempo médio até que um evento ocorra, ou a probabilidade do sistema estar em um determinado estado s em um tempo t .

Diante deste contexto, este trabalho propõe um modelo de Cadeia de Markov absorvente capaz de representar o progresso dos estudantes de um curso ao longo dos anos. O modelo permite representar o momento em que o estudante passa a ser considerado “retido”, passando a ter uma maior probabilidade de evasão. Este modelo permite aferir métricas como o tempo e a probabilidade de um estudante se graduar ou evadir. Além disso, associado à análise de sobrevivência é possível observar se dois ou mais cursos possuem o mesmo comportamento, através de testes de hipóteses. Outra vantagem da utilização de uma técnica de modelagem com a Cadeia de Markov é a possibilidade de realizar análises, alterando-se os parâmetros do modelo. Assim, é possível observar como a mudança na probabilidade de evasão nos primeiros períodos pode impactar na probabilidade de conclusão.

Para avaliar o modelo proposto, foram realizadas análises em dois cursos de graduação de uma instituição de ensino superior pública. Os resultados apresentam uma diferença no comportamento da evasão dos dois cursos analisados, porém uma semelhança para o término do vínculo. Também foram realizados testes alterando as probabilidades de evasão e/ou de retenção nos primeiros períodos. Foi identificado que, comparado com o controle da evasão prematura, o controle da retenção se mostra mais eficaz para diminuir a evasão global dos cursos.

2. Trabalhos Relacionados

A análise da evasão e retenção em estudantes universitários têm crescido nos últimos anos, como pode ser observado nos diversos estudos disponíveis nos anais do Simpósio Brasileiro de Informática na Educação – SBIE, [Lima and Coutinho 2019, Ramos et al. 2017,

Queiroga et al. 2017, Dos Santos and Falão 2017, Carrano et al. 2019, Beltran et al. 2019, Manhães et al. 2012]. O uso de técnicas de modelagem [Araújo 2019, Boumi and Vela 2020, Brezavšček et al. 2017], análise de sobrevivência [Campos 2016, Saccaro et al. 2019, Costa et al. 2018], mineração de dados [Carrano et al. 2019, Rigo et al. 2014, Manhães et al. 2012], aprendizagem de máquina [Ramos et al. 2017, Queiroga et al. 2017, De Jesus et al. 2021] e pesquisas quantitativas e qualitativas [Lima and Coutinho 2019, Rodrigues et al. 2015], permitem analisar fatores que influenciam na retenção, evasão, conclusão e vínculo estudantil.

Nos trabalhos de [Lima and Coutinho 2019] e [Rodrigues et al. 2015], os autores realizam uma análise descritiva sobre a evasão e o desempenho acadêmico em cursos da área de tecnologia, através de dados de notas, frequências, reprovações e vínculo, com o intuito de investigar o comportamento de turmas ao longo do tempo, para encontrar causas que levam os estudantes a evadir do curso. Além de analisar o problema da evasão na perspectiva dos alunos. Os resultados apontaram um alto percentual de estudantes em atraso no curso, incompatibilidade entre os horários ofertados e que mesmo um alto índice de frequência por parte dos estudantes, nem sempre é convertido em boas notas.

Diante disso, técnicas de mineração de dados já foram aplicadas em alguns trabalhos, com o objetivo de aferir indicadores e características que contribuem no entendimento da evasão [Carrano et al. 2019, Rigo et al. 2014, Manhães et al. 2012]. Alguns algoritmos de aprendizado de máquina, como o Random Forest, SVM e o KNN, são bastante utilizados com o objetivo de criar grupos com características similares. No contexto da educação, estes algoritmos podem analisar a importância de atributos como cor/raça, gênero, cidade, estado civil, desempenho dos alunos no primeiro semestre, assiduidade e satisfação, são os mais relevantes e prever a probabilidade de evasão dos estudantes. Alguns trabalhos apresentam uma acurácia que varia de 60% até 85% [De Brito et al. 2020, Dos Santos and Falão 2017], por exemplo.

