



**UNIVERSIDADE
FEDERAL RURAL
DE PERNAMBUCO**



Lhaíslla Eduarda Cavalcanti Rodrigues da Silva

Uso de Análise de Sobrevivência como Ferramenta na Obtenção de Indicadores de Permanência no Ensino Superior

Recife

2024

Lhaíslla Eduarda Cavalcanti Rodrigues da Silva

Uso de Análise de Sobrevivência como Ferramenta na Obtenção de Indicadores de Permanência no Ensino Superior

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Universidade Federal Rural de Pernambuco – UFRPE
Departamento de Estatística e Informática
Curso de Bacharelado em Sistemas de Informação

Orientador: Gabriel Alves de Albuquerque Júnior

Recife
2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

L688Cav
alcanti
Rodrigues
s da
Silvau

Cavalcanti Rodrigues da Silva, Lhaislla Eduarda
Uso de Análise de Sobrevivência como Ferramenta na Obtenção de Indicadores de Permanência no
Ensino Superior / Lhaislla Eduarda Cavalcanti Rodrigues da Silva. - 2024.
66 f. : il.

Orientador: Gabriel Alves de Albuquerque Junior.
Inclui referências e apêndice(s).

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Sistemas da Informação, Recife, 2024.

1. Análise de sobrevivência. 2. Business Intelligence. 3. Permanência estudantil. I. Junior, Gabriel Alves
de Albuquerque, orient. II. Título

CDD 004

Uso de Análise de Sobrevivência como Ferramenta na Obtenção de Indicadores de Permanência no Ensino Superior

Monografia apresentada ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Sistemas de Informação.

Aprovada em: 06 de Março de 2024

BANCA EXAMINADORA

Gabriel Alves de Albuquerque Júnior

Departamento de Estatística e Informática

Universidade Federal Rural de Pernambuco – UFRPE

Cláudio Tadeu Cristino

Departamento de Estatística e Informática

Universidade Federal Rural de Pernambuco – UFRPE

Agradecimentos

Agradeço de coração a todos que estiveram ao meu lado durante esta jornada acadêmica, pois ninguém faz nada sozinho e sem apoio de cada um não teria sido possível alcançar o sucesso acadêmico. Primeiramente, expresso minha gratidão a Deus por guiar meus passos ao longo desta trajetória. Ao meu orientador, sou imensamente grata pela sua paciência, orientação e pelos materiais indicados que contribuíram para meu crescimento intelectual, pessoal e profissional. Aos meus familiares, amigos e até mesmo meu gato, agradeço pelo suporte emocional, financeiro e moral que me proporcionaram durante os desafios da graduação. Não posso deixar de mencionar meus professores, por todo conhecimento compartilhado ao longo da jornada acadêmica. Cada gesto de encorajamento foi um impulso em direção ao meu objetivo.

“Ensinar não é transferir conhecimento, mas criar as possibilidades para a sua própria produção ou a sua construção.”
(Paulo Freire)

Resumo

Este trabalho propõe uma abordagem integrada de técnicas estatísticas, *Business Intelligence* e conceitos de *Data Science* para criar um modelo de análise de sobrevivência visando compreender a permanência dos estudantes no ensino superior. Utilizando cursos de computação como estudo de caso, diferentes grupos são comparados para calcular a probabilidade de permanência dos discentes até o término do curso. O trabalho faz uso da plataforma SABIA para dar suporte à gestão orientada a dados, destacando a importância das ferramentas tecnológicas na análise acadêmica. Os resultados revelam padrões entre os cursos de uma mesma área ao considerar a evasão como evento de interesse, apresentando quedas mais acentuadas nos períodos iniciais, especialmente em momentos críticos, como no segundo semestre, que apresentam probabilidades de permanência mais baixas em relação ao primeiro semestre, assim como o a modalidade de ingresso *Outros* que nos períodos finais refletem ocorrência da evasão, evidenciando a importância de intervenções personalizadas para evitar a evasão e desperdícios de recursos, contribuindo para uma gestão mais eficiente e eficaz das instituições de ensino superior.

Palavras-chave: Análise de sobrevivência, Business Intelligence, Permanência estudantil.

Abstract

This paper proposes an integrated approach of statistical techniques, business intelligence and data science concepts to create a survival analysis model aimed at understanding student retention in higher education. Using computer science courses as a case study, different groups are compared in order to calculate the probability of students staying until the end of the course. The work makes use of the SABIA platform to support data-driven management, highlighting the importance of technological tools in academic analysis. The results show patterns between courses in the same area when considering dropout as an event of interest, with steeper drops in the initial periods, especially at critical moments, such as the second semester, which show lower probabilities of permanence compared to the first semester, as well as the *Other* entry modality, which in the final periods reflects the occurrence of dropout, highlighting the importance of personalised interventions to prevent dropout and waste of resources, contributing to more efficient and effective management of higher education institutions.

Keywords: Survival analysis, business intelligence, Student stay.

Lista de ilustrações

Figura 1 – Taxa de conclusão de estudantes que ingressaram em programa de licenciatura ou equivalente, 2017.	20
Figura 2 – Ciclo de matrícula.	20
Figura 3 – Etapas da análise estatística.	22
Figura 4 – Principais áreas relacionadas com a EDM.	26
Figura 5 – Etapas de uma EDM.	28
Figura 6 – Etapas de desenvolvimento de um Academic Analytics.	29
Figura 7 – Etapas da elaboração da plataforma de BI.	37
Figura 8 – Input dos dados.	42
Figura 9 – Output dos dados.	42
Figura 10 – Visualização inicial do SABIA.	43
Figura 11 – Filtros da plataforma SABIA.	44
Figura 12 – Filtros do modelo de sobrevivência.	44
Figura 13 – Gráfico resultante da função de sobrevivência.	45
Figura 14 – Gráficos de sobrevivência para os estudantes que ingressaram nos cursos de computação por semestres.	49
Figura 15 – Gráficos de sobrevivência para os estudantes por modalidade de ingresso nos cursos de computação.	52
Figura 16 – Gráficos de sobrevivência para os estudantes que ingressaram nos cursos de computação permaneceram matriculados no período analisado.	55
Figura 17 – Gráficos de sobrevivência para os estudantes que ingressaram por todos no evento de morte <i>FORMADO</i> e <i>EVADIDO</i>	59

Lista de tabelas

Tabela 1 – Ciclo do Learning Analytics	27
Tabela 2 – Funções do odelo de análise de sobrevivência	38
Tabela 3 – Kaplan meier lifelines	41
Tabela 4 – Probabilidades de evadidos por semestre.	51
Tabela 5 – Probabilidades de evadidos por modalidade de ingresso.	54
Tabela 6 – Probabilidades de sobrevivência por todos.	57
Tabela 7 – Probabilidades de sobrevivência - todos por evento de morte formado e evadido.	61

Lista de abreviaturas e siglas

IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas
SABIA	Academic Business Intelligence and Analytics
ENADE	Exame Nacional de Desempenho dos Estudantes
OECD	Organização para a Cooperação e Desenvolvimento Económico
BI	Business Intelligence
LA	Learning Analytics
AA	Academic Analytics
EDM	Educational Data Mining
BSI	Bacharelado em Sistemas de Informação
BCC	Bacharelado em Ciência da Computação
LC	Licenciatura em Computação
SISTEC	Sistema Nacional de Informações da Educação Profissional e Tecnológica
OLAP	Analytical Processing
DW	Data Warehouse
DM	Data Mart
IGC	Índice Geral de Cursos
TSD	Taxa de Sucesso de Disciplina

Sumário

	Lista de ilustrações	8
1	INTRODUÇÃO	13
1.0.1	Motivação e Justificativa	15
1.0.2	Objetivos	17
1.0.3	Organização do trabalho	18
2	REFERENCIAL TEÓRICO	19
2.0.1	Contextualizando a evasão no ensino superior	19
2.0.2	Análise de sobrevivência	21
2.0.3	Fundamentos do Business Intelligence	25
2.0.4	Data Science	26
2.0.4.1	Learning Analytics (LA) e Educational Data Mining (EDM)	26
2.0.4.2	Academic Analytics (AA)	28
2.0.5	SABIA	30
3	TRABALHOS RELACIONADOS	31
3.0.1	Análise de sobrevivência	31
3.0.2	Business Intelligence	33
4	MATERIAIS E MÉTODOS	36
4.0.1	Definição e formatação dos dados	36
4.0.2	Modelo de análise de sobrevivência	37
5	RESULTADOS	43
5.0.1	Plataforma BI	43
5.0.1.1	Aplicações da ferramenta	45
5.0.2	Curvas de sobrevivência	46
5.0.3	Análise de sobrevivência para os cursos de computação	47
5.0.3.1	Semestre	49
5.0.3.2	Modalidade de Ingresso	52
5.0.3.3	Todos	55
5.0.3.4	Todos por evento de morte formado e evadido	58
5.0.4	Considerações Finais	62
6	DISCUSSÕES E TRABALHOS FUTUROS	64

REFERÊNCIAS	66
--------------------------	-----------

1 Introdução

A evasão é um problema recorrente no contexto do ensino superior brasileiro, impactando negativamente o progresso acadêmico (SOUZA; KOMATI; ANDRADE, 2022). Este fenômeno não se restringe apenas às instituições do Brasil; ocorre também internacionalmente, afetando os sistemas educacionais. Além dos prejuízos financeiros para as instituições, a evasão afeta a entrada de profissionais qualificados no mercado de trabalho (HOED, 2016).

Pesquisas internacionais enfatizam a complexidade do fenômeno do abandono universitário (BEHR et al., 2020). Na Alemanha essa tendência é que cerca de 28% dos estudantes de qualquer ano desistam dos estudos em programas de bacharelado em instituições de ensino superior (HEUBLEIN, 2014). Enquanto na Espanha a taxa de abandono precoce, conforme revelada pelo Inquérito às Forças de Trabalho é 16%, acima da meta estabelecida pela União Europeia (SANTOS-VILLALBA et al., 2023).

A evasão refere-se à saída antes da conclusão do ano, série ou ciclo, por desistência, independentemente do motivo. Assim, a evasão representa uma falha em impulsionar o aluno para uma condição superior de profissional qualificado, limitando o acesso ao aprimoramento de conhecimentos, ao desenvolvimento cognitivo de habilidades e competências no respectivo nível de ensino (INEP, 2017). Quanto ao tipo de evasão, pode estar ligada a um curso, a uma instituição específica ou até mesmo ao sistema educacional como um todo (SANTOS et al., 2022).

O desenvolvimento de políticas e ações eficazes para reduzir o abandono é um desafio constante enfrentado pelos gestores universitários, presente em problemas vinculados ao curso, como adaptação ao ambiente acadêmico, falta de suporte acadêmico, conflitos entre trabalho e estudo, e dificuldades de acesso à tecnologia (SILVA, 2013). Fatores socioeconômicos incluindo questões pessoais e financeiras impõem desafios à continuidade dos estudos (SACCARO; FRANÇA; JACINTO, 2019). A investigação das causas relacionadas à evasão utiliza diferentes abordagens que consideram aspectos quantitativos e qualitativos (CAMPOS, 2016).

A análise de sobrevivência é uma ferramenta estatística usada para entender padrões de evasão no ensino superior, investigando o tempo até a ocorrência de um evento (HOED, 2016). Embora o estimador de Kaplan-Meier seja uma abordagem não paramétrica comumente empregada nesse contexto, é importante notar que há também técnicas paramétricas disponíveis para análise de sobrevivência. Em métodos não paramétricos, como o Kaplan-Meier, não se assume uma forma específica para a distribuição das probabilidades dos eventos. Em contraste, abordagens paramétri-

cas se baseiam em modelos que fazem suposições sobre a distribuição (COLOSIMO; GIOLO, 2021). Isso ajuda a identificar características que afetam a probabilidade de permanência ao longo do tempo.

Complementando essa perspectiva, o *Business Intelligence (BI)* destaca-se como uma das abordagens empregadas no processamento de dados no ensino superior (ALMEIDA; CAMARGO, 2015). Ao integrar as informações coletadas pela análise de sobrevivência com dados de diversas fontes, junto a técnicas de processamento e visualização de dados, possibilita a criação de *dashboards* interativos. Adicionalmente, a aplicação de *Data Science* na educação amplia a capacidade analítica, permitindo uma compreensão mais profunda dos padrões de permanência.

Este trabalho propõe a criação de um modelo de análise de sobrevivência a ser incorporado em uma plataforma de BI, com visualizações e análises de dados, de forma ágil e intuitiva. O foco das análises se dá em um dos painéis dedicado à análise de permanência dos discentes no ensino superior. A plataforma em questão é o *Academic Business Intelligence and Analytics (SABIA)* (MARQUES et al., 2023). O painel específico é denominado *Permanência*, nele, são exibidas as probabilidades de sobrevivência dos discentes ao longo do curso, filtradas por grupos da instituição. Dessa forma, busca-se fazer uso de uma solução que traga mais eficiência para a tomada de decisões na gestão acadêmica, que possibilite identificar características mais relevantes na evasão dos alunos, realizando análises por meio de comparações e uma gestão fundamentada em evidências.

O processo de construção no âmbito da plataforma SABIA envolve o desafio de integrar diversas fontes de dados. Além das ferramentas de análise de sobrevivência e *BI*, o projeto incorpora conceitos fundamentais de *Data Science* na educação, tendo como subcampos o *Learning Analytics (LA)*, *Educational Data Mining (EDM)* e *Academic Analytics (AA)*. O *LA*, visa extrair inferências a partir de dados educacionais, utilizando métodos estatísticos e algoritmos para compreender o desempenho dos alunos e otimizar o processo de ensino-aprendizagem (PALOMINO et al., 2022). O *EDM* se concentra na aplicação de técnicas de mineração de dados e análise de dados em ambientes educacionais para melhorar o processo de ensino e aprendizagem e informar práticas pedagógicas mais eficazes e personalizadas (CARMINATI et al., 2020). Por outro lado, o *AA* concentra-se na análise de dados institucionais para melhorar a eficácia e a eficiência das práticas acadêmicas (ANDRADE; FERREIRA, 2016).

No decorrer deste trabalho, serão apresentados detalhes do processo de construção do painel web na plataforma SABIA, destacando as etapas que conduziram à concepção e implementação do painel. Isso incluirá desde a seleção e preparo dos dados até a elaboração do modelo, implementação e visualização dos resultados. O escopo deste processo aborda não apenas a construção técnica da página, mas também

o compartilhamento de algumas análises de sobrevivência em cursos de computação, referentes aos grupos, modalidade de ensino, semestre e todos.

Espera-se, com este trabalho, utilizar uma ferramenta de monitoramento para obter indicadores de sucesso estudantil, com o objetivo de auxiliar gestores e coordenadores na redução de desperdícios de recursos associados à evasão dos alunos nos cursos. Ao questionar se cursos da mesma área demonstram comportamentos similares, a ferramenta SABIA é empregada para comparar grupos e obter curvas de permanência, permitindo analisar tendências representadas pela proporção de alunos que permanecem no curso em cada intervalo para observar variáveis de interesse.

Identificando os fatores que influenciam a permanência dos alunos nos cursos, os gestores podem tomar decisões informadas para prevenir a evasão, resultando em uma utilização mais eficiente dos recursos disponíveis. A comparação entre grupos e o cálculo das curvas de permanência para cada um possibilitam discernir se cursos similares compartilham comportamentos similares, contribuindo para uma compreensão mais abrangente da dinâmica estudantil. A análise de tendências e probabilidades de sucesso em intervalos de tempo específicos proporciona uma visão detalhada da variação da permanência dos alunos ao longo do tempo e entre diferentes grupos, auxiliando nas decisões. Embora as curvas de permanência e as probabilidades de sucesso forneçam uma compreensão agregada do comportamento dos alunos, é importante ressaltar que essas análises não identificam alunos individuais que continuam no curso, mas oferecem uma visão geral do sucesso estudantil que justifica a proposição de intervenções acadêmicas.

1.0.1 Motivação e Justificativa

A análise internacional dos desafios educacionais ultrapassa fronteiras e tem se tornado cada vez mais relevante para governos e instituições de ensino, à medida que buscam desenvolver políticas eficazes para melhorar as perspectivas sociais, econômicas e educacionais dos indivíduos. Nesse contexto, a Organização para a Cooperação e Desenvolvimento Econômico (OECD) fornece indicadores quantitativos internacionais para comparação, por meio de relatórios anuais como o *Education at a Glance*, que compilam e orientam a formulação de políticas educacionais em nível global (OECD, 2021).

Observando o panorama brasileiro, o abandono universitário é classificado como uma preocupação persistente, que vai além de questões individuais, refletindo também desafios estruturais e socioeconômicos. Apesar dos avanços registrados nos percentuais de conclusão do ensino superior, com aumento de 44% para 67% em 2021, ainda persistem desafios, especialmente no primeiro ano. A pesquisa aponta para razões como desinteresse nos estudos e a necessidade de ingresso precoce no mercado de

trabalho, o mesmo padrão é encontrado em outros países da América Latina e no mundo (OECD, 2021). De acordo com o Censo da Educação Superior divulgado pelo MEC, que apresenta indicadores de trajetória dos cursos anualmente, as taxas de desistência acumulada atingiram 59% na rede pública em 2022 (INEP, 2023).

A formação de profissionais qualificados é fundamental para a manutenção econômica de um país, uma vez que a qualidade do trabalho desempenhado está relacionada ao grau de conhecimento e formação. Contudo, o abandono universitário representa uma interrupção no processo de desenvolvimento de aprendizado. À medida que a desistência permanece elevada, o potencial de desenvolvimento e inovação é prejudicado, comprometendo a capacidade do país de atender às demandas do mercado de trabalho e de se manter relevante em setores estratégicos.

Embora a experiência de trabalho possa aprimorar as habilidades e conhecimentos práticos dos alunos, uma carga de trabalho pesada pode impedir que os alunos se comprometam totalmente com seus estudos e, como resultado, muitos optam por abandonar o curso em detrimento do emprego. Garantir que os estudantes recebam o apoio de que precisam pode não só ajudá-los a tirar o máximo partido de sua experiência universitária, adquirir competências e conhecimentos avançados que impulsionarão suas vidas pessoais e profissionais, mas também pode contribuir para reduzir a evasão, diminuir o tempo de conclusão dos cursos e, principalmente, evitar o desperdício de investimentos por parte das instituições de ensino.