Alguns modelos utilizam técnicas de aprendizagem de máquina, como Multilayer Perceptron e Regressão Logística, para prever a evasão dos estudantes, tanto em ambientes EAD, quanto presenciais [Ramos et al. 2017, Queiroga et al. 2017, De Jesus et al. 2021]. Os resultados apontam um desempenho entre 75% e 95% na predição de estudantes em risco de evasão no primeiro semestre. Isso permite, conforme visto em [Beltran et al. 2019] e [De Almeida Teodoro and Kappel 2020], monitorar o risco de evasão dos estudantes e oferecer auxílio aos que apresentam maior risco para diminuir a probabilidade de evasão.

Em [Brezavšček et al. 2017], [Boumi and Vela 2020] e [Araújo 2019], o objetivo foi o de se desenvolver um modelo estocástico para estimar e monitorar continuamente vários indicadores de qualidade e eficácia no ensino superior. Os modelos de Cadeia de Markov foram aplicados para simular o progresso acadêmico em instituições de ensino superior, onde o progresso dos alunos foi estimado considerando a duração prevista do estudo, os créditos acadêmicos ou a quantidade de reprovações para representar os estados. Os dados obtidos através dos modelos permitem estimar a progressão dos alunos entre os diferentes estágios, o tempo esperado que um aluno passa em um determinado estágio, a duração prevista para concluir o curso, a probabilidade de conclusão ou evasão, prever a futura matrícula dos estudantes e simular a chance de evasão dos estudantes, através de alterações nas probabilidades de transição do modelo. Desta forma, determina-se ações

para diminuir a evasão.

Dentre os estudos citados acima, [Araújo 2019] e [Brezavšček et al. 2017] têm a representação do estado trancado, onde o aluno pode transitar para este estado a partir de qualquer estado que representa o progresso estudantil. Porém, ao seguir a matriz de transição, o estudante poderia sair do estado trancado para voltar em um estado diferente ao qual se encontrava antes de atingir o estado trancado. Portanto, no presente trabalho o trancamento está incluído, porém de forma implícita, já que o progresso acadêmico aqui simulado será representado pelos semestres em que o estudante está no curso, o que inclui o tempo em que ele passa trancado.

Apesar de [Brezavšček et al. 2017], [Araújo 2019] e [Boumi and Vela 2020] representarem o progresso acadêmico com Cadeia de Markov, eles não demonstram a criação de estados específicos para o progresso dos alunos retidos. Dessa forma, todos os estudantes têm o progresso acadêmico representado pelos mesmos estados e a retenção é representada por *self-loop*. Isso não demonstra o real comportamento dos estudantes retidos e pode fazer com que os dados que o modelo gera sejam irreais em relação ao tempo e probabilidade do estudante para a conclusão ou evasão. Portanto, no presente trabalho, o progresso estudantil é definido em relação à quantidade de semestres em que o estudante está vinculado ao curso. Isso permite criar progressos alternativos entre os alunos retidos e não retidos.

O uso de análise de sobrevivência permite avaliar a probabilidade de eventos em relação ao tempo. Neste contexto, os trabalhos [Saccaro et al. 2019], [Campos 2016] e [Costa et al. 2018] têm o objetivo de analisar variáveis que influenciam no tempo de permanência e probabilidade de conclusão e evasão para cursos em instituições públicas e privadas. Foram utilizadas técnicas de análise de sobrevivência, tais como Kaplan-Meier, teste Log-rank e modelos de regressão semi-paramétricos, levando em consideração o número de semestres do curso, gênero, idade, forma de ingresso, turno, desempenho acadêmico (média das notas) e a existência de reprovação e trancamento entre os estudantes. Como resultado, percebe-se que a evasão é maior nos primeiros períodos e em instituições privadas.