A desistência da graduação nas Instituições de Ensino Superior não apenas impacta a economia, mas também perpetua desigualdades sociais ao limitar o acesso de certos grupos a oportunidades educacionais. Investigar e analisar os motivos que contribuem para este fenômeno é determinante para implementar estratégias que visem reduzir desistências e promover uma educação superior mais inclusiva, acessível e menos custosa às instituições.

A falta de acompanhamento preciso e personalizado da trajetória do discente ao longo de sua permanência na instituição dificulta a identificação da evasão estudantil, acarretando custos sociais e econômicos, além de ser um fator que influencia negativamente o quantitativo de profissionais devidamente capacitados e formados para atender às demandas do mercado de trabalho. Diante disso, é relevante que haja meios de monitorar e avaliar os índices de desenvolvimento da Educação que possibilitem realizar comparações ao longo do tempo e forneçam medidas para apoio, informações de melhores políticas e práticas, para o reconhecimento de padrões e fatores que levam à evasão e à formação.

Nesse contexto, a compreensão das características que levam ao fim do vínculo acadêmico torna-se um importante indicador para pesquisas de permanência estudantil na educação superior. O presente trabalho busca investigar as estatísticas que refle-

tem o problema ao longo dos anos para fornecer insumos que embasem a elaboração de estratégias de otimização de recursos.

Assim, a utilização de um modelo de análise de permanência estudantil em um painel de gestão representa um diferencial potencializador na obtenção de indicadores. O painel proporciona a reunião de dados de diferentes fontes para uma visualização centralizada, oferecendo uma apresentação intuitiva por meio de gráficos e elementos visuais, o que facilita a compreensão. Adicionalmente, possibilita análises comparativas dos resultados entre diferentes grupos e permite aos usuários interagirem com os dados, filtrando informações e explorando diferentes perspectivas, facilitando a identificação de tendências e padrões.

Enquanto a análise de sobrevivência permite calcular probabilidades relacionadas à ocorrência de eventos ao longo do tempo e ajuda a identificar fatores que influenciam a probabilidade de ocorrência de um evento específico, também possibilita modelar o comportamento de eventos ao longo do tempo, levando em consideração variáveis relevantes. Por fim, permite analisar dados históricos para entender como a permanência dos estudantes evolui ao longo do tempo. O uso dos agrupamentos proporciona explorar diferentes perspectivas e assim, avaliar o impacto de intervenções ou estratégias na ocorrência de eventos. Essa visão mais completa e estratégica atribui a gestores e tomadores de decisão não apenas a compreensão da situação atual de forma mais clara, mas também a possibilidade de antecipar cenários futuros comparando situações do passado a atuais e implementar medidas preventivas e corretivas de maneira mais eficaz.

1.0.2 Objetivos

O objetivo deste trabalho é proporcionar uma ferramenta de monitoramento que indique a probabilidade de permanência na graduação por meio da análise de sobrevivência, investigando as variáveis de influência que impactam a evasão estudantil no ensino superior. Como forma de garantir o alcance do objetivo geral definido, alguns pontos específicos são levados em consideração no processo:

- Criar um modelo de análise de Sobrevivência;
- Identificar fatores relevantes para a análise de sobrevivência;
- Implementar um painel de *Business Intelligence* para visualização de resultados na plataforma SABIA;
- Elaborar estudos de caso em cursos de computação;
- Realizar análises sobre diferentes grupos: semestre, modalidade de ingresso e todos.

Ao abordar esses pontos específicos, este trabalho visa contribuir para uma compreensão mais aprofundada da permanência no ensino superior. Pretende-se comparar o comportamento de grupos entre cursos da mesma área. As análises auxiliarão na obtenção de probabilidades de permanência para a formulação de políticas e estratégias ao lidar com a evasão estudantil.

1.0.3 Organização do trabalho

Este trabalho é dividido em seis seções. A Seção 1 apresentou a introdução ao objeto de trabalho, incluindo a justificativa e os objetivos que motivaram sua realização. Na Seção 2, são apresentados os conceitos fundamentais utilizados neste trabalho. Na seção 3 é realizada uma análise de trabalhos relacionados, abordando pesquisas anteriores sobre o uso de análise de sobrevivência, BI na visualização de informações e análise de dados educacionais no ensino superior. A seção 4 detalha os materiais e métodos empregados em sua condução. A seção 5 descreve os resultados obtidos, em seguida a seção 6 apresenta as considerações finais e discute possíveis trabalhos futuros.

2 Referencial Teórico

Nesta seção, são apresentados conceitos importantes relacionados ao tema tratado. Os tópicos abordam: evasão estudantil; análise de sobrevivência, que explora a duração até um evento ocorrer; *Data science*, que utiliza métodos científicos, processos, algoritmos e sistemas para extrair conhecimento a partir de dados; e *Business Intelligence*, um conjunto de ferramentas para transformar dados brutos em informações estratégicas. Esses conceitos são essenciais para embasar e executar este trabalho, contribuindo para uma análise fundamentada.

2.0.1 Contextualizando a evasão no ensino superior

A evasão é um desafio persistente em diversas universidades brasileiras, representa um problema de proporções internacionais e acarreta prejuízos para os países e instituições de ensino. A desistência dos estudantes de seus cursos impacta as IES, a economia, desenvolvimento social e a gestão dos recursos investidos. Existem diferentes formas de interpretar o problema, podendo ser uma evasão de curso, quando o estudante se desliga do curso superior em situações diversas, como abandono, desistência, transferência ou reopção, exclusão por norma institucional; evasão da instituição, quando o estudante se desliga da instituição na qual está matriculado; e evasão do sistema, quando o estudante abandona de forma definitiva ou temporária o ensino superior (ANDIFES, 1996).

Do ponto de vista econômico, os alunos desvinculados geram uma importante perda de receitas, uma vez que o investimento feito em estudantes que vêm a evadir não apresenta retorno. Neste contexto, é considerada evasão quando um estudante abandona definitivamente seu curso, em qualquer estágio do processo educacional.

A OECD, partindo de uma análise do número de concluintes, identificou uma taxa de formação de 44% no Brasil, nos cursos de licenciatura, ainda pequena se comparada a países como Reino Unido e Irlanda, conforme representado na Figura 1. O que sobra corresponde a 66%, representando a evasão no país (OECD, 2021).

Para mensurar o desempenho acadêmico e obter indicadores do fluxo educacional nas instituições, são estabelecidos critérios para as situações acadêmicas como forma de representar o estado atual do estudante e identificar sua trajetória no curso. Tanto o INEP quanto o Sistema Nacional de Informações da Educação Profissional e Tecnológica (SISTEC) registram os diferentes estados do último vínculo do aluno no curso, fornecendo subsídios para o desenvolvimento de indicadores. A Figura 2 representa o mapeamento ao longo do ciclo de matrícula.

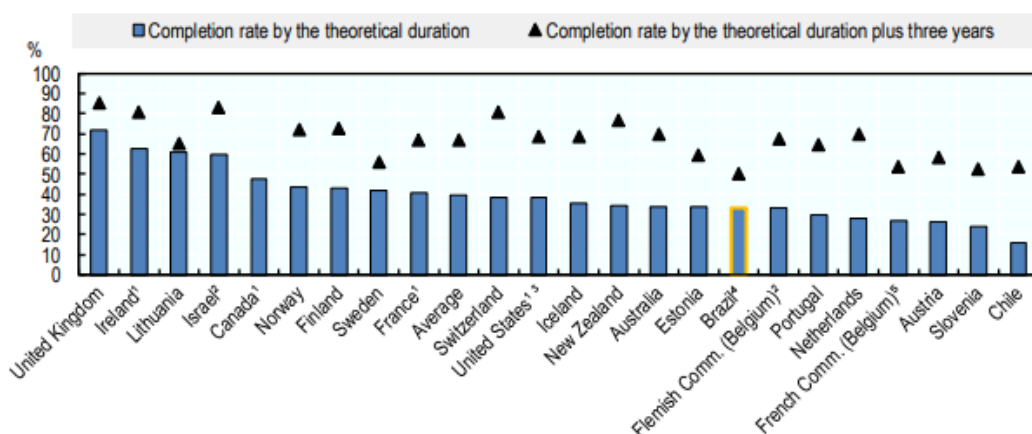


Figura 1 – Taxa de conclusão de estudantes que ingressaram em programa de licenciatura ou equivalente, 2017.

Fonte: (OECD, 2021).

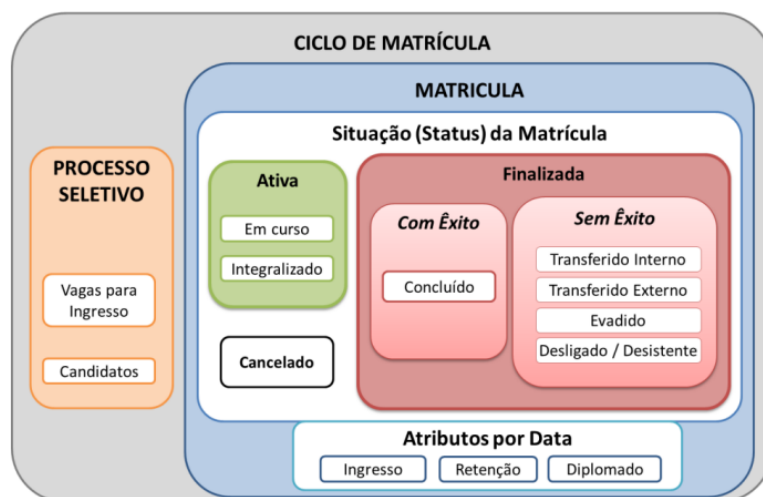


Figura 2 – Ciclo de matrícula.

Fonte: (Ministério da Educação - Secretaria de Educação Profissional e Tecnológica, 2014).

Para compreender os diferentes estados em relação à formação superior, é importante considerar conceitos como permanência, desistência e conclusão. O sucesso e o insucesso representam o encerramento do percurso educacional. Em alguns casos, o conceito de permanência e retenção podem se confundir, contudo, a retenção muitas vezes é usada no sentido de o aluno não progredir na velocidade ideal, devido a trancamentos ou reprovação em disciplinas (INEP, 2017). Para os propósitos deste trabalho, a permanência refere-se ao período de tempo em que um aluno permanece matriculado e ativo em um curso, incluindo o tempo desde a matrícula inicial até a desvinculação.

Ao analisar os dados do Censo, identificam-se diferentes situações dos estudantes no contexto do ensino superior. A categoria *FORMADO* abrange alunos que

completaram todos os créditos acadêmicos necessários para a titulação no curso até a data de referência do Censo, sem necessariamente terem participado da colação de grau ou do Exame Nacional de Desempenho de Estudantes (Enade). Por outro lado, a situação *CURSANDO* refere-se aos alunos que ainda não alcançaram a carga horária necessária para concluir o curso no ano de referência do Censo. Além disso, o Censo registra casos de desvinculação do curso, que podem ocorrer devido à evasão, abandono, desligamento ou transferência para outra IES. Outras categorias incluem alunos com matrícula trancada na data de referência do Censo, indicando uma interrupção temporária na sua vinculação com a instituição. Também são registrados alunos transferidos para outro curso de graduação na mesma IES, representando uma mudança de área de estudo dentro da mesma instituição.

Para este trabalho, será adotada a definição de evasão como o desligamento do estudante do curso. Os conceitos servem para mensurar o tempo que um estudante leva até realizar a transição para um estado definitivo, seja deixando o curso, independentemente da forma como essa transição ocorre. No entanto, é importante não apenas acompanhar o fluxo de estudantes na instituição, mas também identificar as variáveis associadas ao abandono do curso e o momento em que os alunos estão mais propensos a desistir. Essa análise é fundamental para prevenir a evasão estudantil, sendo o foco deste trabalho o uso da análise de sobrevivência para explorar os aspectos temporais desse fenômeno.

2.0.2 Análise de sobrevivência

Uma abordagem estatística amplamente empregada em diversas áreas é a análise de sobrevivência. Essa técnica é utilizada em áreas, como engenharia, ciências sociais e medicina (HABERMAN, 1981). A análise de sobrevivência concentra-se no estudo do tempo de sobrevivência, analisando o intervalo entre eventos (KLEINBAUM et al., 2012). O tempo varia conforme a área de estudo: na engenharia, pode representar o tempo até a falha de um componente; nas ciências sociais, os estudos focam em situações que têm como resposta o tempo entre eventos, como estudar a duração dos casamentos e os fatores que influenciam a estabilidade ou a ruptura das relações conjugais, examinar a duração do emprego em diferentes setores e ocupações, ou estudar movimentos sociais, para examinar a duração e a persistência dos movimentos sociais, bem como identificar os fatores que influenciam sua mobilização e desmobilização; na medicina, pode representar o tempo até a morte de um paciente (COLOSIMO; GIOLO, 2021).

Durante a condução de análises estatísticas, seguem-se etapas em um processo estruturado, conforme mostrado na Figura 3. Na fase inicial, são formuladas hipóteses de interesse e ocorre a definição das questões a serem abordadas, estabe-

lecendo hipóteses ao fenômeno em estudo. Posteriormente, ocorre a coleta de dados e a aplicação de métodos estatísticos para validar ou refutar as hipóteses formuladas. Esse processo é contínuo, permitindo o aprimoramento das análises e ajustes à medida que as informações são obtidas.

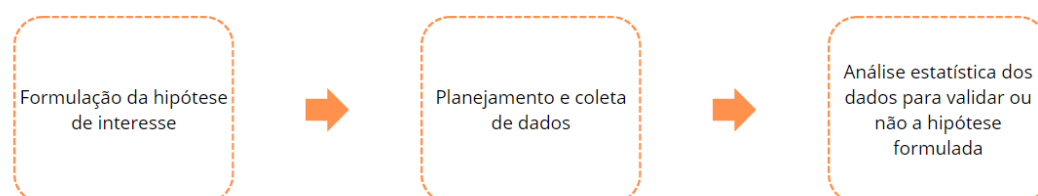


Figura 3 – Etapas da análise estatística.
Fonte: Adaptado de Colosimo e Giolo (2021).

Com o uso da Análise de Sobrevivência, a pesquisa direciona-se para situações em que a variável resposta é o tempo até a ocorrência de um evento de interesse. O tempo na estimativa do evento é considerado como intervalo desde o início da observação até a ocorrência do evento analisado. Nesse contexto, a Análise de Sobrevivência é representada pela função de sobrevivência. Esta função, simbolizada por $S(t)$, expressa a probabilidade de o evento observado e não falhar ao longo do tempo, ou seja, de *sobreviver*, levando em consideração que o tempo de sobrevivência seja superior a um determinado instante t (BASTOS; ROCHA, 2006).

A função de sobrevivência, denotada por $S(t)$, expressa a probabilidade de que um evento não tenha ocorrido até um tempo específico t . Pr representa a probabilidade, e T é a variável aleatória que indica o tempo de ocorrência (COLOSIMO; GIOLO, 2021), conforme descrito na Equação 2.1.

$$S(t) = Pr(T > t) \quad (2.1)$$

Essa técnica baseia-se em conceitos fundamentais, como tempo, evento, escala e censura. Como característica dos dados de análise de sobrevivência, o tempo de início é a origem do estudo, enquanto o evento refere-se ao fenômeno de interesse que está sendo estudado. A escala refere-se ao intervalo ou unidade de medida utilizada na análise, geralmente expressa em tempo real, como dias, meses ou anos, representando o período de observação (BASTOS; ROCHA, 2006).

Além da função de sobrevivência, outro conceito essencial é a taxa de falha, que representa a probabilidade instantânea de falha no tempo t . Ela corresponde à variação do número de falhas ao longo do tempo até a ocorrência do evento. A taxa de

falha é definida como o limite inverso da função de sobrevivência dividido pelo intervalo de tempo, e à medida que esse intervalo tende a zero, torna-se útil para identificar períodos de maior ou menor risco de ocorrência (COLOSIMO; GIOLO, 2021).

A censura pode se manifestar de diferentes formas. Ela ocorre quando não há registro da ocorrência do evento durante o período de análise, sendo esse registro o último no tempo. Essa situação pode surgir quando o tempo de acompanhamento é restrito ou quando ocorrem impedimentos durante a ocorrência do evento. As amostras utilizadas abrangem o fenômeno de estudo, desde casos em que o evento não ocorreu até situações com informações parciais sobre o evento de interesse, incluindo casos em que há perda de acompanhamento ou o evento não se concretiza até o final do estudo (BASTOS; ROCHA, 2006).

Pode-se categorizar a censura como: a censura à direita, que ocorre quando o tempo de acompanhamento de um indivíduo termina antes que o evento de interesse ocorra; a censura à esquerda que trata-se da censura dos dados desde o início do estudo, ou seja, o tempo de ocorrência do evento é desconhecido para alguns indivíduos porque ocorreu antes do início do estudo; e a censura aleatória, que ocorre por diferentes motivos aleatórios, como a remoção de uma amostra durante o estudo, perda de dados ou abandono do estudo, é uma combinação de censura à direita e à esquerda (HABERMAN, 1981).

Ao lidar com dados robustos, há situações em que nem todos os eventos são observados. Nestes casos, é necessário distinguir entre técnicas paramétricas, não paramétricas e semi-paramétricas. As técnicas não paramétricas, como o estimador de Kaplan-Meier (KMF), oferecem flexibilidade sem impor restrições sobre a distribuição do tempo de vida na população da amostra (KAPLAN; MEIER, 1958). Já as semi-paramétricas, como a regressão de Cox, não impõem uma estrutura rígida à distribuição, permitindo a análise de múltiplas variáveis e oferecendo equilíbrio, requerendo o uso de técnicas para avaliar a sua adequação. Por outro lado, as técnicas paramétricas impõem uma forma específica às distribuições de tempo de vida (COLOSIMO; GIOLO, 2021).

Seguindo as definições de técnicas não paramétricas, o estimador *Kaplan-Meier* (KMF) é utilizado para calcular a função de sobrevivência, também conhecido como estimador de produto-limite, considera a sobrevivência em cada intervalo de tempo (KAPLAN; MEIER, 1958). Particularmente, o KMF é aplicado em situações onde a forma da distribuição é incerta ou variável, quando não se conhece a frequência exata associada a cada intervalo, devido à sua flexibilidade na modelagem da distribuição dos tempos de vida. A função de sobrevivência obtida pelo estimador de Kaplan-Meier, representada pela Fórmula 2.2, mostra a proporção de eventos que ocorreram sem falhas em relação ao tamanho da amostra (COLOSIMO; GIOLO, 2021). Essas etapas

de sobrevivência são geradas a partir da ordenação dos tempos de falha, calculando as probabilidades condicionais de sobrevivência em cada intervalo de tempo.

$$S(t_i) = \prod_{t_j < t_i} \left(1 - \frac{d_j}{n_j} \right) \quad (2.2)$$

O método de *Kaplan-Meier* elimina a necessidade de fazer suposições sobre a distribuição uniforme das censuras ao longo do intervalo de observação (KAPLAN; MEIER, 1958). Portanto, é considerado o momento exato da ocorrência do evento, em vez do número fixo de intervalos.

A função de sobrevivência resultante é como uma escada, com uma série de degraus que diminuem ao longo do tempo, representando as estimativas de sobrevivência em intervalos específicos (HABERMAN, 1981). O *Kaplan-Meier* possui como propriedades não ser tendencioso em grandes amostras, a medida que a quantidade de dados disponíveis aumenta a estimativa tende a se aproximar cada vez mais do verdadeiro valor. Quando o tamanho da amostra tende ao infinito, a distribuição do estimador se aproxima de uma distribuição normal e busca calcular a função de sobrevivência que maximiza a probabilidade de observar os eventos na amostra (COLOSIMO; GIOLO, 2021).

Na área educacional, a função de sobrevivência é aplicada na análise do tempo que os alunos levam para concluir um curso, levando em conta fatores como desistências, transferências entre cursos, entre outros (HOED, 2016). Com base nos propósitos deste trabalho, adota-se a abordagem de censura à direita. Isso implica que os estudantes são considerados censurados quando atingem um tempo de observação definido pelo usuário, tendo o tempo de acompanhamento de um estudante terminado antes que o evento de interesse ocorra. No contexto do painel utilizado, o período de observação será definido como vinte semestres para a conclusão do curso.

O presente estudo, alinhado às etapas de análise estatística, utiliza o método de análise de sobrevivência com o estimador não paramétrico Kaplan-Meier como indicador para compreender a dinâmica temporal da trajetória estudantil e, por meio do monitoramento contínuo, identificar padrões de desistência ao longo do tempo. Avalia-se também o impacto de variáveis acadêmicas na probabilidade de evasão, além de determinar os principais pontos de inflexão onde os alunos são mais propensos a abandonar seus estudos.

Com uma abordagem de monitoramento de indicadores, as instituições de ensino podem implementar estratégias direcionadas, como programas de apoio acadêmico, orientação personalizada e políticas de permanência, visando mitigar os fatores de risco e promover a conclusão bem-sucedida dos estudos pelos alunos. Em suma, a análise de sobrevivência no contexto da educação oferece uma ferramenta para iden-

tificar, monitorar e entender a evasão, baseando-se no comportamento das curvas de sobrevivência para embasar decisões direcionadas à melhoria do sucesso dos alunos.

2.0.3 Fundamentos do Business Intelligence

O Business Intelligence (BI) é uma metodologia que envolve a coleta, organização e análise de dados para oferecer informações úteis aos gestores e tomadores de decisão. Seu propósito é viabilizar o acesso interativo aos dados, permitir a manipulação e servir como ferramenta para os profissionais realizarem análises precisas (TURBAN et al., 2009).

A transformação de grandes volumes de dados convertidos em informações pode beneficiar atividades de uma organização. Ambientes de integração, que possibilitam a consolidação e o uso estratégico dos dados, servem para armazenar e processar essas informações por meio de ferramentas especializadas (ALMEIDA; CAMARGO, 2015).

Na arquitetura dos componentes do BI estão relacionados a economia de tempo, o aprimoramento de estratégias e planos, tomada de decisões mais assertivas, redução de custos e processos mais eficientes. As áreas mais comuns de aplicação do BI incluem relatórios gerais, análise de vendas e marketing, planejamento e previsão, consolidação financeira, relatórios regulamentares, orçamento e análise de rentabilidade (BOTELHO; FILHO, 2014).

Os principais componentes da arquitetura de BI guiam o desenvolvimento de aplicações. As fontes de informação são os locais de origem dos dados, como bancos de dados operacionais, sistemas transacionais, planilhas e arquivos de texto. O processo de extração, transformação e carga (ETL) consiste na extração de dados das fontes, sua transformação para análise e carga no *Data Warehouse (DW)*. Este, por sua vez, é um repositório centralizado de dados históricos e consolidados, projetado para suportar análises e consultas de negócios. Os dados são organizados para facilitar análises multidimensionais e geração de relatórios (CANO, 2007).

Os *Data Marts (DM)* são subconjuntos do *Data Warehouse*, contendo dados específicos de áreas ou departamentos para análises direcionadas. Por fim, a área de apresentação é a interface onde os usuários interagem com os dados para tomada de decisões. Isso inclui ferramentas como relatórios, *dashboards* e *Online Analytical Processing (OLAP)*, que tornam os dados acessíveis e compreensíveis aos usuários finais (TURBAN et al., 2009).

2.0.4 Data Science

Data Science ou Ciência de dados utiliza princípios, metodologias e orientações direcionadas às etapas de transformação, validação e análise de dados para inferir conhecimento em conjuntos de dados (FILATRO, 2020). Esta abordagem científica é constituída em torno de hipóteses que podem ser testadas, partindo de um problema para chegar à compreensão ou solução. O foco no aprimoramento do processo decisório envolve o fenômeno em estudo por meio de análise de dados, combinando indivíduos com habilidades analíticas, processos bem estruturados, etapas até a visualização dos dados, com plataformas para obtenção de relatórios e capacidade de um processo escalonável (PAIXÃO; SILVA; TANAKA, 2015).

Visto que se trata de um campo de estudo que se estende por diferentes disciplinas, são combinadas técnicas como estatística, *Machine Learning*, matemática e *Data Mining* para processar um grande volume de dados e extrair conhecimento, tendo como etapas a definição do problema, coleta dos dados, preparação dos dados, exploração dos dados, modelagem dos dados, comunicação dos dados e automatização das análises.

São encontradas nos campos de estudo da ciência de dados metodologias que se adaptam a diferentes setores e necessidades. No âmbito da educação, destacam-se áreas como *Learning Analytics*, *Educational Data Mining (EDM)* e *Academic Analytics*. Estas subdisciplinas aplicam os princípios e técnicas para lidar com informações e analisar dados educacionais que podem apresentar relações entre si (RODRIGUES et al., 2014), como abordado na Figura 4.

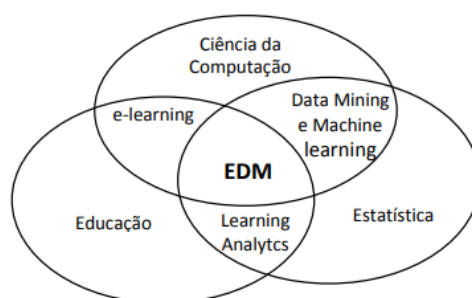


Figura 4 – Principais áreas relacionadas com a EDM.

Fonte: Rodrigues et al. (2014).

2.0.4.1 Learning Analytics (LA) e Educational Data Mining (EDM)

Para compreender o progresso dos alunos ao longo do curso, disciplina ou programa, há a medição, coleta, análise e reporte dos dados. Nesse contexto, o *Learning Analytics* e o *Educational Data Mining (EDM)* destacam-se como componentes da ciência de dados aplicada à educação. Empregando técnicas para identificar padrões de aprendizado, áreas de melhoria e até mesmo antecipar possíveis desafios. Enquanto

o LA foca na análise do comportamento dos alunos e na otimização do processo de ensino (PALOMINO et al., 2022), o EDM visa extrair informações a partir dos dados educacionais para melhorar os processos pedagógicos (CARMINATI et al., 2020).

Direcionados à compreensão embasada em dados, o LA e o EDM proporcionam análises ao longo do tempo, buscando compreender o passado. Detalhes como relatórios e descrição dos dados com análises abrangentes do que ocorreu anteriormente. E assim, o entendimento e orientação, explorando o motivo de determinados eventos ocorrerem, utilizando modelos e explicações para analisar o presente, fundamentando a melhor ação a ser tomada no futuro e antecipando o que é provável que aconteça, seja por meio de previsões, simulações de caminhos alternativos de ação ou identificação do caminho ideal a ser seguido (FILATRO, 2020).

O *Learning Analytics* contribui para a tomada de decisões informadas por parte dos educadores e gestores educacionais por meio da evolução do desempenho acadêmico, proporcionando uma base sólida para aprimorar estratégias pedagógicas e garantir um ambiente educacional mais eficiente e personalizado, promovendo o sucesso acadêmico nas instituições educacionais (MARQUES et al., 2023).

As etapas do ciclo de vida do LA descritas na Tabela 1 referem-se ao processo de análise sistemática do desempenho e progresso dos alunos ao longo do tempo, visando à melhoria do ensino e aprendizagem (CLOW, 2012).

Tabela 1 – Ciclo do Learning Analytics

Processo	Descrição
Learners	Coleta de dados de múltiplas fontes, como desempenho acadêmico e indicadores relevantes para a análise do processo de ensino e aprendizagem.
Data	Análise e apresentação dos dados coletados, fornecendo <i>insights</i> sobre o desempenho no processo de aprendizagem.
Metrics/Analytics	Identificação de potenciais problemas de aprendizagem com base nos <i>insights</i> gerados na etapa de dados, visando intervenções preventivas para melhorar o desempenho acadêmico.
Intervention	Implementação das intervenções identificadas para fornecer recursos personalizados no ambiente acadêmico.

Da mesma forma, o processo do *Educational Data Mining (EDM)* permite a utilização dos resultados por *stakeholders*, como desenvolvedores de software educacional, professores, gestores e pesquisadores na área da educação para o aprimoramento nos processos de ensino-aprendizagem, assim como nos mecanismos de gestão aca-

dêmica e pedagógica das instituições de ensino, promovendo, assim, uma abordagem mais eficaz e embasada em dados no campo educacional (RODRIGUES et al., 2014). O fluxo da Figura 5 representa este processo.

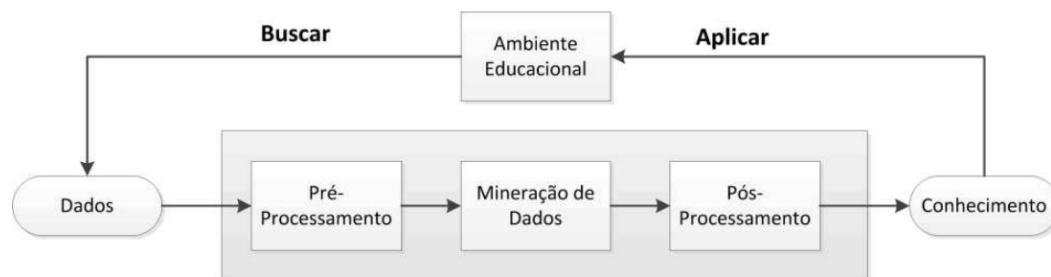


Figura 5 – Etapas de uma EDM.

Fonte: Rodrigues et al. (2014).

No processo de *EDM* ocorre o pré-processamento de dados, que engloba a limpeza, integração e transformação dos dados para prepará-los para análise. Em seguida, avança-se para a análise exploratória de dados, onde são identificadas tendências e anomalias. Posteriormente, na etapa de modelagem de dados, são desenvolvidos e aplicados modelos estatísticos e algoritmos, e realizada a avaliação do modelo para garantir a precisão. Por fim, os resultados são interpretados e visualizados. Ambos os processos estruturam etapas para otimização analítica no âmbito educacional, abordando questões como o tipo de dados coletados, a razão para sua análise, a metodologia de análise e os resultados obtidos.

2.0.4.2 Academic Analytics (AA)

O objetivo principal do *Academic Analytics* é utilizar dados educacionais para melhorar o desempenho dos alunos, otimizar o processo de ensino e orientar decisões estratégicas nas instituições de ensino, empregando análises estatísticas para identificar padrões e tendências. Isso não apenas permite que educadores adaptem seus métodos de ensino para atender às necessidades específicas dos alunos, mas também que gestores e coordenadores desenvolvam estratégias eficazes, revisem políticas institucionais e aprimorem a qualidade global do ambiente educacional (PAZ; CAZELLA, 2021).

Para utilizar este processo, é necessário conhecer as etapas, que incluem desde a coleta até a interpretação de dados provenientes de múltiplas fontes (ANDRADE; FERREIRA, 2016). A Figura 6 apresenta visualmente as cinco etapas centrais, desde o acesso aos dados até a exploração na gestão da IES.

A etapa inicial do processo envolve a obtenção dos dados educacionais necessários para análise, que podem ser provenientes de diversas fontes, como sistemas de gestão acadêmica e bancos de dados. Em seguida, na etapa de transformação, os

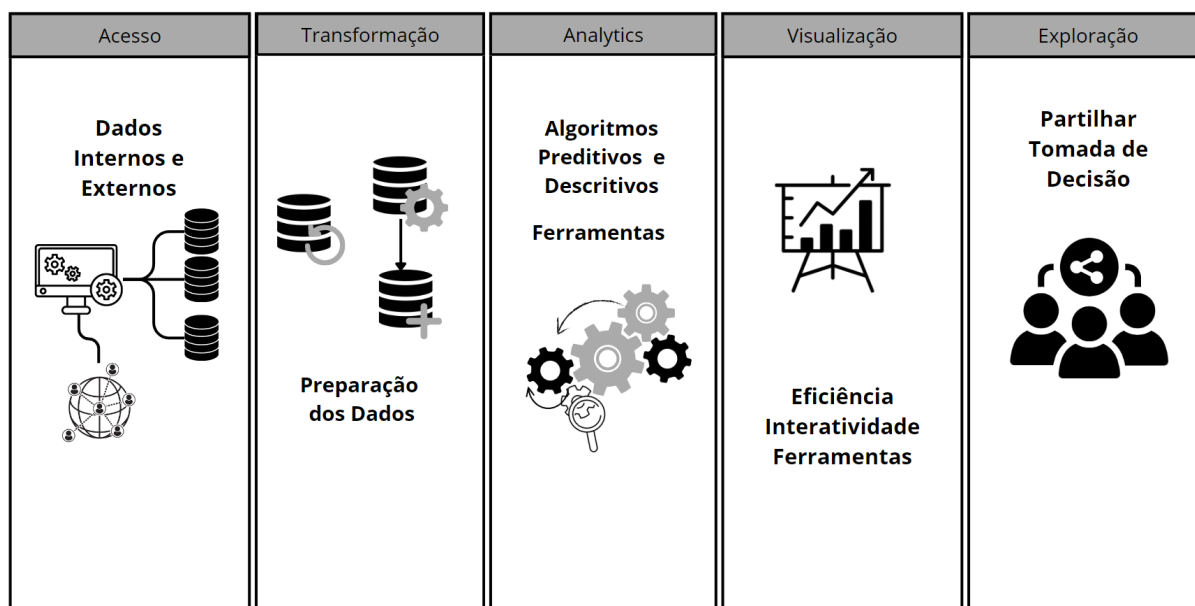


Figura 6 – Etapas de desenvolvimento de um Academic Analytics.
Fonte: Adaptado de [Andrade e Ferreira \(2016\)](#).

dados brutos são processados e preparados para análise, através da identificação e correção de inconsistências, erros e valores duplicados, além da organização e formatação adequada para facilitar interpretações precisas.

Posteriormente, na etapa de *Analytics*, os dados preparados são submetidos à análise estatística utilizando algoritmos preditivos ou descritivos, como análise de sobrevivência e regressão, com o objetivo de identificar padrões e correlações para compreender melhor o desempenho acadêmico. Os resultados são então interpretados e apresentados visualmente, utilizando gráficos e *dashboards*, para comunicar as revelações extraídas e facilitar a compreensão dos dados. Na etapa de exploração, os resultados são disponibilizados para a elaboração de planos estratégicos e ações direcionadas às necessidades da instituição de ensino superior, contribuindo para a organização, aprimoramento do conhecimento e tomada de decisões mais precisas baseadas em informações fundamentadas em dados ([ANDRADE; FERREIRA, 2016](#)).

Os resultados são disponibilizados para a elaboração de planos estratégicos e ações direcionadas às necessidades da IES. Esse processo pode contribuir para aprimorar o fluxo de conhecimento, promover diálogo, possibilitar comparações com outras instituições, proporcionar uma percepção de sucesso na trajetória acadêmica do discente, reduzir custos e permitir um destino mais eficaz dos recursos financeiros, baseado em informações fundamentadas em dados. Dessa forma, busca-se tomar decisões mais precisas dentro da instituição.

2.0.5 SABIA

O *Academic Business Intelligence and Analytics (SABIA)*¹ é uma plataforma utilizada por gestores e coordenadores para acompanhar informações educacionais e determinar questões como cálculo de vagas a serem disponibilizadas, abrangendo indicadores de desempenho da instituição, cursos, disciplinas e estudantes. Estruturada com base nos processos empregados na construção de sistemas de *Business Intelligence*, a plataforma integra conceitos de *Learning Analytics* e *Academic Analytics*, com o propósito de oferecer suporte à gestão fundamentada em evidências nas Instituições de Ensino Superior (IES) (MARQUES et al., 2023).

Em funcionamento na Universidade Federal Rural de Pernambuco, essa plataforma conta com recursos que exibem dados gerais e funcionalidades mais específicas. Utilizando como fonte de informações na plataforma os dados do INEP, junto a dados internos da universidade, que passam por processos de validação e segurança, para garantir a proteção de informações sensíveis. Os usuários acessam essas funcionalidades por meio de login, com opções de filtragem disponíveis em painéis dinâmicos. Esses painéis utilizam gráficos, tabelas, mapas de calor e *cards* para converter dados em informações que podem ser compartilhadas e interpretadas.