A probabilidade de evasão maior durante o primeiro ano do curso foi um fator recorrente durante os trabalhos analisados [Campos 2016, Saccaro et al. 2019, Costa et al. 2018]. Ela demonstra, entre outros fatores, um efeito relacionado ao sistema de seleção unificada (SISU), onde estudantes que entram na universidade por meio deste apresentam uma probabilidade de evadir maior em comparação a alunos que entram por outros meios, como avaliação seriada [Saccaro et al. 2019]. Essa observação demonstra a importância de análises para avaliar como a diminuição da evasão e retenção nos primeiros anos pode impactar na evasão do curso como um todo.

Nesse contexto, pode-se observar que o estudo da evasão é de grande relevância. Ademais, as Cadeias de Markov permitem a modelagem e análise de sistemas de eventos discretos. Assim, elas são capazes de representar de forma eficaz o progresso de alunos em cursos de graduação, permitindo ainda a parametrização do modelo e a realização de experimentos. Estes modelos permitem, por exemplo, observar o comportamento da taxa de evasão global ao mudar a retenção em um período específico.

A análise de sobrevivência é bastante utilizada para apresentar a probabilidade

de um evento relacionado ao tempo de ocorrência. Essas técnicas destacam-se entre as demais utilizadas para avaliar a evasão, pois permitem criar um modelo generalizado capaz de prever o comportamento de estudantes em relação à probabilidade e o tempo, com a possibilidade de ser aplicado em diversos contextos sem necessidade de alteração do modelo, apenas de seus parâmetros. Desta forma, esta pesquisa pretende realizar uma modelagem do progresso dos estudantes de graduação utilizando Cadeia de Markov com a criação de estados específicos para representar os alunos retidos, o que ainda não foi feito nos trabalhos anteriores. Além do auxílio de gráficos de análise de sobrevivência de Kaplan-Meier para analisar as probabilidades dos eventos evasão, conclusão e desvinculação conforme os semestres do curso.

3. Metodologia

O modelo proposto neste artigo é definido por meio de uma Cadeia de Markov Absorvente. Esse é um processo estocástico, composto por estados, onde a probabilidade de transição ocorre observando apenas o estado atual e evoluindo conforme um tempo discreto. Os estados do modelo foram definidos considerando os semestres em que o estudante está vinculado ao curso, com o objetivo de simular o progresso acadêmico.

Entre as situações acadêmicas definidas como estados neste estudo, o estado retido é determinado como quando o estudante é reprovado em uma quantidade de disciplinas maior que a quantidade de um período ou caso ele tenha trancado em algum semestre, o que normalmente o levaria a não conseguir se formar no tempo ideal do curso. O estado evadido indica que o estudante que abandonou o curso, enquanto o estado concluído, representa o momento em que o estudante finaliza o curso com sucesso. Por fim, o estudante que se evadiu ou concluiu o curso é considerado como desvinculado.

O modelo, representado por uma versão simplificada na Figura 1, com omissão das probabilidades de transição, é composto por estados de transição representados por S_n , onde n é a duração em períodos do vínculo e S_xR demonstra o vínculo quando o estudante é considerado retido, onde x é a duração em períodos do vínculo retido. Essa diferença entre os estados retidos e não retidos é necessária, já que esses estudantes apresentam comportamentos distintos na progressão durante os períodos do curso de graduação. Entre os estados de absorção, o estado C representa o estudante concluído, e o estado E representa o estudante evadido.

A transição pelos estados do modelo ocorre a partir do estado inicial representado pelo primeiro semestre do curso, onde o aluno pode realizar uma transição para o semestre seguinte, retido ou não, ou para o estado de evasão. A partir dos estados que representam o período padrão de conclusão do curso, também é possível atingir o estado de conclusão ou continuar progredindo pelos estados dos semestres seguintes.

As probabilidades de transição entre os estados são determinadas pela matriz de transição. Essas probabilidades utilizadas como entrada no modelo proposto foram extraídas a partir de dados de 1918 alunos de dois cursos da área de computação de uma universidade pública. Foram considerados estudantes que ingressaram pelo Sistema de Seleção Unificada (SISU) em 10 semestres diferentes.