¹ <https://ufrpe-odg.github.io/site/sabia.html>

3 Trabalhos Relacionados

Esta seção descreve trabalhos relevantes que abordam o fenômeno da evasão no ensino superior. Esses estudos aplicam diversas metodologias no âmbito educacional, contribuindo para a investigação dos fatores de influência na evasão estudantil. Ao analisar a literatura existente, é possível identificar reflexões acerca do tema que auxiliam na formulação de políticas e práticas mais eficazes, com o intuito de mitigar o problema da evasão. Alguns desses estudos destacam a importância da utilização de técnicas de análise de sobrevivência e *Business Intelligence (BI)* como ferramentas relevantes na gestão acadêmica.

3.0.1 Análise de sobrevivência

O estudo de (HOED, 2016) aborda a evasão nos cursos de graduação da área de computação, utilizando dados do INEP e da Universidade de Brasília (UnB). Por meio de análises quantitativas, são calculadas as taxas anuais de evasão e aplicadas técnicas estatísticas, como análise de sobrevivência e mineração de regras de associação utilizando o algoritmo apriori. Além disso, é conduzida uma análise qualitativa por meio de questionários aplicados a alunos evadidos, com o objetivo de identificar suas causas. Os resultados evidenciam diferenças entre instituições públicas e privadas e destacam a relação entre a evasão e áreas de conhecimento específicas, como Ciências, Matemática e Computação. Um estudo de caso é realizado, analisando os cursos de Bacharelado em Ciência da Computação, Licenciatura em Computação, Engenharia de Software e Engenharia de Computação, revelando como a relação candidatos por vagas e variáveis como sexo, forma de ingresso e cotas impactam na evasão. Neste trabalho, serão utilizados os semestres e modalidades de ensino para verificar como impactam na permanência estudantil.

No mesmo contexto, a pesquisa de (SACCARO; FRANÇA; JACINTO, 2019) concentrou-se em variáveis relacionadas à evasão de estudantes no ensino superior nas áreas das ciências naturais e engenharias. Utilizando técnicas de análise de sobrevivência e dados do INEP, a pesquisa revelou que a evasão é mais alta em instituições privadas. Constatou-se que ser homem e ter uma idade mais avançada reduzem o tempo de permanência no ensino superior, enquanto estudantes que recebem apoio financeiro apresentam probabilidades mais altas de permanência no curso. Campos (2016) também utilizou análise de sobrevivência em sua pesquisa e por meio da coleta de dados institucionais, considerou as variáveis sexo, idade ao ingressar, forma de ingresso no curso, situação do aluno no curso e área de conhecimento, para o acompa-

nhamento do tempo de permanência na análise de alunos evadidos do curso, obtendo como resultados diferenças entre percentuais de evasão de acordo com a escolha da variável no estudo. Enquanto os estudos anteriores abordaram a evasão de forma mais abrangente, analisando diversas áreas, este trabalho se concentrará exclusivamente na análise das áreas de computação em uma universidade de ensino superior. Utilizará as etapas de ETL no tratamento dos dados, incorporando técnicas de *Learning Analytics* e *Academic Analytics* direcionadas ao compartilhamento dos resultados para auxiliar decisões e promover o sucesso dos estudantes.

Trabalhos como o de (SILVA, 2013) e (SOUZA; KOMATI; ANDRADE, 2022) também investigaram os fatores determinantes no abandono estudantil. Enquanto o estudo de (SILVA, 2013) trata de uma análise de dados de uma instituição privada para identificar correlações entre evasão e demanda por cursos, destacando a falta de programas institucionais profissionalizados no combate à evasão, o trabalho de (SOUZA; KOMATI; ANDRADE, 2022) se concentra em um estudo de caso específico em um curso de graduação em Sistemas de Informação. Ambos os estudos reconhecem a importância dos primeiros semestres como períodos críticos para a evasão, destacando que é nesse momento que os alunos enfrentam maior risco de desistência e apontam a relevância de fatores como questões acadêmicas, expectativas dos alunos e integração com a instituição como determinantes para explicar os percentuais de evasão. Souza, Komati e Andrade (2022) ainda identifica características que ressaltam as disciplinas dos dois primeiros semestres com aproximadamente 50% de retenção da população estudantil. Neste trabalho, será realizada a modelagem do progresso acadêmico de acordo com o vínculo do discente ao longo dos semestres.

Assim como os demais trabalhos, (SANTOS et al., 2022) aborda a análise de evasão, utilizando análise de sobrevivência. Como diferencial, também utiliza Cadeias de Markov Absorventes e analisa o impacto da retenção de estudantes no ensino superior. O estudo, realizado em cursos das áreas de agrárias, computação e saúde em uma universidade pública brasileira, concentrou-se na análise do impacto da retenção de estudantes. A metodologia empregou a construção de um modelo estocástico de Cadeia de Markov, juntamente com a aplicação da análise de sobrevivência para avaliar o tempo e a probabilidade de eventos como evasão, conclusão e desvinculação dos estudantes. Resultando em diferenças no comportamento de evasão entre os cursos e evidenciando a relevância do semestre em que os estudantes foram retidos. Neste trabalho, será analisada a permanência estudantil, considerando o vínculo do discente até a evasão, comparando probabilidades de sucesso entre cursos da área de computação, para grupos relacionados ao ingresso do aluno na IES.

3.0.2 Business Intelligence

Decisões baseadas em fatos e informações oportunas e em tempo hábil fornecem às empresas agilidade e respaldo na tomada de decisões mais assertivas. O uso de aplicativos, ferramentas e metodologias para a coleta, tratamento, armazenamento, recuperação e disseminação de informações visa auxiliar o processo decisório em organizações complexas. Segundo [Botelho e Filho \(2014\)](#), é possível adotar diferentes abordagens de acordo com as situações de aplicação e os diversos níveis hierárquicos, ampliando assim a adaptabilidade do processo decisório em diferentes contextos organizacionais.

Na educação, o *Educational Data Mining (EDM)* ou Mineração de Dados Educacionais (MDE) se tem como objetivo transformar os dados brutos provenientes de sistemas educacionais em informações práticas e úteis. Neste contexto, [\(MANHÃES et al., 2011\)](#) e [\(CARMINATI et al., 2020\)](#) exploram o uso de algoritmos de mineração de dados para identificar o risco de evasão de alunos em instituições de ensino superior. Enquanto [\(MANHÃES et al., 2011\)](#) destacam a importância da aplicação de técnicas de mineração de dados em conjunto com ferramentas de análise de dados para identificar riscos nos cursos de graduação, por meio de experimentos para comparar o desempenho dos algoritmos, [\(CARMINATI et al., 2020\)](#) concentram-se na análise da base de dados financeira da instituição acadêmica para obter indicadores de evasão. Há ainda o trabalho de [\(RODRIGUES et al., 2014\)](#) realizam uma revisão literária das diferentes abordagens na área de EDM, abrangendo técnicas, modelos, objetivos educacionais e ferramentas utilizadas. Os estudos ressaltam a relevância dos algoritmos voltados ao processamento dos dados. Além disso, [Manhães et al. \(2011\)](#) ainda explicitam a importância do uso de ferramentas de Business Intelligence para identificar e analisar esses padrões.

Explorando o cenário das plataformas que incorporam conceitos de *Business Intelligence*, tanto o estudo de [\(PALOMINO et al., 2022\)](#) quanto o de [\(MARQUES et al., 2023\)](#) abordam a importância do uso de ferramentas de visualização voltadas à análise de dados educacionais para o alcance de uma gestão baseada em evidências. Enquanto [\(MARQUES et al., 2023\)](#) apresenta uma ferramenta, intitulada *SABIA*, que agrega conceitos de *Learning Analytics (LA)* e *Academic Analytics (AA)* para dar suporte à gestão em IES, fornecendo painéis com visões e detalhamentos sobre os cursos e disciplinas da instituição acadêmica, além de indicadores-chave como Índice Geral de Cursos (IGC) e Taxa de Sucesso de Disciplina (TSD), [\(PALOMINO et al., 2022\)](#) realiza um estudo de plataformas atuais que fornecem visualização de dados educacionais, focando nas metas do Plano Nacional de Educação (PNE), para o desenvolvimento de futuras soluções. A partir das ferramentas apresentadas, os trabalhos destacam que a utilização do BI auxilia a gestão acadêmica.

Andrade e Ferreira (2016) discorre sobre questões relevantes, reflexões, aplicações e protótipos de sistemas utilizando o *Academic Analytics*. Com a leitura de bases de dados, registros de diferentes fontes são integrados para gerar relatórios e análises. Essa abordagem utiliza dados para apoiar a gestão da Instituição de Ensino Superior (IES), fornecendo à administração indicadores históricos sobre vários aspectos da organização. Assim como (PAZ; CAZELLA, 2021) que aplica um modelo de Analítica Acadêmica em uma Instituição de Ensino Superior, junto a técnicas de BI para otimização do tempo e qualidade nas informações, resultando em uma boa avaliação por parte dos coordenadores em relação ao apoio à tomada de decisão e auxílio do processo de gestão.

Além do tema do BI, o estudo de (ALMEIDA; CAMARGO, 2015) aborda o monitoramento e controle da evasão escolar no ensino superior na obtenção de indicadores relevantes. Destacando o uso de um sistema de monitoramento contínuo de dados para identificar e compreender problemas relacionados à evasão. A aplicação de ferramentas analíticas e técnicas de BI permite que as instituições de ensino superior visualizem as informações armazenadas, visando otimizar o uso dos recursos da instituição e fornecer informações direcionadas aos gestores educacionais.

Diante do exposto, é notório que diversos autores têm realizado pesquisas direcionadas à educação no ensino superior, desses estudos, várias tendências e aspectos fundamentais emergem: fatores determinantes para a evasão, variações nas taxas entre instituições públicas e privadas, desafios na implementação de programas de combate à evasão e o papel das ferramentas analíticas na gestão educacional. Fomentar a análise de dados em diversas instituições acadêmicas possibilita a descoberta de informações relevantes em cada uma delas, permitindo comparações com dados em diferentes escalas de análise. Neste estudo, as etapas de tratamento de dados acadêmicos, que envolvem a exploração, transformação e carga dos dados, atuam na viabilidade de análises estatísticas em uma plataforma de BI. O grande diferencial dessa abordagem reside na integração de um painel direcionado à permanência estudantil, que possibilita a personalização das análises e o compartilhamento contínuo de resultados.

Neste cenário, o presente trabalho fornecerá um painel web interativo que permite a comparação entre grupos, como tipo de ingresso, curso, campus acadêmico, semestre de ingresso, turno e área de estudo, conforme a situação de vínculo do discente. Utilizando a plataforma *System of Academic Intelligence and Analytics (SABIA)*, o objetivo é apoiar instituições de ensino superior a tomar decisões mais informadas (MARQUES et al., 2023). Essa abordagem de monitoramento de indicadores vai além da modelagem do tempo de permanência do estudante na instituição, verificando diferentes comportamentos associados aos grupos. Isso inclui a capacidade de detectar

períodos específicos onde o comportamento dos alunos difere e grupos de alunos que enfrentam maior risco de evasão. Compreender esses padrões capacita as instituições a reconhecer e tomar medidas em períodos críticos, como oferecer suporte adicional aos alunos ou ajustar a oferta de disciplinas, visando o sucesso acadêmico dos estudantes e otimizando os recursos da instituição.

4 Materiais e Métodos

A partir da revisão bibliográfica e de trabalhos relacionados, adquiriu-se conhecimento para planejar e executar o método e a pesquisa adotados por este trabalho. No desenvolvimento, foram empregadas técnicas de análise estatística de sobrevivência e *Business Intelligence*. A implementação seguiu etapas que incluíram seleção, processamento, transformação e interpretação dos dados. Desta forma, o desenvolvimento seguiu a estrutura de:

- Definição da base de dados utilizada;
- Definição da ferramenta para criação da *dashboard*;
- Formatação dos dados, processamento e transformação dos dados;
- Implementação do algoritmo de sobrevivência;
- Visualização e interpretação dos dados.

4.0.1 Definição e formatação dos dados

A principal fonte de dados utilizada nesta pesquisa foi o *INEP*. Os dados referem-se à página de microdados, reunindo um conjunto de informações detalhadas relacionadas a pesquisas, exames e avaliações educacionais, referentes ao período de 2010 a 2023.

A junção dos dados oriundos do *INEP* e informações internas à instituição foi utilizada na criação de *Data Marts*, exclusivos destinados a cada painel do *SABIA*, abrangendo informações sobre uma instituição pública federal de ensino superior. Esses *Data Marts* foram desenvolvidos seguindo os princípios do *pipeline de dados ETL*, permitindo um fluxo automatizado de diferentes tipos de dados da origem ao destino, visando um processo de pipeline de dados tolerante a falhas, conforme descrito por (RAJ et al., 2020).

Tendo em conta que os dados pertinentes a IES não mudam com alta frequência ou imediatamente, o processamento *real-time* do *pipeline de dados* não é trivial para os propósitos deste trabalho. Atualizações e análises dos dados são realizadas a cada semestre. A Figura 7 representa as etapas da elaboração da plataforma de BI.

As etapas para elaboração da plataforma BI, utilizaram a linguagem Python¹ e bibliotecas direcionadas à manipulação de grandes registros de dados, como Pandas²,

¹ <https://www.python.org/>

² <https://pandas.pydata.org/>

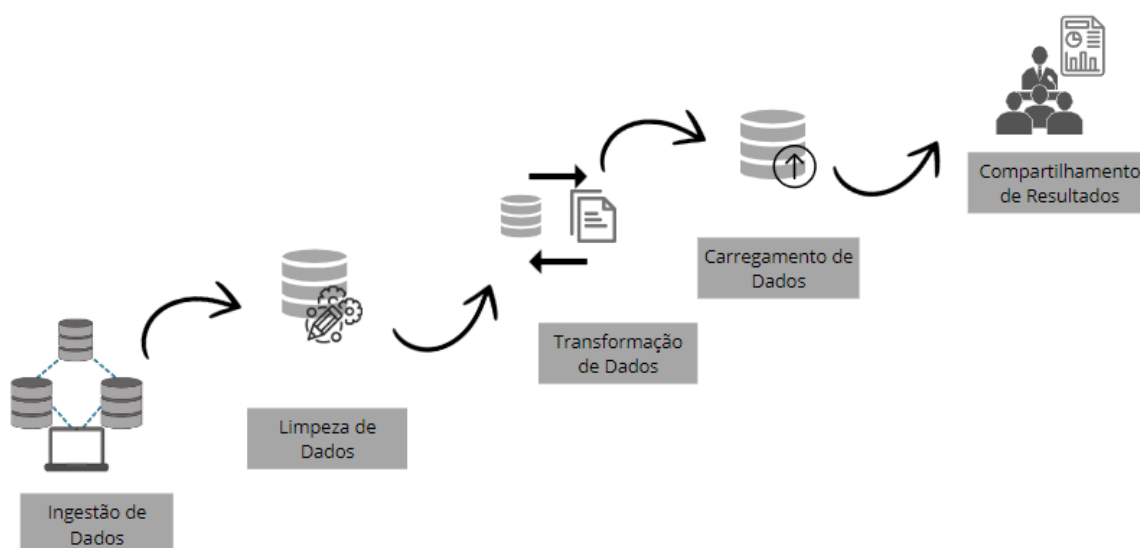


Figura 7 – Etapas da elaboração da plataforma de BI.
Fonte: Adaptado de (MARQUES et al., 2023).

que serviu na manipulação e análise de dados, enquanto o Numpy³ proporcionou suporte em operações matemáticas. A ingestão dos dados envolveu a extração e coleta das informações, provenientes de múltiplas fontes, incluindo acessos internos e externos do ambiente de trabalho, advindos de planilhas, sistemas de gestão, resultados do ENADE, entre outros.

Na fase de limpeza de dados, foram removidas inconsistências, redundâncias e dados duplicados, com o tratamento de erros, para garantir qualidade e uniformidade das informações. Na etapa de transformação para construção do painel de análise de permanência, os dados foram preparados, adaptados de diferentes fontes. Foram aplicadas técnicas como agregações, filtragens, criação de novos campos com base em outros, ajustes nos nomes ou tipos dos campos e tratamento dos valores. Em seguida, na etapa de carregamento, os dados são integrados em um único *Data Mart*, nomeado como *vinculo_periodo.parquet*, com extensão do arquivo no formato *Parquet* para otimizar o tamanho do arquivo e a eficiência de manipulação.

4.0.2 Modelo de análise de sobrevivência

O método empregado neste trabalho visa analisar a probabilidade de permanência dos estudantes ao longo da duração de seus vínculos. Para isso, desenvolveu-se um modelo de análise de sobrevivência com base nos dados de uma universidade pública federal.

O modelo proposto segue um processo de análise estatística, visando consoli-

³ <https://numpy.org/pt/>

dar informações sobre a permanência dos alunos no ensino superior. Isso implica na observação dos padrões de vínculo, visando detectar e monitorar características comuns e discrepâncias ao analisar as variáveis de interesse. O modelo foi elaborado por meio da linguagem de programação Python e implementado em um painel na plataforma de *Business Intelligence* intitulada SABIA.

O conjunto de dados criado nas etapas de extração, transformação e carregamento (*ETL*) é carregado no modelo gerado. As funções e métodos que estruturam o modelo estão presentes na Tabela 2. Inicialmente, os dados referentes ao vínculo e período, coletados e transformados em *Data Mart*, são carregados na plataforma utilizando a classe *DataLoader*. No *build_body*, uma cópia do *dataframe* original é feita para garantir a preservação dos dados. Em seguida, cria-se a coluna de modalidade de ingresso, derivada do tipo de admissão, segmentando as modalidades regulares de ingresso na graduação de outros tipos de admissão e as funções para o cálculo e exibição das estimativas de sobrevivência são chamadas. A página é iniciada com as opções de filtros gerais para seleção de: curso, campus, semestre de ingresso, tipo de ingresso, turno, ano de início e ano de fim.