Além da disponibilidade e confiabilidade dos dados, os dois cursos analisados foram escolhidos por possuírem características similares, como a área de conhecimento

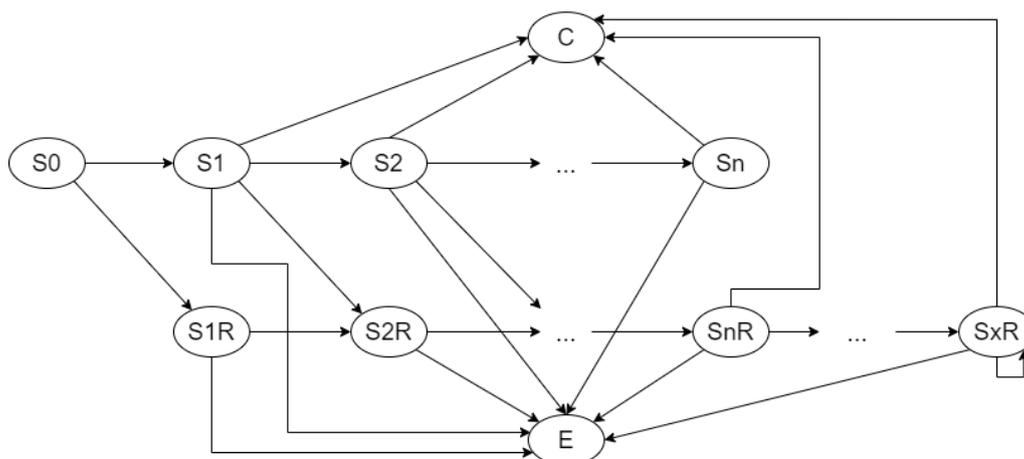


Figura 1. Modelo de Cadeia de Markov para Análise de Sobrevivência para representar o progresso acadêmico.

em que estão classificados, disciplinas que fazem parte do currículo acadêmico e o perfil dos docentes. Isto facilita algumas análises e comparações, auxiliando a validação do método proposto neste trabalho.

Ao considerar os dados apresentados dos dois cursos selecionados, o comportamento de progressão dos estudantes ocorre de forma semelhante entre os cursos agrupados e apresenta uma taxa de evasão maior que a real, uma vez que desconsidera estudantes que entraram por outra forma de ingresso. Para análises, variações das probabilidades de retenção e evasão podem ser passadas por parâmetros em diferentes semestres. Esses parâmetros, conforme são alterados, permitem simular e avaliar diferentes aspectos do progresso dos estudantes durante um curso.

A simulação será realizada de forma a representar a trajetória de 10000 estudantes, valor esse suficiente para que as métricas aferidas entrem em estabilidade, desde o primeiro semestre do curso até sua conclusão ou evasão, podendo passar por estados que refletem o comportamento de estudantes retidos. Ao final da simulação são gerados dados referentes ao tempo e probabilidade de retenção, evasão, conclusão e desvinculação. Então, os tempos para se tornar evadido, concluído e desvinculado são aplicados na análise de sobrevivência de Kaplan-Meier.

4. Resultados

Os resultados obtidos através das análises que serão apresentados a seguir permitem avaliar retenção, evasão, conclusão e vínculo. Isso ocorre por meio de simulações do progresso dos estudantes com Cadeia de Markov em conjunto com análise de sobrevivência.

As Figuras 2(a), 2(b) e 2(c), apresentam a análise de sobrevivência dos eventos evasão, conclusão e desvinculação, para os dois cursos agrupados nos dados. O progresso estudantil é representado até o vigésimo semestre e observa-se que a evasão e desvinculação apresentam um comportamento semelhante em relação ao tempo e probabilidade, conforme disponível na Tabela 1. Apesar disso, a conclusão apresenta valores que divergem, principalmente a partir do semestre quinze. Ao realizar testes de hipótese, o resultado para desvinculação é um p -valor igual a 0,55, onde aceita a hipótese nula de

que os dois cursos são equivalentes. Porém, para a conclusão e evasão, o valor do p -valor é inferior a 0,05, consequentemente rejeita a hipótese nula.

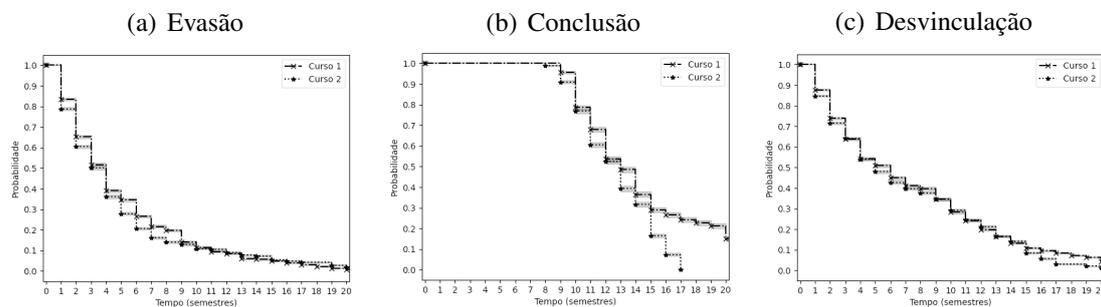


Figura 2. Análise de sobrevivência da evasão, conclusão e desvinculação para o Curso 1 e 2.

Tabela 1. Probabilidade e tempo por curso.

	Probabilidade (%)		Tempo (semestres)	
	Curso 1	Curso 2	Curso 1	Curso 2
Retenção	73,66	66,18	3,11	3,39
Evasão	73,92	70,12	4,95	4,37
Conclusão	21,20	27,94	13,05	12,75
Desvinculação	95,12	98,06	6,76	6,75

Ao utilizar os dados agrupados dos dois cursos apresentados anteriormente, a Figura 3(a) apresenta a análise de sobrevivência dos eventos evasão, conclusão e desvinculação, representando o progresso estudantil até o vigésimo semestre. Nesta análise, que será utilizada como cenário de referência, observa-se que a maior parte dos estudantes se desvinculam antes mesmo de algum deles concluir o curso. No quinto semestre, cerca de 50% dos estudantes não estão mais vinculados ao curso e após o décimo sexto período menos de 10% ainda mantém o vínculo. Em relação à conclusão, essa começa a ocorrer a partir do oitavo semestre, enquanto no décimo sexto semestre cerca de 80% dos estudantes que irão graduar, já o terão feito. Além disso, a maior probabilidade de evasão ocorre nos primeiros semestres, com aproximadamente 63% dos estudantes que irão evadir, evadindo até o final do quarto semestre, valor esse muito superior por semestre se comparado com os períodos seguintes.

Na Figura 3(b), é possível analisar as probabilidades de evasão, conclusão e vínculo em relação ao total de estudantes. Até o décimo quinto semestre 20% do total de estudantes já estão graduados, com acréscimo de apenas 4% até o vigésimo semestre. Além disso, 45% dos estudantes evadem até o final do quarto semestre. Dessa forma, propor formas de diminuir a evasão nos primeiros semestres pode impactar positivamente em todo o progresso estudantil.

Considerando que durante os quatro primeiros semestres a probabilidade de evadir é maior, na análise a seguir, a proposta é alterar a probabilidade de evasão para 50% em relação ao cenário de referência nesses semestres. Pela Figura 4(a), é possível constatar uma diminuição da probabilidade de evasão nos semestres iniciais que impacta em cerca de 7% na evasão final, além de apenas 40% dos estudantes terem evadido até o quarto