Tabela 2 – Funções do odelo de análise de sobrevivência

Função	Definição
<code>DataLoader</code>	Carrega os Data Marts.
<code>build_body</code>	Monta o painel, exhibe os filtros e resultados das estimativas de sobrevivência em forma de gráfico, analisando o tempo até a ocorrência de eventos em que um estudante permanece no sistema durante o período.
<code>sensor_subjects</code>	Determina se uma observação deve ser considerada censurada ou não.
<code>create_controls</code>	Gera controles interativos que permitem ao usuário filtrar e personalizar os dados exibidos.
<code>prepare_survival_df</code>	Adequa o DataFrame para análise de sobrevivência, aplicando filtros selecionados pelo usuário e preparando os dados.
<code>__prepare_df_survival</code>	Calcula as probabilidades de sobrevivência para cada grupo e retorna o DataFrame preparado para análise de sobrevivência.
<code>plot_survival</code>	Plota o gráfico de probabilidades de sobrevivência.

Na construção do modelo, a função *create_controls* é responsável por gerar os controles que permitem aos usuários filtrar e personalizar os dados exibidos. Esses controles adicionam componentes visuais na tela e os organizam para facilitar a interação e personalização do modelo. Em seguida, definem-se o agrupamento no modelo, que se refere a variável pela qual os dados serão agrupados para análise e são exibidos em menus suspensos *selectbox*. Estruturam-se as opções de seleção dos eventos de morte utilizando um menu de seleção múltipla *multiselect*, onde o usuário pode

escolher entre *EVADIDO*, *FORMADO* ou ambas as opções. Para completar as funcionalidades de filtro e seleção associadas ao modelo de análise de sobrevivência, um controle deslizante *slider* condiciona o valor definido com base na duração máxima de vínculo, ajustável conforme necessário, para estabelecer um limite temporal da duração analisada.

Os parâmetros da função são definidos para os cálculos das curvas de sobrevivência. Neste modelo, a escala de tempo utilizada é a duração de vínculo, indicando o tempo decorrido desde o início do vínculo do estudante até o momento da análise, medida em períodos. O tempo de início é o momento em que cada estudante inicia seu vínculo com a instituição de ensino e é representado pelo valor inicial na coluna *duracao_vinculo*. O evento ou falha de interesse corresponde à saída do sistema educacional, definida com base nas situações de morte fornecidas pelo usuário, tais como *EVADIDO*, *FORMADO* ou ambos. O evento é identificado pela coluna *evento*, onde o valor 1 representa o evento de falha (saída do sistema) e o valor 0 indica que o estudante ainda está no sistema. A censura ocorre quando o evento de interesse (saída do sistema) não é observado para um determinado estudante até o final do período de observação. Neste trabalho, verifica-se se a duração do vínculo do estudante é menor ou igual à duração máxima especificada para análise. Se a condição for verdadeira, o estudante é considerado censurado.

Durante a preparação dos dados, a partir da função *prepare_survival_df*, adequa-se o DataFrame para análise de sobrevivência com base nos filtros selecionados pelo usuário, de agrupamento, evento de morte e duração máxima do vínculo presentes na função *create_controls*. A base de dados é filtrada para garantir apenas observações de duração de vínculo válidas, levando em consideração apenas as últimas entradas de cada estudante no sistema e, assim, evitando estados de vínculo anteriores ou futuros.

A função *sensor_subjects* é então chamada para determinar se uma observação específica deve ser considerada censurada ou não, utilizando o período máximo definido pelo usuário no *slider*, que por padrão na plataforma inicia limitando a vinte períodos. Primeiramente, é avaliado se a duração do vínculo do estudante é menor ou igual à duração máxima de vínculo especificada. Se a duração do vínculo for menor ou igual ao valor máximo, a observação é marcada como censurada nesse critério. Outra abordagem é verificar se o último vínculo do período do estudante não está entre os eventos de término de interesse, *FORMADO*, como estudante que concluiu o curso com êxito ou *EVADIDO*, o estudante que desistiu do curso, o que configura o abandono e insucesso do aluno. Neste trabalho, investiga-se o vínculo dos discentes matriculados, categorizados como *CURSANDO* no momento inicial da análise de permanência.

Para efeitos deste trabalho, considera-se como evadido o aluno que estava matriculado em um semestre letivo específico, mas não renovou a matrícula no semestre seguinte. Os alunos que trancaram a matrícula são classificados como evadidos, enquanto aqueles que trancaram a matrícula mas retornaram aos estudos foram considerados em curso. Os estudantes que concluíram a graduação com êxito são identificados como formados. Optou-se por um modelo não paramétrico para evitar a influência de variáveis desconhecidas no fenômeno de interesse, utilizando a duração do vínculo como escala de tempo, que indica o tempo decorrido desde o início do vínculo do estudante até o momento da análise (medida em períodos).

Após as operações de preparação, o *DataFrame* resultante é ajustado para conter apenas as colunas relevantes (duração do vínculo, evento, censurado, grupo e identificação dos discentes). A censura é aplicada com base nas colunas de duração do vínculo e identificação dos discentes, utilizando o período máximo definido pelo usuário. Por fim, o método privado `__prepare_df_survival` é chamado para calcular as probabilidades de sobrevivência para cada grupo e retornar o *DataFrame* preparado para análise de sobrevivência com informações sobre o tempo decorrido e as probabilidades de sobrevivência.

Os cálculos resultantes da análise de sobrevivência fornecem informações sobre as probabilidades de permanência ao longo do tempo para cada grupo. Essas informações são então agregadas e armazenadas em um *DataFrame* auxiliar. Durante essa operação, são registrados todos os estudantes que experimentaram o evento de interesse ao longo da duração do vínculo (medida em períodos), até o limite máximo estabelecido pelo usuário para as observações. Os registros incluem:

1. Todos os estudantes que experimentaram o evento de interesse durante o período de observação, seja evasão, formação ou ambas.
2. Todos os estudantes que ainda estão no sistema educacional, ou seja, aqueles que permanecem vivos por mais tempo do que o período de observação.
3. Os estudantes que permaneceram no sistema até o final do período de observação, sendo considerados censurados à direita.

Para o terceiro caso, o evento de interesse não ocorreu dentro do período de observação. Neste modelo, encerram-se as observações quando um aluno atinge o limite de períodos selecionado pelo usuário, caracterizando-se como censura à direita. Considera-se *vivo* o estudante que permaneceu até o final do período de observação e *morto* o estudante que experimentou o evento durante o período observado, como a evasão ou a formação de um aluno, correspondendo a finalização do vínculo.

Para implementar o método estatístico *Kaplan-Meier* e obter a função de sobrevivência em relação ao tempo, utilizou-se a biblioteca *Lifelines*⁴, conforme representado na Tabela 3. O modelo *Kaplan-Meier* ajusta os dados utilizando os tempos de sobrevivência e a ocorrência de eventos. Neste trabalho, os parâmetros para a função de sobrevivência são definidos como a duração do vínculo, representada por T , e o estado do vínculo, que fornece informações sobre o estado de cada período. Uma coluna de evento é criada para registrar a última ocorrência associada ao estado de falha ou censura. O evento, denotado por e , resulta do filtro do último vínculo relacionado à evasão ou formação. Com a definição da duração e do evento, é chamada a função `__prepare_df_survival` para calcular as probabilidades de sobrevivência.

Tabela 3 – Kaplan meier lifelines

```
from lifelines import KaplanMeierFitter

kmf = KaplanMeierFitter()

kmf.fit(T, event_observed=E)
```

A escolha desse método ocorreu devido à sua capacidade de lidar com dados de censura, sendo comum em estudos educacionais nos quais nem todos os alunos são acompanhados até a conclusão do curso. Por ser um método não paramétrico, torna-se possível aplicá-lo em diferentes cenários, sem a necessidade de assumir uma distribuição específica para os dados, uma vez que a permanência dos alunos pode ser influenciada por uma variedade de fatores complexos que podem não seguir uma distribuição pré-definida. A biblioteca *Lifelines* foi escolhida devido à sua eficiência e facilidade de uso na implementação do método de *Kaplan-Meier*, proporcionando uma solução robusta para a análise de sobrevivência.

O cálculo é conduzido individualmente para cada grupo presente no agrupamento selecionado. Um exemplo com dados fictícios simula o *input* e *output* dos dados, tendo como entrada os dados representados na Figura 8 e após a preparação em `prepare_survival_df`, é chamado o método privado `__prepare_df_survival` para o cálculo das probabilidades de sobrevivência, resultando na geração de um *DataFrame* com as probabilidades de permanência de cada grupo, representado na Figura 9, tomando como exemplo o agrupamento turno. Essa abordagem proporciona uma visão detalhada e comparativa das trajetórias de sobrevivência. A probabilidade é calculada como o produto das probabilidades em intervalos de tempo sucessivos. Quanto maior a proporção de sobreviventes, maior é a probabilidade de sobrevivência, e vice-versa. No contexto deste trabalho, essa probabilidade representa a chance de um estudante

⁴ <https://lifelines.readthedocs.io/>

permanecer na instituição até a quebra do vínculo, fornecendo assim informações ao longo do tempo.

	duracao_vinculo	evento	censurado	grupo_id_discente
0	13	1	<input type="checkbox"/>	3
1	24	0	<input checked="" type="checkbox"/>	4
2	5	0	<input type="checkbox"/>	5
3	29	0	<input checked="" type="checkbox"/>	6
4	26	0	<input type="checkbox"/>	7
5	21	1	<input checked="" type="checkbox"/>	8
6	1	1	<input type="checkbox"/>	9
7	15	0	<input checked="" type="checkbox"/>	10
8	20	0	<input type="checkbox"/>	11
9	15	0	<input checked="" type="checkbox"/>	12

Figura 8 – Input dos dados.

	index	grupo	time	prob	ci_lower	ci_upper
0	1	NOTURNO	4	1	1	1
1	2	MATUTINO	6	0.9995	0.9991	0.9991
2	3	NOTURNO	3	0.999	0.9982	0.9982
3	4	NOTURNO	14	0.9985	0.9973	0.9973
4	5	NOTURNO	4	0.998	0.9964	0.9964
5	6	INTEGRAL	4	0.9975	0.9955	0.9955
6	7	INTEGRAL	4	0.997	0.9946	0.9946
7	8	MATUTINO	11	0.9965	0.9937	0.9937
8	9	VESPERTINO	5	0.996	0.9928	0.9928
9	10	MATUTINO	13	0.9955	0.9919	0.9919

Figura 9 – Output dos dados.

Finalmente, os resultados são compartilhados por meio do algoritmo de análise de sobrevivência implementado. O gráfico é construído com base no *DataFrame* contendo as probabilidades, referentes ao agrupamento selecionado no filtro, o evento de morte e a duração máxima de vínculo a ser analisada. Utilizando a biblioteca Plotly⁵, responsável por recursos visuais e interativos, é exibida a visualização dinâmica e informativa das trajetórias de sobrevivência para os grupos, representadas como linhas no gráfico. Cada linha é colorida de forma distinta para facilitar a identificação, e o gráfico inclui informações detalhadas sobre a probabilidade de sobrevivência em cada ponto temporal. A *dashboard* incorpora filtros interativos de curso, campus, turno, período de início e término, os quais adicionam segmentações de seleção aos dados.

Destacando a combinação de técnicas visuais e estatísticas na interpretação e comunicação dos resultados, o uso do SABIA permite que os usuários monitorem como diferentes grupos evoluem em termos de probabilidades de sobrevivência. Os cursos analisados foram escolhidos com base na área de conhecimento em que estão classificados, o que facilita algumas análises e comparações, auxiliando na validação do método proposto neste trabalho para análise de permanência e comparações que justifiquem investimentos e desperdícios de recursos.

⁵ <https://plotly.com/python/>

5 Resultados

Nesta seção, são apresentados os resultados obtidos a partir do modelo de análise de sobrevivência proposto. Esses resultados decorrem da análise dos agrupamentos e fornecem uma compreensão do estado de vínculo dos estudantes até a interrupção do evento, que pode ser *EVADIDO*, *FORMADO* ou ambos. Ao explorar o comportamento dos eventos, examina-se a probabilidade de permanência em relação à duração do vínculo dos estudantes e realizam-se comparações entre grupos para avaliar a permanência. As análises revelaram padrões distintos conforme o grupo acompanhado, com comportamentos semelhantes entre cursos pertencentes a área de computação, nos estágios iniciais observados. Esses resultados fornecem uma base para discussões mais aprofundadas e podem ajudar a mitigar investimentos associados a problemas de evasão estudantil.

5.0.1 Plataforma BI

A plataforma SABIA apresenta painéis que fornecem informações específicas sobre a instituição, cursos, disciplinas e estudantes. O SABIA foi desenvolvido com o uso da biblioteca Streamlit¹, voltada para a criação de aplicativos web. A Figura 10 apresenta a página inicial desta plataforma, exibindo as opções de painéis, incluindo o painel *Permanência*, resultante deste trabalho.



Figura 10 – Visualização inicial do SABIA.

O painel deste trabalho integra a plataforma de análises acadêmicas, com foco na análise de permanência dos estudantes. Os filtros de seleção que constam no projeto personalizam a visualização dos dados, incluindo opções para escolha dos cursos,

¹ <https://streamlit.io/>

seleção do ano de início e término, filtros por campus, turno e semestre de ingresso como ilustrado na Figura 11. Os detalhes na abordagem refinam a apresentação dos resultados para compreensão dos padrões presentes nos dados analisados.

ANÁLISE DE PERMANÊNCIA

Este painel apresenta informações de permanência dos discentes em relação ao vínculo estudantil.

Filtrar

TIPO DO SEMESTRE

REGULAR x

CAMPUS

Choose an option

TIPO DO CURSO

REGULAR x

CURSO

Choose an option

TURNO

Choose an option

ANO DE INGRESSO INICIAL

2010

ANO DE INGRESSO FINAL

2023

SEMESTRE DE INGRESSO

Choose an option

Figura 11 – Filtros da plataforma SABIA.

Ao iniciar o painel, além dos filtros gerais, estão disponíveis os filtros de agrupamento de seleção, que definem a variável de influência observada no evento. O agrupamento inclui curso, campus, semestre de ingresso, modalidade de ingresso, turno, área e todos (correspondente aos estudantes que sobreviveram ao experimento e concluíram a formação). Além disso, é possível definir o evento de término da observação e controlar a duração máxima observada com um controle deslizante, conforme ilustrado na Figura 12.

ANÁLISE DE SOBREVIVÊNCIA

A análise de sobrevivência, utilizada na investigação da permanência de estudantes no ensino superior, é empregada para calcular a probabilidade de um estudante permanecer matriculado em um curso ao longo do tempo, permitindo comparar ocorrências de um evento.

AGRUPAMENTO

TURNO

EVENTO DE MORTE

EVADIDO x FORMADO x

DURAÇÃO MÁXIMA

1 20 26

Figura 12 – Filtros do modelo de sobrevivência.

Por fim, ao carregar o painel com a seleção dos filtros, os resultados das informações são apresentados em um gráfico. Nesse contexto, são exibidas as probabilidades

de sobrevivência em relação à duração do vínculo, calculadas a partir da função de sobrevivência, como observado na Figura 13.

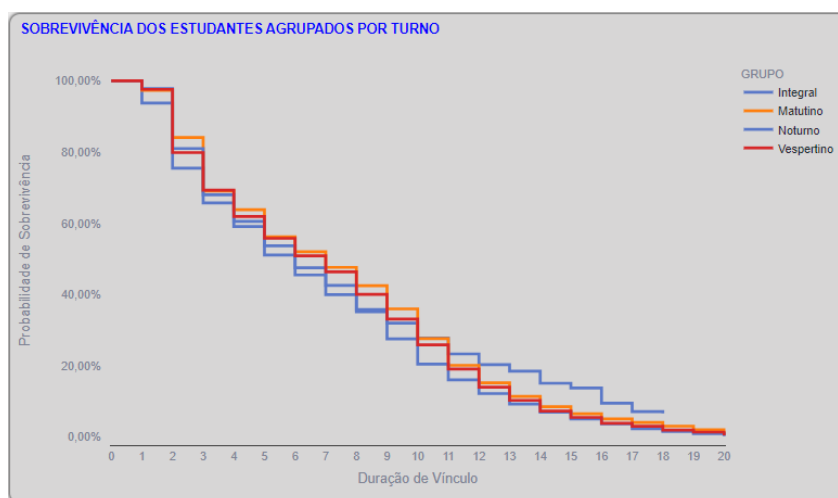


Figura 13 – Gráfico resultante da função de sobrevivência.

O gráfico gerado pelo modelo de análise de sobrevivência representa as curvas de sobrevivência, retratando a probabilidade acumulada de permanência na graduação ao longo do tempo, e a variação da curva indica como essa probabilidade se modifica com o decorrer do tempo. No eixo vertical, encontra-se representada a probabilidade de um estudante permanecer na graduação ao longo do tempo. No início do curso, essa probabilidade é alta (100%) devido à ausência de abandono; à medida que o tempo avança, a probabilidade de permanência na graduação tende a diminuir. O eixo horizontal, referente à escala, indica a duração em que o vínculo estudantil está sendo observado.

5.0.1.1 Aplicações da ferramenta

Como possibilidades de utilização da ferramenta de análise de sobrevivência no painel de permanência, os usuários podem calcular as curvas de permanência dos estudantes obtendo indicadores do sucesso acadêmico, analisando os fatores que influenciam a permanência nos cursos, o que auxilia gestores a tomar decisões informadas para evitar a evasão a partir desses dados e com isto, reduzir desperdício de recursos associados ao abandono estudantil.

A comparação de grupos e das curvas resultantes permite identificar se cursos da mesma área demonstram comportamentos similares. Além disso, a análise das tendências de probabilidades ajuda a entender como a permanência dos alunos varia ao longo do curso e em diferentes grupos. Observando a variação e as probabilidades referentes a cada grupo, também é possível identificar a velocidade dessa variação.