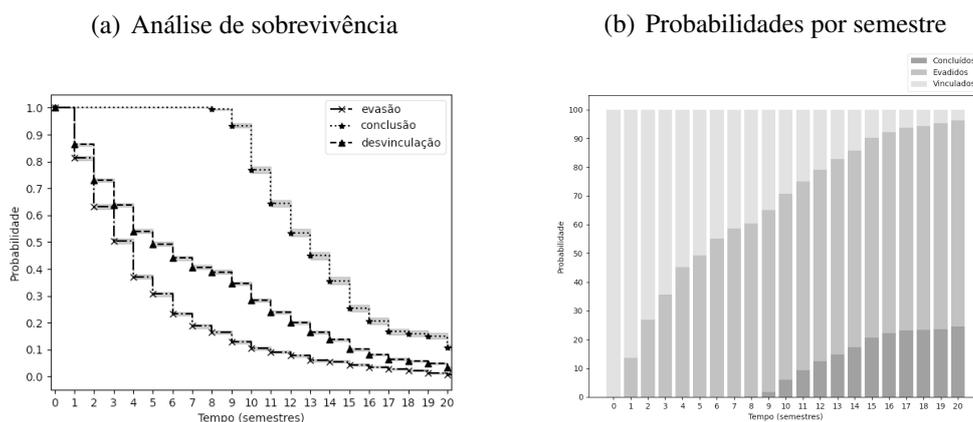


Figura 3. Análise de sobrevivência da evasão, conclusão e desvinculação e probabilidades por semestre para concluídos, evadidos e vinculados.

período. Consequentemente essas alterações influenciam na probabilidade final de conclusão e no vínculo dos estudantes conforme a probabilidade de evasão. Isso constata que diminuir em 50% a evasão apenas nos quatro primeiros períodos já apresenta um impacto positivo suficiente para modificar a evasão total. Portanto, isso permite planejar ações direcionadas para atender esse grupo de estudantes.

A retenção também pode ser considerada um fator importante na influência no vínculo do estudante. Na análise apresentada a seguir, a proposta é alterar a probabilidade de retenção para 50% do cenário de referência nos quatro primeiros períodos. Pela Figura 4(b), observa-se uma influência maior da retenção na evasão, conclusão e vínculo se comparado com a análise anterior da probabilidade de evasão. Isso acontece devido ao comportamento dos alunos retidos, que apresentam maior probabilidade de evasão. Apesar da evasão ainda se manter elevada nos dois primeiros períodos de curso, ao longo dos semestres seguintes ela diminui e apresenta uma redução final de aproximadamente 12%. Isso representa uma diferença de 5% em comparação com a análise anterior.

Uma alternativa as análise anteriores é alterar a probabilidade de ambos, retenção e evasão, para 50% do cenário de referência nos quatro primeiros períodos. Dessa forma, conforme observado na Figura 4(c), isso implica em uma diminuição ainda maior da probabilidade de evasão por semestre, com uma menor probabilidade de desvinculação nos primeiros períodos.

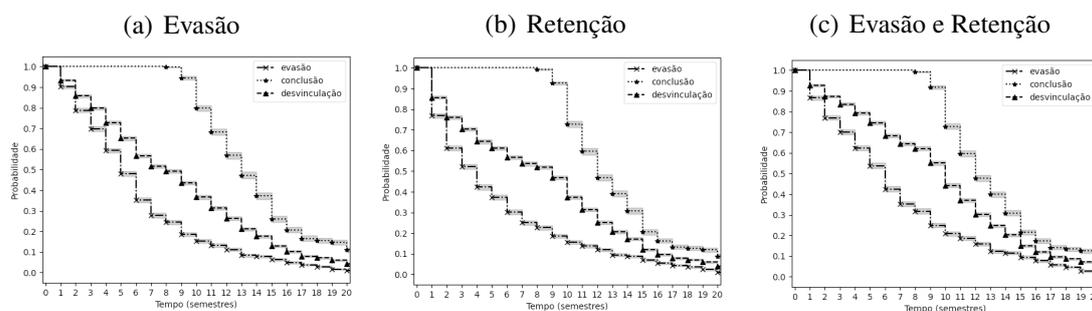


Figura 4. Análise de sobrevivência da evasão, conclusão e desvinculação para uma redução de 50% na evasão ou retenção nos quatro primeiros semestres.

Os resultados disponíveis na Tabela 2, demonstram como a alteração das probabilidades apenas nos primeiros quatro períodos pode influenciar nas probabilidades finais de retenção, evasão, conclusão e vínculo. Ao diminuir a evasão, houve um aumento da retenção, enquanto em todas as análises existiu uma maior probabilidade de conclusão e diminuição da evasão. Porém, essas alterações tiveram pouco impacto na probabilidade de desvinculação, já que mesmo ao diminuir a retenção ou evasão a maior parte dos estudantes se desvinculam muito antes do tempo de censura de 20 períodos.

Tabela 2. Probabilidade das análises para os quatro primeiros semestres de curso.

	Probabilidade (%)			
	Referência	50% da Evasão	50% da Retenção	50% da Retenção e Evasão
Retenção	70,80	79,40	60,50	68,80
Evasão	72,40	65,60	60,20	53,40
Conclusão	24,10	29,40	34,80	40,60
Desvinculação	96,50	95,00	95,00	94,00

Na Tabela 3, observa-se os tempos por semestre ao realizar as análises nos quatro primeiros períodos. Constata-se, uma diminuição do tempo para conclusão ao reduzir a retenção, assim como um aumento do tempo para desvinculação ao diminuir tanto a retenção, quanto a evasão. Dessa forma, é possível criar um planejamento melhor elaborado para organizar vagas extras nas turmas, com base na previsão de como esses fatores podem influenciar no tempo de vínculo do estudante.

Tabela 3. Tempo das análises para os quatro primeiros semestres de curso.

	Tempo (semestres)			
	Referência	50% da Evasão	50% da Retenção	50% da Retenção e Evasão
Retenção	3,20	3,15	4,72	4,63
Evasão	4,74	6,06	5,18	6,63
Conclusão	13,00	13,07	12,53	12,62
Desvinculação	6,81	8,23	7,87	9,22

5. Considerações Finais

Este estudo apresentou um modelo de Cadeia de Markov para simular o progresso dos estudantes durante um curso de graduação. As análises realizadas foram avaliadas através de análise de sobrevivência e permitiram verificar como a probabilidade de evasão e retenção nos primeiros semestres de curso influencia na evasão final. Além de observar o comportamento entre os dois cursos analisados.

Nas principais contribuições, pode-se verificar como a probabilidade de evasão e retenção nos quatro primeiros semestres do curso influenciam as probabilidades e os tempos de conclusão, evasão e desvinculação. Dessa forma, adotar políticas que foquem na diminuição da evasão nesses semestres de curso contribui para uma redução significativa da evasão final, apesar de um pequeno aumento na retenção e nos tempos de evasão,

retenção e vínculo. Portanto, permite criar um planejamento para direcionar esforços em relação a esse grupo específico de estudantes. Enquanto a retenção tem um efeito ainda mais positivo nas probabilidades de evasão, conclusão e desvinculação dos estudantes, com um impacto menor no tempo.

Como trabalho futuro, pretende-se acrescentar categorias, tais como cidade, gênero, estado civil e cor/raça. Com objetivo de analisar como esses aspectos influenciam no tempo e na probabilidade de evasão, conclusão e vínculo. Além de criar os parâmetros dos modelos conforme grupos de estudantes, produzidos a partir de algoritmos de aprendizagem de máquina. Isso permite previsões mais precisas e individualizadas conforme as características dos estudantes.