A velocidade das variações nas curvas de sobrevivência pode indicar a ocorrência de eventos ou fatores críticos que influenciam a decisão dos alunos de continuar ou

interromper seus estudos, indicando se os estudantes enfrentam maior risco de evasão ou estabilidade da permanência. O acompanhamento da sobrevivência também pode ser em relação à latitude e longitude nas curvas resultantes de cada grupo, a latitude indica a probabilidade de um aluno permanecer em um determinado ponto do tempo, enquanto a longitude representa a duração em que essa probabilidade se mantém.

As variáveis podem ser agrupadas para identificar a permanência em diferentes variáveis, localizações geográficas com base no campus, analisar a influência do semestre de ingresso na permanência, investigar o impacto do horário das aulas por turno e avaliar se diferentes formas de entrada dos alunos afetam a permanência. A evolução da ferramenta de obtenção de indicadores, pode incluir outros agrupamentos para análises futuras, o que permite explorar a permanência dos alunos em mais grupos, como perfil socioeconômico e histórico acadêmico, auxiliando a identificação de perfis de risco e a avaliação de políticas e programas.

A inclusão de agrupamentos mais direcionados a fatores socioeconômicos e histórico acadêmico futuros contribuirão para aprofundar a compreensão dos fatores que influenciam a permanência dos alunos. As diferenças nas probabilidades de permanência conseguem comparar o desempenho de alunos de diferentes faixas de renda. Também se tem potencial para determinar quais grupos apresentam maior risco de evasão e quais características, investigar como diferentes variáveis se cruzam e impactam a permanência, como o efeito do perfil socioeconômico em conjunto com o histórico acadêmico, e verificar o impacto de políticas e programas de apoio à permanência em diferentes grupos de estudantes. Com a inclusão de novos agrupamentos, a ferramenta oferece uma visão mais completa da permanência, com probabilidades em relação a etnia, renda, desempenho, notas em exames de ingresso, participação em atividades extracurriculares, região, cidade, permitindo aos gestores tomar decisões mais assertivas.

5.0.2 Curvas de sobrevivência

As curvas de sobrevivência representam, no presente trabalho, a permanência dos estudantes no ensino superior ao longo do tempo. Essas curvas são obtidas por meio da análise de dados acadêmicos utilizando o modelo de análise de sobrevivência desenvolvido. Graficamente, as curvas representam a probabilidade acumulada de permanência dos estudantes na graduação em relação à duração do vínculo estudantil. Essas curvas revelam como a probabilidade de permanência na graduação varia à medida que o tempo avança, refletindo os eventos que ocorrem. À medida que o tempo passa, essa probabilidade tende a diminuir.

As curvas são interpretadas de acordo com suas características, levando em consideração a inclinação, ponto de estabilidade, início da curva, declínio ao longo do

tempo e interseção com o eixo vertical. No contexto deste trabalho, a inclinação da curva indica a chance de mudança na probabilidade de permanência dos estudantes ao longo do tempo. Curvas mais íngremes indicam probabilidades mais rápidas de evasão ou queda na permanência. Isso significa que os estudantes estão deixando o curso a uma probabilidade mais alta em comparação com uma curva menos íngreme. Por outro lado, uma curva menos íngreme refere-se a uma chance de evasão mais lenta ou uma diminuição mais gradual na probabilidade de permanência ao longo do tempo.

O ponto de estabilidade presente nas curvas de sobrevivência corresponde ao ponto onde a curva atinge um patamar horizontal, indicando a probabilidade de que os estudantes que chegaram até esse ponto permaneçam no curso até o final do período observado. Este patamar representa a estabilidade de permanência, quanto mais alto, maior é a proporção de estudantes que permanecem no curso em comparação com aqueles que deixam. O início da curva representa o ponto inicial da jornada dos estudantes no curso, onde a probabilidade de permanência é frequentemente alta, quase 100%. Este estágio inicial sugere que os estudantes estão ingressando no curso e ainda não enfrentaram os desafios que podem levar à evasão.

Ao tratar do declínio ao longo do tempo, a curva reflete a tendência de diminuição na probabilidade de permanência dos estudantes à medida que o tempo avança e pode ocorrer de maneira gradual ou abrupta. Dependendo de cada curso e dos fatores que influenciam a evasão, uma diminuição gradual representa que os estudantes estão enfrentando desafios ao longo do tempo que afetam sua capacidade de permanecer no curso, enquanto um declínio abrupto pode indicar a ocorrência de eventos específicos que levam à evasão em massa. A interseção da curva com o eixo vertical, que ocorre no ponto onde a probabilidade é de 100% no início do período observado, serve como ponto de referência inicial da análise, representando o ponto de partida da trajetória estudantil.

No painel de permanência, as curvas podem ser exploradas usando filtros de seleção que permitem personalizar a visualização dos dados de acordo com suas necessidades específicas. Os resultados oferecem uma visão detalhada da permanência dos estudantes no ensino superior e são direcionados à gestão acadêmica e ao planejamento estratégico das instituições de ensino superior.

5.0.3 Análise de sobrevivência para os cursos de computação

Para demonstrar as funcionalidades do modelo de análise de sobrevivência resultante é feita a análise dos cursos de computação da *Universidade Federal Rural de Pernambuco (UFRPE)*. Dado que nem todos os campi apresentam os três cursos de computação, manteve-se o foco das análises nos cursos do campus da Sede para os

períodos de 2010.1 a 2023.1.

Para os fins deste trabalho, a censura utilizada é a direita, o que significa que as observações são encerradas quando um evento de interesse não ocorre dentro do período de observação, neste caso, quando um aluno atinge o limite de períodos selecionado pelo usuário. Para os estudantes que não se formaram nem evadiram, ou seja, que permaneceram na instituição até o final do período de observação, se aplica a censura à direita. Isso ocorre porque não houve o evento de interesse (*FORMADO* ou *EVADIDO*) dentro do período de análise, e portanto, eles são censurados à direita.

O evento de morte refere-se ao estado final do aluno em relação ao seu vínculo com a instituição de ensino. No contexto deste trabalho, o evento de morte é utilizado para representar a ocorrência do evento de interesse, que é a evasão dos estudantes. Portanto, quando um aluno evadiu, seu vínculo é registrado como *EVADIDO*. Esse evento indica que o aluno deixou de renovar a matrícula em um semestre letivo específico, sem se matricular no semestre seguinte. Os estudantes que evadiram dentro do período de observação são considerados como eventos de interesse mortos, pois atingiram o resultado final de *EVADIDO*. Os estudantes que permanecem matriculados no final do período de observação são os casos censurados à direita, pois seus destinos finais (se evadirão ou se formarão) são desconhecidos dentro do período de análise, não há alunos considerados como mortos devido à formatura, pois o evento de morte está limitado apenas aos casos de evasão.

O painel ainda fornece a opção de seleção do evento de morte para *FORMADO* e *EVADIDO* simultaneamente, nesse caso, a censura à direita ocorrerá para os estudantes que ainda estão matriculados e o destino final (evadidos ou formados) é desconhecido no momento da análise. Nesse caso, os estudantes que se formaram dentro do período de observação são considerados como eventos de interesse mortos, pois atingiram o resultado final de *FORMADO*. Os estudantes que evadiram dentro do período de observação também são considerados como eventos de interesse mortos, pois atingiram o resultado final de *EVADIDO*. E os estudantes que permanecem matriculados no final do período de observação são os casos censurados à direita, pois não se tem conhecimento dos seus destinos finais (se serão formados ou evadirão) são desconhecidos dentro do período de análise.

O estudo de caso realizado nos cursos de Bacharelado em Sistemas de Informação (BSI), Bacharelado em Ciência da Computação (BCC) e Licenciatura em Computação (LC) da sede, ocorreu por meio das comparações dos agrupamentos semestre, modalidade de ingresso e todos. Para estes, serão apresentados os resultados dos cursos, analisando como se comportam em cada agrupamento, quando comparados entre si, evidenciando graficamente as probabilidades de permanência quando o evento de morte é *EVADIDO*. Para demonstrar o comportamento entre áreas distintas, será

feito a comparação entre os cursos de computação e o cursos de Bacharelado em Medicina Veterinária (VET) por agrupamento todos com o evento de morte *EVADIDO* e *FORMADO* ao mesmo tempo. Outros agrupamentos, como turno, campus e área, estavam restritos a apenas um dos cursos analisados, limitando sua aplicabilidade para uma análise comparativa abrangente entre os cursos.

5.0.3.1 Semestre

O ano letivo é dividido em dois semestres principais: o primeiro semestre, que ocorre no início do ano, e o segundo semestre, que se desenvolve nos meses seguintes. Cada semestre é caracterizado por um conjunto específico de disciplinas ou cursos oferecidos aos estudantes, juntamente com atividades acadêmicas. Essas atividades são distribuídas entre aulas, projetos, exames e avaliações, contribuindo para a organização do currículo, a gestão do tempo e a progressão dos alunos.

Na Tabela 4, observa-se que Sistemas de Informação e Ciência da Computação apresentam probabilidades iniciais de evasão mais baixas em relação à Licenciatura em Computação. Enquanto nos primeiros períodos de vínculo as probabilidades de sucesso para BSI e BCC permanecem entre os 90%, logo no 2º período LC apresenta probabilidade de 89,39% de permanência. Embora o comportamento dos gráficos seja semelhante entre os cursos, com valores cada vez mais reduzidos de permanência até a evasão, as probabilidades de permanência no 2º semestre são menores do que as do 1º semestre. Os gráficos de sobrevivência na Figura 14 ilustram visualmente as tendências de evasão por curso para os estudantes ingressantes no 1º semestre, destacando a diferença de como as probabilidades de permanência diminuem em cada curso ao longo do tempo até o fim do período observado.

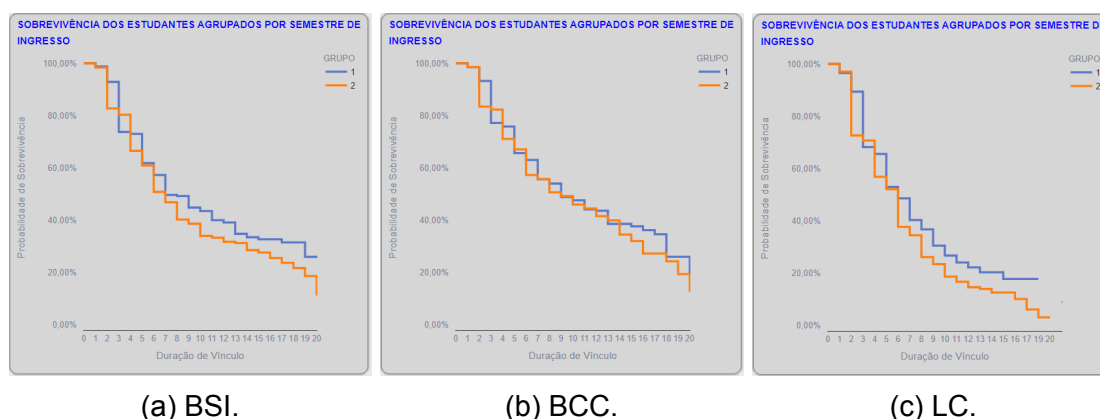


Figura 14 – Gráficos de sobrevivência para os estudantes que ingressaram nos cursos de computação por semestres.

Na interpretação das curvas de sobrevivência, observa-se que no 1º semestre tanto em Sistemas de Informação quanto em Ciência da Computação, as curvas apresentam uma inclinação relativamente suave, indicando uma diminuição gradual na

probabilidade de permanência dos estudantes ao longo do tempo. Apesar dos desafios que afetam a permanência, a evasão não ocorre em um ritmo acelerado. O ponto de estabilidade é alcançado por volta do 6º período, estabilizado em torno de 60-65%, expressando que a maioria dos estudantes que chega a esse estágio tende a permanecer no curso até o final do período observado. O início da curva mostra uma probabilidade próxima de 100%. Na Licenciatura em Computação, nota-se uma inclinação mais acentuada, referente a uma diminuição mais rápida na probabilidade de permanência dos estudantes. Apesar disso, o ponto de estabilidade é atingido em torno do 4º período, com uma probabilidade de permanência em torno de 60-70%, sugerindo que, mesmo com a inclinação íngreme, uma proporção significativa de estudantes ainda permanece no curso nesse estágio.

Ao comparar os semestres, de forma geral, percebe-se que os estudantes que ingressaram no 1º semestre apresentam probabilidades de permanência mais altas em comparação com aqueles que ingressaram no 2º semestre, o que significa uma maior evasão para este último. No entanto, há variações ao longo do tempo e entre os cursos. Embora os estudantes do 1º semestre geralmente comecem com chances de permanência mais baixas, essa diferença diminui à medida que o tempo passa, com alguns cursos do 2º semestre alcançando ou até ultrapassando as probabilidades de permanência dos seus pares do 1º semestre em períodos posteriores. Também é perceptível quedas mais abruptas nas probabilidades de permanência nos períodos iniciais, indicando uma maior incidência de evasão, onde no 2º semestre esse fenômeno é percebido primeiro do que no 1º semestre, isso pode estar relacionado a uma fase de adaptação dos estudantes no curso, mas que evidenciam os três primeiros períodos de vínculo como pontos críticos que devem ser acompanhados, curso a curso.

Na Licenciatura em Computação, as chances de permanência são mais baixas em comparação com Sistemas de Informação e Ciência da Computação. O 9º período reflete a disparidade nas probabilidades entre os cursos, onde as diferenças entre os cursos são: BSI com 40,18%, BCC com 50,62% e LC com 26,06%. Essas diferenças destacam a importância de investigações mais aprofundadas por parte dos coordenadores de curso, com o objetivo de compreender melhor as razões. Além disso, as diferenças nas probabilidades de permanência entre os cursos quando se analisados os semestres acadêmicos podem levantar questões sobre a adequação do currículo, os métodos de ensino utilizados e o ambiente acadêmico em geral, sugerindo áreas que podem ser melhoradas para promover o sucesso dos estudantes em todos os cursos.

Semestre

Probabilidade de Evasão para os estudantes que ingressaram no 1º Semestre														
Período	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Sistemas de Informação	100%	98,77%	92,89%	73,70%	72,97%	61,82%	57,27%	49,62%	49,16%	44,77%	43,43%	39,91%	39,05	34,71%
Ciência da Computação	100%	98,56%	93,51%	77,16%	75,81%	61,65%	63,01%	55,65%	53,96%	48,81%	47,57%	44,07%	43,56%	38,53%
Licenciatura da Computação	100%	96,52%	89,39%	68,17%	65,53%	52,95%	48,54%	40,23%	36,69%	30,47%	26,66%	23,99%	22,15%	20,25%
Probabilidade de Evasão para os estudantes que ingressaram no 2º Semestre														
Período	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Sistemas de Informação	100%	98,34%	82,75%	80,28%	66,45%	60,85%	50,78%	46,80%	40,18%	38,59%	33,94%	33,25%	31,70%	31,21%
Ciência da Computação	100%	98,57%	83,42%	82,28%	71,04%	67,07%	57,25%	55,62%	50,62%	49,20%	45,94%	44,41%	41,48%	39,88%
Licenciatura da Computação	100%	97,05%	72,64%	70,69%	56,81%	52,04%	37,65%	34,46%	26,06%	23,39%	18,57%	16,62%	14,54%	13,8%

Tabela 4 – Probabilidades de evadidos por semestre.

5.0.3.2 Modalidade de Ingresso

A modalidade de ingresso na graduação refere-se à classificação dos processos de admissão em uma instituição de ensino superior, sendo ela dividida em dois grupos: *Regular* e *Outros*. A modalidade *Regular* abrange os métodos de admissão Sistema de Seleção Unificada (SISU), Vestibular, e também modalidades especiais relacionadas ao SISU. A modalidade *Outros* engloba todas as outras formas de ingresso, que podem variar de acordo com políticas específicas da instituição, como seleções por transferência, reintegração ou processos seletivos internos.

A Tabela 5 expressa as probabilidades de permanência, ao comparar as modalidades de ingresso *Regular* e *Outros*, nota-se ocorrência de queda nas chances de permanência na modalidade *Regular*, com casos de evasão já registrados após o primeiro período em todos os cursos de computação analisados. Por outro lado, na modalidade *Outros*, os períodos iniciais apresentam chance de permanência um pouco superior. Entretanto, ao comparar os períodos finais, observa-se que no 9º período, a probabilidade de permanência em *Regular* cai abaixo de 50% nos cursos, com Licenciatura em Computação registrando a menor chance de sucesso.

As curvas de sobrevivência na Figura 15 demonstram, que em Sistemas de Informação, na modalidade *Regular*, a probabilidade de evasão diminui gradualmente ao longo do tempo, estabilizando-se em torno de 33,24% no 13º período. Por outro lado, na modalidade *Outros*, a probabilidade de permanência também decresce ao longo do tempo, iniciando em 100% e estabilizando-se em cerca de 27,40% no 13º período. Os estudantes na modalidade *Outros* alcançam uma estabilidade mais rapidamente em comparação com os da modalidade *Regular*.

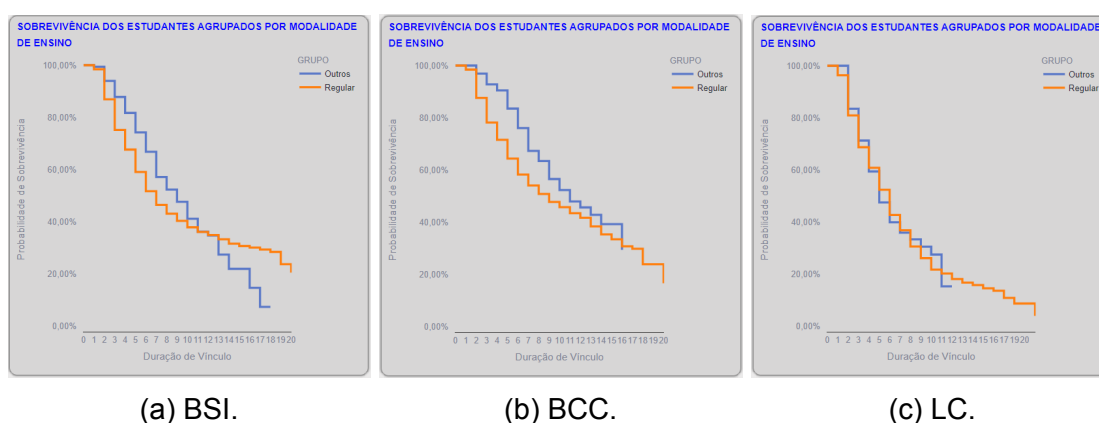


Figura 15 – Gráficos de sobrevivência para os estudantes por modalidade de ingresso nos cursos de computação.

Em Ciência da Computação, na modalidade *Regular*, a probabilidade de permanência diminui gradualmente, estabilizando-se em cerca de 38,39% no 13º período. Na modalidade *Outros*, a probabilidade de permanência segue uma trajetória seme-

lhante, estabilizando-se em aproximadamente 42,79% no 13º período. Embora a estabilidade seja alcançada mais cedo na modalidade *Outros*, esta apresenta uma chance de evasão mais alta em comparação com a modalidade *Regular*. Já em Licenciatura em Computação, na modalidade *Regular*, a probabilidade de evasão diminui de forma mais acentuada, estabilizando-se em 16,71% no 13º período. Na modalidade *Outros*, percebe-se um declínio similar na evasão, estabilizando-se em 0% no 13º período. Nesse caso, os estudantes na modalidade *Outros* demonstram probabilidades de permanência menores em comparação com a modalidade *Regular*. Embora os estudantes na modalidade *Outros* alcancem uma estabilidade mais rápida, a probabilidade de evasão é geralmente mais alta nos períodos iniciais do que na modalidade *Regular*, exceto na Licenciatura em Computação, onde a chance de permanência atinge 0% na modalidade *Outros*.

Ao considerar todas as modalidades em conjunto por curso, conclui-se que Bacharelado em Ciência da Computação se destaca como a opção mais favorável em termos de permanência dos alunos, apresentando melhores chances de permanência, uma vez que registra as probabilidades mais altas tanto na modalidade *Regular* quanto na modalidade *Outros*. A probabilidade de evasão aumenta ao longo do tempo, à medida que as chances de permanência são reduzidas, torna-se evidente a dificuldade em manter os alunos matriculados até a conclusão do curso. Ao comparar as modalidades, é perceptível a influência mais imediata da saída do estudante do sistema, associada à modalidade *Regular* nos períodos iniciais, até o 3º período há uma tendência a um declínio abrupto das probabilidades, tendo maior estabilidade nos períodos finais. Enquanto que a modalidade *Outros* embora sofra influência do evento de interesse nos períodos iniciais, as chances de permanência decaem mais rapidamente, também com maiores variações nos períodos finais, tendo suas últimas ocorrências observadas antes da modalidade *Regular*, que por sua vez permanece mais tempo.

Modalidade de Ingresso

Probabilidade de evasão por modalidade de ingresso <i>Regular</i>														
Período	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Sistemas de Informação	100%	98,44%	86,92%	75,19%	67,64%	59,06%	51,70%	46,45%	43,08%	40,29%	37,85%	36,07%	34,90%	33,24%
Ciência da Computação	100%	98,44%	87,61%	78,19%	71,55%	64,35%	54,22%	54,05%	50,79%	47,76%	45,73%	43,44%	41,67%	38,39%
Licenciatura da Computação	100%	96,37%	80,92%	68,71%	60,80%	52,36%	42,69%	36,80%	30,56%	26,10%	21,68%	20,22%	18,07%	16,71%
Probabilidade de evasão por modalidade de ingresso <i>Outros</i>														
Período	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Sistemas de Informação	100%	99,40%	93,95%	87,83%	81,70%	74,21%	66,79%	57,13%	52,30%	47,63%	41,14%	36,15%	34,71%	27,40%
Ciência da Computação	100%	100%	96,91%	92,78%	90,46%	83,51%	76,03%	67,30%	63,42%	56,53%	52,29%	47,93%	45,65%	42,79%
Licenciatura da Computação	100%	100%	83,49%	71,27%	59,39%	47,51%	39,85%	35,86%	33,30%	30,53%	27,47%	15,26%	15,26%	0%

Tabela 5 – Probabilidades de evadidos por modalidade de ingresso.

5.0.3.3 Todos

A seleção do agrupamento *Todos* na plataforma corresponde a todos os participantes que permaneceram *Vivos* no evento acompanhado, ou seja, todos os estudantes que mantiveram o vínculo estudantil. Neste contexto, *Vivos* refere-se a todos aqueles que permaneceram no sistema, não sofreram o evento de morte e consequentemente não evadiram durante o tempo de observação. A Figura 16 corresponde às probabilidades de permanência dos cursos que constam na Tabela 6 e serão utilizadas para as análises deste agrupamento.

Os resultados obtidos apontam o curso de Ciência da Computação com chances mais elevadas de permanência, cerca de 9% acima de Sistemas de Informação, enquanto Licenciatura em Computação apresenta probabilidades mais baixas. No 13º período, as probabilidades de permanência são de 32,49% para Sistemas de Informação, 38,64% para Ciência da Computação e 16,57% para Licenciatura em Computação. Esses números refletem a probabilidade dos alunos continuarem em seus respectivos cursos até o 13º período.

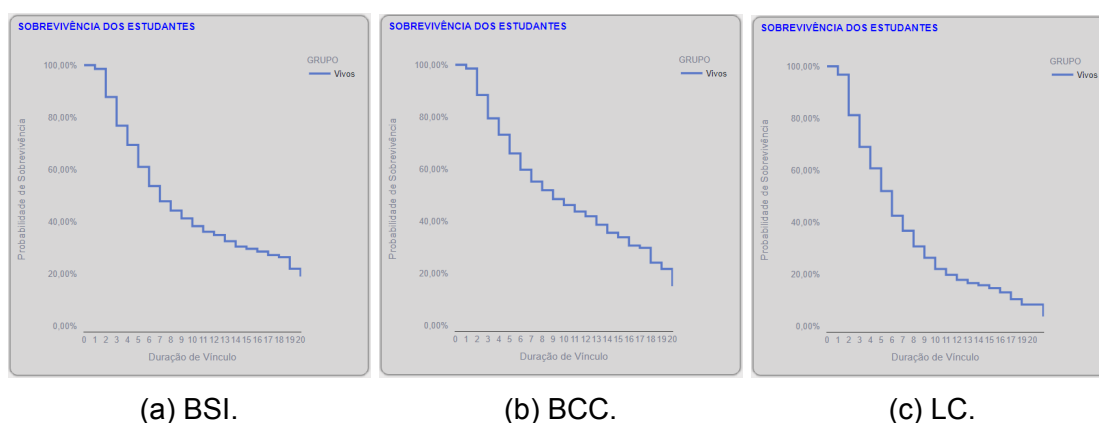


Figura 16 – Gráficos de sobrevivência para os estudantes que ingressaram nos cursos de computação permaneceram matriculados no período analisado.

A interpretação das curvas de sobrevivência evidencia que os estudantes de Ciência da Computação apresentam uma maior probabilidade de permanecer matriculados no curso até o término do vínculo. Nesse curso, a inclinação da curva segue uma tendência de diminuição gradual, com uma queda acentuada, seguida por uma diminuição mais suave. Sendo o 3º período o mais crítico, com probabilidade de sobrevivência de 88,39%, a maior variação ocorre entre o 3º e o 4º período, com uma queda de aproximadamente 8,94% e atingindo o ponto de estabilidade apenas no 9º período, com cerca de 51,85%.

Em Sistemas de Informação, nota-se um padrão similar, com o ponto de estabilidade ocorrendo mais tarde no curso, por volta do 9º período, com uma probabilidade de cerca de 44,26%. O período mais crítico é também o 3º período, onde a chance de

sobrevivência é de 87,79%. A maior variação é observada entre o 3º e o 4º período, com uma queda de aproximadamente 11,02%. No curso de Licenciatura em Computação, a queda nas chances de sobrevivência é mais pronunciada, resultando em uma curva mais acentuada. O ponto de estabilidade é alcançado no 9º período, com uma probabilidade de 30,70%. Assim como nos outros cursos, o 3º período é identificado como o mais crítico, com a maior variação ocorrendo entre o 3º e o 4º período, com uma queda de aproximadamente 12,21%, sendo a maior entre os cursos de computação.

Esses comportamentos podem ser atribuídos a uma série de fatores. Nos cursos de Ciência da Computação e Sistemas de Informação, os desafios iniciais podem estar relacionados à adaptação dos alunos ao rigor do currículo, à dificuldade dos conceitos abordados ou até mesmo à falta de preparação prévia em disciplinas fundamentais. Por outro lado, na Licenciatura em Computação, além desses desafios, os alunos podem enfrentar dificuldades adicionais relacionadas à combinação de conteúdo técnico com habilidades pedagógicas necessárias para a carreira de professor, o que pode explicar a queda mais expressiva nas probabilidades de sobrevivência.

Todos

Todos os estudantes que permaneceram vivos														
Período	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Sistemas de Informação	100%	98,57%	87,79%	76,77%	69,42%	61,01%	53,65%	47,83%	44,26%	41,23%	38,28%	36,10%	34,89%	32,49%
Ciência da Computação	100%	98,57%	88,39%	79,45%	73,15%	65,98%	59,74%	55,17%	51,85%	48,46%	46,21%	43,63%	41,86%	38,64%
Licenciatura da Computação	100%	96,77%	81,20%	68,99%	60,74%	52%	42,49%	36,73%	30,70%	26,34%	22,01%	19,77%	17,82%	16,57%

Tabela 6 – Probabilidades de sobrevivência por todos.

5.0.3.4 Todos por evento de morte formado e evadido

Para investigar o comportamento dos cursos quando se trata de evasão e formação de alunos, foi realizado um comparativo entre os cursos de computação, Bacharelado em Sistemas de Informação (BSI), Bacharelado em Ciência da Computação (BCC) e Licenciatura em Computação (LC) e representando uma área de estudo diferente dos demais o curso de Bacharelado em Medicina veterinária (VET). O objetivo é analisar o comportamento entre diferentes áreas acadêmicas.

Na medicina veterinária, a probabilidade de permanência dos alunos diminui progressivamente ao longo do tempo, com um declínio gradual, conforme representado na Tabela 7, começando em 100% e reduzindo para 20,53% ao final do 13º período de observação, a estabilização ocorre por volta do 8º período, com 71,31%, com a maior variação entre o 10º e o 11º período, tendo seu ponto mais crítico no 11º período com probabilidade de permanência de 40,07%. Isso indica que a maioria dos alunos permanece no curso nos estágios iniciais, mas uma proporção significativa deles evade ou se forma nos períodos seguintes.

Da mesma forma que na medicina veterinária, a probabilidade de permanência diminui ao longo do tempo nos cursos de Sistemas de Informação e Ciência da Computação. Começando em 100% no primeiro período, essa probabilidade cai para 22,44% no 13º período para Sistemas de Informação, e para 23,91% no mesmo período para Ciência da Computação. Já na Licenciatura em Computação, observa-se uma diminuição constante na probabilidade de permanência. Iniciando em 100% no primeiro período, essa probabilidade reduz para 12,94% no décimo terceiro período.

Em Sistemas de Informação o ponto de estabilidade ocorre por volta do período 9, com 44,01%, o período 1º corresponde ao maior ponto crítico, sendo representado por uma curva decrescente e gradual e apresentando nos períodos iniciais variações de 10,78% no 1º período e 11,02% no 2º período. Tendo ocorrido o ponto de estabilidade após o 3º período. Ciência da Computação também apresenta um declínio gradual, com maior variação entre o 4º e o 5º período, com uma diferença de 13,67% e o ponto de estabilidade só ocorre após o 5º período. Já o período mais crítico é o 1º, onde é possível perceber um declínio abrupto das probabilidades de permanência. Em Licenciatura da computação, a curva resultante é a mais acentuada, a maior variação ocorre entre o 1º e o 2º, com uma diferença de 14,95%, o ponto mais crítico ocorre no 5º período, onde a probabilidade de sobrevivência é de 51,77%, o que indica uma alta chance de evasão ou desistência no curso nesse período. É possível notar a ocorrência do declínio abrupto logo no 2º período e que embora nos outros cursos também demonstrem essa tendência, na Licenciatura em Computação este comportamento é mais expressivo

As curvas de sobrevivência resultantes de cada curso, representadas na Figura 17, oferecem uma visão das trajetórias dos alunos ao longo de seus programas acadêmicos. Em Medicina Veterinária, Sistemas de Informação, Ciência da Computação e Licenciatura em Computação, observa-se uma tendência comum de queda na probabilidade de permanência ao longo do tempo. No entanto, o declínio varia entre os cursos. Em Medicina Veterinária, a diminuição é gradual, sugerindo uma saída gradual de alunos ao longo dos períodos. Sistemas de Informação e Ciência da Computação exibem uma queda moderada, indicando desafios contínuos ao longo do curso. Por outro lado, na Licenciatura em Computação, a queda é mais acentuada, correspondendo a uma chance mais rápida de desistência ou conclusão do curso.

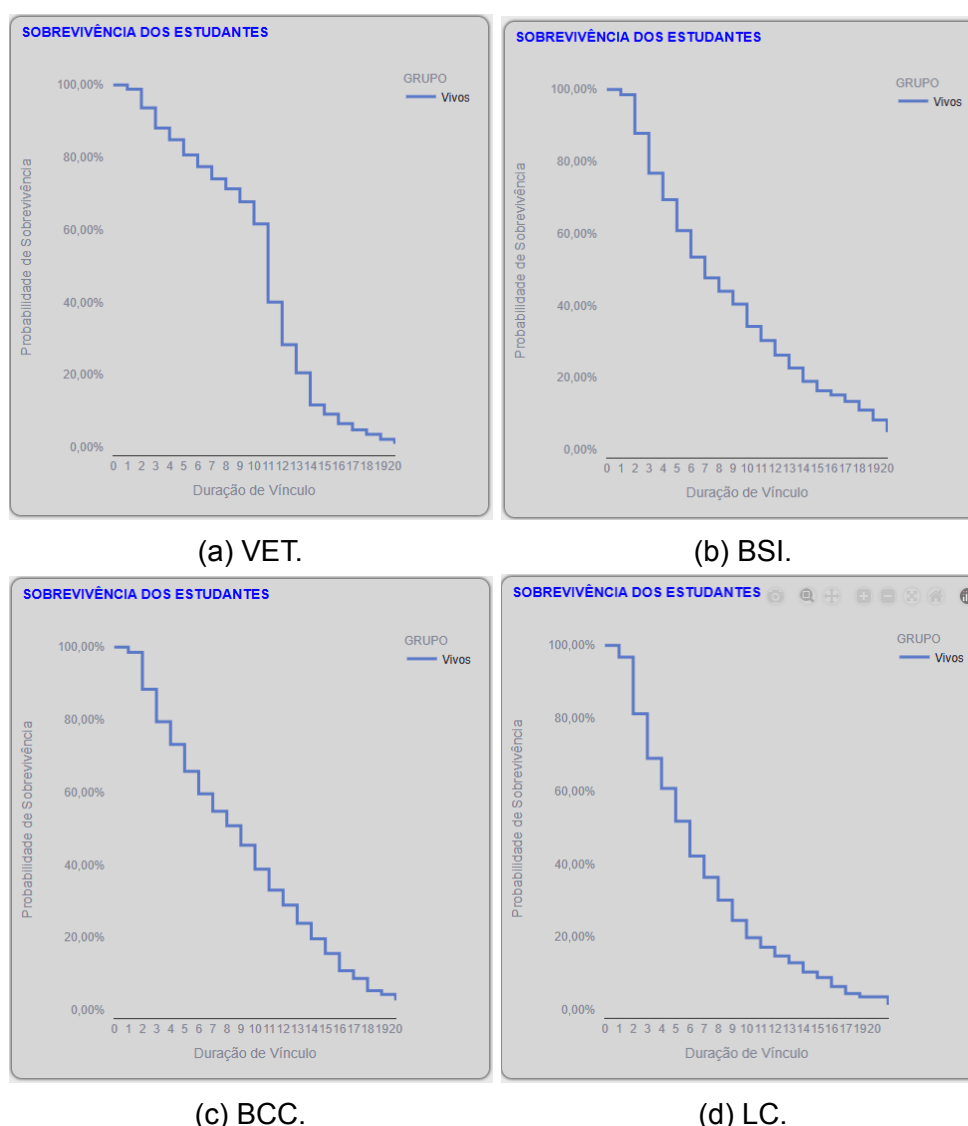


Figura 17 – Gráficos de sobrevivência para os estudantes que ingressaram por todos no evento de morte *FORMADO* e *EVADIDO*.

Quando considerado o evento de interesse como formados e evadidos, no 20º período as probabilidades de permanência estão próximas a 0%. Isso se deve ao fato de que a maioria dos alunos já concluiu o curso ou desistiu até esse ponto. No 20º

período, em Medicina veterinária, a probabilidade de permanência é de 1,36%, em Sistemas de Informação é de 5,32%, em Ciência da Computação é de 3,08% e em Licenciatura da Computação é de 3,59%. Os cursos de computação têm valores de probabilidade de permanência maiores no vigésimo período, indicando que uma proporção relativamente maior de alunos de computação continua no curso por períodos mais longos, além do período de formação estabelecido, em comparação com os alunos de veterinária. Isso pode refletir diferenças na complexidade dos cursos, desistências, na demanda de mercado, em demandas específicas de cada disciplina e nos diferentes níveis de preparação e interesse dos alunos ou em outras variáveis que influenciam a trajetória educacional dos alunos em cada área.

Todos por evento de morte formado e evadido

Período	Todos por evento de morte formado e evadido													
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Medicina Veterinária	100%	98,83%	93,67%	88,08%	84,08%	80,70%	77,42%	74,07%	71,31%	67,76%	61,67%	40,07%	28,30%	20,53%
Sistemas de Informação	100%	98,57%	87,79%	76,77%	69,42%	60,82%	53,48%	47,68%	44,01%	40,39%	34,23%	30,33%	26,27%	22,44%
Ciência da Computação	100%	100%	98,57%	88,39%	79,45%	65,78%	59,56%	54,78%	50,77%	45,42%	38,81%	33,05%	28,92%	23,91%
Licenciatura da Computação	100%	96,77%	81,82%	68,99%	60,74%	51,77%	42,18%	36,33%	30,07%	24,53%	19,80%	17,18%	14,76%	12,94%

Tabela 7 – Probabilidades de sobrevivência - todos por evento de morte formado e evadido.