Referências

- Araújo, D. F. d. (2019). *Análise e avaliação da trajetória de estudantes de graduação baseadas em modelagem por cadeias de Markov*. PhD thesis, Programa de Pós-Graduação em Biometria e Estatística Aplicada. Departamento de Estatística e Informática.
- Beltran, C. A. R., Xavier-Júnior, J. C., Barreto, C. A., and Neto, C. O. (2019). Plataforma de aprendizado de máquina para detecção e monitoramento de alunos com risco de evasão. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 1591.
- Boumi, S. and Vela, A. E. (2020). Improving graduation rate estimates using regularly updating multi-level absorbing markov chains. *Education Sciences*, 10(12):377.
- Brezavšček, A., Bach, M. P., and Baggia, A. (2017). Markov analysis of students' performance and academic progress in higher education. *Organizacija*, 50(2):83–95.
- Campos, J. D. d. S. (2016). Fatores explicativos para a evasão no ensino superior através da análise de sobrevivência: o caso da ufpe. Master's thesis, Universidade Federal de Pernambuco.
- Carrano, D., De Albergaria, E. T., Infante, C., and Rocha, L. (2019). Combinando técnicas de mineração de dados para melhorar a detecção de indicadores de evasão universitária. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 1321.
- Costa, F. J. d., Bispo, M. d. S., and Pereira, R. d. C. d. F. (2018). Dropout and retention of undergraduate students in management: a study at a brazilian federal university. *RAUSP Management Journal*, 53(1):74–85.
- Da Silva, A. M. and Santos, B. C. S. (2017). Eficácia de políticas de acesso ao ensino superior privado na contenção da evasão. *Avaliação: Revista da Avaliação da Educação Superior*, 22(3):741–757.
- De Almeida Teodoro, L. and Kappel, M. A. A. (2020). Aplicação de técnicas de aprendizado de máquina para predição de risco de evasão escolar em instituições públicas de ensino superior no brasil. *Revista Brasileira de Informática na Educação*, 28:838–863.
- De Brito, B. C. P., De Mello, R. F. L., and Alves, G. (2020). Identificação de atributos relevantes na evasão no ensino superior público brasileiro. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 1032–1041. SBC.

- De Jesus, H. O., Rodriguez, L. C., and Junior, A. d. O. C. (2021). Predição de evasão escolar na licenciatura em computação. *Revista Brasileira de Informática na Educação*, 29:255–272.
- Dos Santos, D. C. V.-B. and Falão, T. P. (2017). Acompanhamento de alunos em ambientes virtuais de aprendizagem baseado em sistemas tutores inteligentes. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 1267.
- Kleinbaum, D. G. and Mitchel, K. (2010). *Survival analysis*, volume 3. Springer.
- Lima, E. and Coutinho, E. (2019). Uma análise sobre o desempenho de alunos de graduação em disciplinas iniciais de programação. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 1876.
- Manhães, L. M. B., Da Cruz, S. M. S., Costa, R. J. M., Zavaleta, J., and Zimbrão, G. (2012). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In *Brazilian symposium on computers in education (simpósio brasileiro de informática na educação-sbie)*, volume 1.
- Queiroga, E., Cechinel, C., and Araújo, R. (2017). Predição de estudantes com risco de evasão em cursos técnicos a distância. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 1547.
- Ramos, J. L. C., Gomes, A. S., Rodrigues, R., Silva, J., De Souza, F. d. F., De Gouveia Zambom, E., and Prado, L. (2017). Um modelo preditivo da evasão dos alunos na ead a partir dos construtos da teoria da distância transacional. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 1227.
- Rigo, S. J., Cambuzzi, W., Barbosa, J. L., and Cazella, S. C. (2014). Aplicações de mineração de dados educacionais e learning analytics com foco na evasão escolar: oportunidades e desafios. *Revista Brasileira de Informática na Educação*, 22:132.
- Rodrigues, F. S., Brackmann, C. P., and Barone, D. A. C. (2015). Estudo da evasão no curso de ciência da computação da ufrgs. *Revista Brasileira de Informática na Educação*, 23.
- Saccaro, A., França, M. T. A., and Jacinto, P. d. A. (2019). Fatores associados à evasão no ensino superior brasileiro: um estudo de análise de sobrevivência para os cursos das áreas de ciência, matemática e computação e de engenharia, produção e construção em instituições públicas e privadas. *Estudos Econômicos (São Paulo)*, 49(2):337–373.
- Sheskin, T. J. (2016). *Markov chains and decision processes for engineers and managers*. CRC press.
- Xavier, M. and Meneses, J. (2020). A literature review on the definitions of dropout in online higher education. In *EDEN Conference Proceedings*, pages 73–80.