5.0.4 Considerações Finais

Com base nos resultados apresentados, é possível inferir interpretações dos dados referentes aos diferentes agrupamentos analisados nos cursos de computação, Bacharelado em Sistemas de Informação (BSI), Bacharelado em Ciência da Computação (BCC) e Licenciatura em Computação.

Em relação aos semestres, observa-se uma tendência geral de diminuição das chances de permanência ao longo do tempo, sendo o 2º semestre especialmente crítico, com uma queda pronunciada nas probabilidades de permanência. Embora haja variações entre os cursos, os estudantes que ingressaram no 1º semestre tendem a apresentar probabilidades de permanência mais altas em comparação com os que ingressaram no 2º semestre, evidenciando uma maior evasão para este último. No entanto, essa diferença tende a diminuir à medida que o tempo passa, com alguns cursos do 2º semestre alcançando ou até ultrapassando as chances de permanência dos seus pares do 1º semestre em períodos posteriores. É ainda perceptível que em ambos os semestres os períodos iniciais apresentam pontos mais críticos, com variações mais altas nas probabilidades de permanência, sendo o 2º e 3º período os mais afetados em relação a evasão.

Quanto à modalidade de ingresso, os resultados indicam que os estudantes de Ciência da Computação têm as probabilidades mais elevadas de permanência, tanto na modalidade *Regular* quanto na modalidade *Outros*, enquanto os alunos de Licenciatura em Computação apresentam as probabilidades mais baixas de permanência. Isso sugere uma maior probabilidade de os estudantes de Ciência da Computação permanecerem matriculados no curso até o fim do vínculo. Comparando as modalidades, a ocorrência da evasão é mais rápida nos períodos iniciais na modalidade *Regular*, enquanto ocorre de forma mais tardia na modalidade *Outros*, tendo um comportamento mais aproximado entre as modalidades no curso de Licenciatura em computação, com declínios abruptos no 2º período de ambos os grupos.

Por fim, ao considerar todos os alunos que permaneceram até o final do período de observação, o curso de Ciência da Computação mantém uma chance de permanência relativamente mais alta, seguida por Sistemas de Informação e Licenciatura em Computação, que registra a menor probabilidade de permanência, com quedas mais acentuadas nos períodos iniciais quando comparado a outros cursos.

Na comparação entre os cursos de computação e um curso de outra área de estudo, utilizando o evento de morte formado e evadido, observou-se que o ponto de declínio varia entre os cursos, refletindo possíveis diferenças na complexidade dos programas, desistência, demanda do mercado e em outras variáveis relevantes. Enquanto em medicina veterinária a diminuição na probabilidade é gradual, sugerindo

uma saída contínua de alunos ao longo do tempo, nos cursos de computação a queda é mais acentuada, indicando desafios enfrentados em toda a trajetória dos estudantes.

As suposições sobre os possíveis problemas enfrentados pelos estudantes incluem dificuldades em disciplinas específicas, que podem contribuir para a evasão, especialmente nos primeiros períodos dos cursos, quando os alunos estão se adaptando a novos conhecimentos e ao ambiente acadêmico. A quantidade de reprovações também pode ser um fator relevante, desmotivando os alunos e levando-os a abandonar o curso.

Problemas financeiros podem afetar a integração dos alunos ao sucesso acadêmico, pois a falta de recursos pode impedir o acesso a materiais didáticos, transporte e alimentação. Além disso, o acesso limitado à internet e aos equipamentos necessários para o desempenho das atividades, especialmente nos cursos de computação, pode dificultar a participação, entrega de atividades e acesso a recursos educacionais digitais. A demanda do mercado de trabalho também pode contribuir para quedas no desempenho dos alunos, pois o tempo e a dedicação precisam ser divididos entre as demandas de trabalho e estudo. Esses fatores combinados podem resultar na evasão dos cursos de computação. Destaca-se, então, a importância de abordagens individualizadas para apoiar os estudantes e melhorar suas chances de sucesso acadêmico.

As diferenças reiteram a importância de uma abordagem individualizada ao analisar o desempenho e a permanência dos alunos em diferentes áreas acadêmicas, levando em consideração as características específicas de cada curso e as necessidades dos estudantes. Essas análises fornecem *insights* para entender os padrões de permanência e podem orientar estratégias para melhorar probabilidades de sucesso de estudantes em diferentes cursos, visando à otimização de recursos no ambiente acadêmico.

6 Discussões e Trabalhos futuros

Os impactos da evasão no ensino superior podem causar prejuízos para as universidades, assim como permanências além do tempo estimado para a conclusão, tanto em termos financeiros quanto estruturais relacionados à manutenção dos cursos. Este trabalho buscou compreender os padrões de permanência dos alunos nos cursos de computação da UFRPE, utilizando análise de sobrevivência e a plataforma *Academic Business Intelligence and Analytics (SABIA)* que engloba *Business Intelligence (BI)* e técnicas como *Learning Analytics (LA)*, *Academic Analytics (AA)* e *Educational Data Mining (EDM)* em sua concepção.

Este trabalho teve como objetivo realizar uma análise da permanência por meio do modelo de análise de sobrevivência proposto, visando entender o comportamento de permanência em relação ao vínculo estudantil. Ao revisar os trabalhos relacionados, foram identificadas técnicas de análise de sobrevivência amplamente utilizadas para investigar a evasão estudantil, bem como plataformas direcionadas à análise de indicadores educacionais. O uso conjunto da análise de sobrevivência e do *Business Intelligence* mostrou-se eficaz para realizar análises detalhadas, representar a permanência acadêmica e identificar intervalos de tempo onde há maior propensão à evasão.

Com o uso da plataforma SABIA como elemento central deste processo, foi possível compartilhar os resultados obtidos, integrar grandes bases de dados, explorar e visualizar os resultados. O uso do BI foi o grande diferencial deste trabalho, permitindo análises contínuas em larga escala. Considerando que tais análises são realizadas semestralmente por parte dos gestores e coordenadores dos cursos, o modelo facilita a elaboração de propostas de intervenção de forma mais ágil e embasada em dados concretos. Isso facilita o acesso aos dados, garante a viabilidade do modelo em um tempo prolongado e fornece uma base sólida para futuras pesquisas e tomadas de decisão.

A investigação das probabilidades de permanência no contexto da educação no ensino superior, utilizando um modelo de análise de sobrevivência, demonstrou ser uma abordagem eficaz para comparar o comportamento de diferentes grupos e identificar aqueles que possuem maior tempo de permanência. Esta duração, muitas vezes, ultrapassa os períodos estimados ideais até a conclusão do curso, resultando em um consumo adicional de recursos da universidade. A análise comparativa entre cursos pertencentes à mesma área revelou padrões de comportamento distintos de acordo com o grupo analisado, indicando possíveis fatores para otimização de recursos. Compreender esses padrões permite às instituições de ensino superior tomar medidas pro-

ativas para oferecer suporte adicional aos alunos ou ajustar suas políticas e práticas acadêmicas, visando melhorar as chances de sucesso dos estudantes.

Nas análises das probabilidades de permanência, observam-se tendências entre os cursos analisados e diferenças entre os agrupamentos. Ao observar a ocorrência do evento de morte, identificou-se que cenários de evasão apresentam quedas mais acentuadas nas chances de permanência, com agrupamentos que resultam em probabilidades mais críticas, como no segundo semestre. Quanto à semelhança de comportamento entre cursos de uma mesma área dentro do mesmo agrupamento, identificam-se padrões distintos, mas que explicitam os três primeiros períodos como mais críticos. Por exemplo, ao analisar a modalidade de ingresso, notam-se variações marcantes entre os cursos de Bacharelado em Sistemas de Informação (BSI), Bacharelado em Ciência da Computação (BCC) e Licenciatura em Computação (LC), evidenciando a influência específica de cada curso na trajetória dos alunos. Essas descobertas destacam a importância de uma abordagem personalizada na gestão dos recursos educacionais, indicando que os investimentos devem ser direcionados a atender demandas específicas de cada programa acadêmico e assim evitar desperdício de recursos.

Como trabalhos futuros, pretende-se utilizar técnicas estatísticas para aprimorar a precisão das análises de permanência dos alunos. Isso inclui a configuração de novos filtros e melhorias nos gráficos, com marcadores de censura e intervalo de confiança, além de gráficos mais voltados à visualização dos dados estudantes, de forma a complementar a informação dos gráficos de sobrevivência e *insights* textuais automáticos. Adicionalmente, utilizar *clusterização* nos agrupamentos ano a ano possibilitará a classificação dos estudantes em *clusters* com base em características semelhantes.

Para revelar padrões de comportamento que influenciam a permanência dos alunos ao longo do tempo, incluindo a realização de análises mais detalhadas e focadas em populações mais vulneráveis, como forma de auxiliar o direcionamento de recursos complementares às ações afirmativas e à detecção de perfis de estudantes com maior risco de evasão, podem ser utilizadas técnicas de *Machine Learning*, juntamente com a análise de sobrevivência, para obter previsões e simulações de cenários futuros que podem ser comparados às probabilidades reais. Ao empregar modelos de *Machine Learning* para prever a permanência dos alunos e identificar os principais impulsionadores da evasão, é possível enxergar cenários futuros e antecipar decisões mais eficazes e direcionadas. Essas abordagens integradas têm o potencial de fornecer análises mais precisas e preditivas da permanência dos alunos nos cursos da IES, contribuindo para a melhoria contínua dos programas educacionais, a gestão eficiente de recursos e investimentos e, conseqüentemente, para o sucesso dos estudantes.

Referências

ALMEIDA, A. M. R.; CAMARGO, S. Aplicando técnicas de business intelligence sobre dados de desempenho acadêmico: Um estudo de caso. *XI Escola Regional de Banco de Dados. Caxias do Sul-RS*, 2015. Citado 3 vezes nas páginas 14, 25 e 34.

ANDIFES. *Diplomação, Retenção e Evasão nos Cursos de Graduação em Instituições de Ensino Superior Públicas*. 1996. Disponível em: <<https://www.gov.br/inep/pt-br/centrais-de-conteudo/acervo-linha-editorial/publicacoes-institucionais/estatisticas-e-indicadores-educacionais/censo-da-educacao-superior-2022-notas-estatisticas>>. Acesso em: 25 de janeiro 2024. Citado na página 19.

ANDRADE, A.; FERREIRA, S. A. Aspectos morfológicos do tratamento de dados na gestão escolar. o potencial do analytics. *Revista Portuguesa de Investigação Educacional*, v. 16, p. 289–316, 2016. Citado 4 vezes nas páginas 14, 28, 29 e 34.

BASTOS, J.; ROCHA, C. Análise de sobrevivência: conceitos básicos. *ArquiMed-Edições Científicas AEFMUP*, 2006. Citado 2 vezes nas páginas 22 e 23.

BEHR, A. et al. Dropping out of university: a literature review. *Review of Education*, Wiley Online Library, v. 8, n. 2, p. 614–652, 2020. Citado na página 13.

BOTELHO, F. R.; FILHO, E. R. Conceituando o termo business intelligence: origem e principais objetivos. *International Institute of Informatics and Systemics*, 2014. Citado 2 vezes nas páginas 25 e 33.

CAMPOS, J. D. d. S. *Fatores explicativos para a evasão no Ensino Superior através da análise de sobrevivência: o caso da UFPE*. Dissertação (Mestrado) — Universidade Federal de Pernambuco, 2016. Citado 2 vezes nas páginas 13 e 31.

CANO, J. L. Business intelligence: competir con información. *Banesto, Fundación Cultur [ie Cultural]*, 2007. Citado na página 25.

CARMINATI, G. et al. Mineração de dados educacionais visando a identificação da evasão no ensino superior. *Anais do Computer on the Beach*, v. 11, n. 1, p. 461–468, 2020. Citado 3 vezes nas páginas 14, 27 e 33.

CLOW, D. The learning analytics cycle: closing the loop effectively. In: *Proceedings of the 2nd international conference on learning analytics and knowledge*. [S.l.: s.n.], 2012. p. 134–138. Citado na página 27.

COLOSIMO, E. A.; GIOLO, S. R. *Análise de sobrevivência aplicada*. [S.l.]: Editora Blucher, 2021. Citado 5 vezes nas páginas 14, 21, 22, 23 e 24.

FILATRO, A. C. *Data Science na Educação: Presencial, a Distância e Corporativa: Presencial, a Distância e Corporativa*. [S.l.]: Saraiva Educação SA, 2020. Citado 2 vezes nas páginas 26 e 27.

HABERMAN, S. Survival models and data analysis. by regina c. elandt-johnson and norman l. johnson. john wiley and sons, new york. 1980. xvi 457 pages. *Journal of the Institute of Actuaries*, v. 108, n. 2, p. 295–296, 1981. Citado 3 vezes nas páginas 21, 23 e 24.

HEUBLEIN, U. Student drop-out from german higher education institutions. *European Journal of Education*, 2014. Citado na página 13.

HOED, R. M. Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de computação. *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, 2016. Citado 3 vezes nas páginas 13, 24 e 31.

INEP. *Metodologia de Cálculo dos Indicadores de Fluxo da Educação Superior*. 2017. Disponível em: <<https://portal.mec.gov.br/component/tags/tag/32044-censo-da-educacao-superior>>. Acesso em: 25 de janeiro de 2024. Citado 2 vezes nas páginas 13 e 20.

INEP. *Censo da Educação Superior 2022*. 2023. Disponível em: <<https://www.gov.br/inep/pt-br/centrais-de-conteudo/acervo-linha-editorial/publicacoes-institucionais/estatisticas-e-indicadores-educacionais/censo-da-educacao-superior-2022-notas-estatisticas>>. Acesso em: 25 de janeiro 2024. Citado na página 16.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, Taylor & Francis, v. 53, n. 282, p. 457–481, 1958. Citado 2 vezes nas páginas 23 e 24.

KLEINBAUM, D. G. et al. Parametric survival models. *Survival Analysis: A Self-Learning Text*, Springer, p. 289–361, 2012. Citado na página 21.

MANHÃES, L. M. B. et al. 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)*. [S.l.: s.n.], 2011. Citado na página 33.

MARQUES, E. et al. Sabia: Uma plataforma para auxiliar a gestão baseada em evidências nas instituições de ensino superior. In: SBC. *Anais do II Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil*. [S.l.], 2023. p. 71–80. Citado 6 vezes nas páginas 14, 27, 30, 33, 34 e 37.

Ministério da Educação - Secretaria de Educação Profissional e Tecnológica. *Documento Orientador Para a Superação da Evasão e Retenção na Rede Federal de Educação Profissional, Científica e Tecnológica*. 2014. Disponível em: <http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=110401-documento-orientador-evacao-retencao-vfinal&category_slug=abril-2019-pdf&Itemid=30192>. Acesso em: 25 de janeiro 2024. Citado na página 20.

OECD. *Education at a Glance 2021*. 2021. Disponível em: <<https://www.oecd-ilibrary.org/content/publication/b35a14e5-en>>. Acesso em: 20 de janeiro 2024. Citado 4 vezes nas páginas 15, 16, 19 e 20.

- PAIXÃO, A. de O.; SILVA, V. A. da; TANAKA, A. De business intelligence a data science: um estudo comparativo entre áreas de conhecimento relacionadas. In: *Congresso de Tecnologia da Informação*. [S.l.: s.n.], 2015. Citado na página 26.
- PALOMINO, P. et al. Plataformas de dados educacionais: Análise com foco no plano nacional de educação. In: SBC. *Anais do I Workshop de Aplicações Práticas de Learning Analytics em Instituições de Ensino no Brasil*. [S.l.], 2022. p. 60–68. Citado 3 vezes nas páginas 14, 27 e 33.
- PAZ, F. J.; CAZELLA, S. C. Reflexões sobre um modelo de analítica acadêmica em instituições comunitárias de ensino superior: estudo piloto. *Revista on line de Política e Gestão Educacional*, p. 2178–2192, 2021. Citado 2 vezes nas páginas 28 e 34.
- RAJ, A. et al. Modelling data pipelines. In: IEEE. *2020 46th Euromicro conference on software engineering and advanced applications (SEAA)*. [S.l.], 2020. p. 13–20. Citado na página 36.
- RODRIGUES, R. L. et al. A literatura brasileira sobre mineração de dados educacionais. In: SOCIEDADE BRASILEIRA DE COMPUTAÇÃO DOURADOS, MS. *Congresso Brasileiro de Informática na Educação*. [S.l.], 2014. p. 621–630. Citado 3 vezes nas páginas 26, 28 e 33.
- SACCARO, A.; FRANÇA, M. T. A.; JACINTO, P. d. A. 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)*, SciELO Brasil, v. 49, p. 337–373, 2019. Citado 2 vezes nas páginas 13 e 31.
- SANTOS, J. F. dos et al. Análise da evasão e do impacto da retenção estudantil no ensino superior com cadeias de markov absorventes. In: SBC. *Anais do XXI Workshop em Desempenho de Sistemas Computacionais e de Comunicação*. [S.l.], 2022. p. 49–59. Citado 2 vezes nas páginas 13 e 32.
- SANTOS-VILLALBA, M. J. et al. Incident factors in andalusian university dropout: A qualitative approach from the perspective of higher education students. In: FRONTIERS. *Frontiers in Education*. [S.l.], 2023. v. 7, p. 1083773. Citado na página 13.
- SILVA, G. P. d. Análise de evasão no ensino superior: uma proposta de diagnóstico de seus determinantes. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, SciELO Brasil, v. 18, p. 311–333, 2013. Citado 2 vezes nas páginas 13 e 32.
- SOUZA, J. A. de; KOMATI, K. S.; ANDRADE, J. O. Análise de sobrevivência: um estudo de caso em um curso de sistemas de informação. In: SBC. *Anais do XXX Workshop sobre Educação em Computação*. [S.l.], 2022. p. 392–403. Citado 2 vezes nas páginas 13 e 32.
- TURBAN, E. et al. *Business intelligence: um enfoque gerencial para a inteligência do negócio*. [S.l.]: Bookman Editora, 2009. Citado na página 25.