

MINERAÇÃO DE DADOS DE ALUNOS DE GRADUAÇÃO QUE PARTICIPARAM DE PROGRAMAS DE MOBILIDADE INTERNACIONAL

*Data Mining of Undergraduate Students Who Participated in International Mobility
Programs*

ALEXANDRE MAGNO GURGEL FIALHO
UFPE

TACIANA DE BARROS JERÔNIMO
UFPE

ROBERTA VANESSA ARAGAO FELIX DA SILVA
UFPE

JOÁS TOMAZ DE AQUINO
UFPE

Comunicação:

O XIII SINGEP foi realizado em conjunto com a 13th Conferência Internacional do CIK (CYRUS Institute of Knowledge), em formato híbrido, com sede presencial na UNINOVE - Universidade Nove de Julho, no Brasil.

MINERAÇÃO DE DADOS DE ALUNOS DE GRADUAÇÃO QUE PARTICIPARAM DE PROGRAMAS DE MOBILIDADE INTERNACIONAL

Objetivo do estudo

O objetivo geral dessa pesquisa é propor um modelo de tomada de decisão utilizando técnicas de mineração de dados, visando tratar e analisar os dados dos alunos de graduação que participaram dos programas de mobilidade internacional da UFPE.

Relevância/originalidade

Este estudo apresenta um modelo de tomada de decisão adaptado do framework CRISP-DM e do processo KDD, utilizando técnicas de mineração de dados. A solução integra fontes de dados fragmentadas para identificar padrões no desempenho dos estudantes, auxiliando na tomada de decisão.

Metodologia/abordagem

Foi adotada a abordagem de estudo de caso descritivo, com metodologia predominantemente quantitativa. A pesquisa envolveu análise bibliográfica, coleta de dados junto à Diretoria de Relações Internacionais da UFPE e aplicação de mineração de dados com os softwares SPSS Statistics e Weka.

Principais resultados

O framework CRISP-DM permitiu identificar padrões significativos de desempenho, destacando fatores-chave como conquistas acadêmicas, participação em programas e características demográficas. Esses insights fornecem uma base para refinar os critérios de seleção e melhorar a alocação de recursos para programas de mobilidade.

Contribuições teóricas/metodológicas

Um modelo adaptado do CRISP-DM e do processo KDD aplicado à gestão universitária, validando o uso de mineração de dados no ensino superior e oferecendo metodologia para análise de informações descentralizadas em instituições públicas, com base em teorias de decisão.

Contribuições sociais/para a gestão

Apresenta uma metodologia replicável para o gerenciamento de dados descentralizados em instituições públicas, aprimorando os processos decisórios e alinhando os programas de mobilidade aos objetivos institucionais. Além de estruturar políticas públicas para internacionalização da pós-graduação e graduação da instituição.

Palavras-chave: CRISP-DM, data mining, KDD, mobilidade estudantil

Data Mining of Undergraduate Students Who Participated in International Mobility Programs

Study purpose

To propose a decision-making model based on data mining techniques to process and analyze information on undergraduate students who participated in UFPE's international mobility programs, supporting strategic decision-making in university management.

Relevance / originality

Presents a model adapted from CRISP-DM and the KDD process, integrating fragmented data to identify academic performance patterns, contributing to more precise decision-making in student mobility management and enhancing institutional efficiency.

Methodology / approach

Descriptive case study with a quantitative approach, involving literature review, data collection from UFPE's International Relations Directorate, and data mining using SPSS Statistics and Weka, applying both predictive and exploratory techniques.

Main results

The CRISP-DM framework revealed performance patterns related to academic achievements, program participation, and demographic factors, enabling refinement of selection criteria and optimization of resource allocation for international mobility programs.

Theoretical / methodological contributions

An adapted model of CRISP-DM and the KDD process applied to university management, validating the use of data mining in higher education and providing a methodology for analyzing decentralized information in public institutions, based on decision-making theories.

Social / management contributions

Presents a replicable methodology for managing decentralized data in public institutions, improving decision-making processes and aligning mobility programs with institutional objectives. Additionally, it structures public policies for the internationalization of the institution's undergraduate and graduate programs.

Keywords: CRISP-DM, data mining, KDD, student mobility

MINERAÇÃO DE DADOS DE ALUNOS DE GRADUAÇÃO QUE PARTICIPARAM DE PROGRAMAS DE MOBILIDADE INTERNACIONAL

1 Introdução

A globalização caracteriza as orientações econômicas e acadêmicas do século XXI, enquanto a internacionalização da educação superior envolve políticas e práticas que adaptam o ambiente acadêmico às demandas globais (Altbach & Knight, 2017; Gacel-Avila, 2005). Nos países em desenvolvimento, como o Brasil, a internacionalização é fundamental para participar do debate global com soberania (Luce, Faundes & Mediel, 2016). Ao dedicar recursos a essa prática, as instituições de ensino superior (IES) promovem colaborações científicas e culturais, incentivando a cooperação acadêmica, a formação de equipes de pesquisa e a mobilidade de servidores e estudantes de graduação e pós-graduação (Oliveira & Freitas, 2016; Oliveira, 2013).

A internacionalização das Instituições de Ensino Superior (IES) vai além das diretrizes institucionais, abrangendo estratégias, iniciativas e diversos agentes que impactam o processo (Oliveira & Freitas, 2016). Na Universidade Federal de Pernambuco (UFPE), a internacionalização foi incorporada ao planejamento estratégico a partir de 2014, sendo monitorada desde então pela Diretoria de Relações Internacionais. Nesse contexto, a gestão dos dados relacionados aos programas de mobilidade internacional revela-se um desafio contínuo, demandando métodos aprimorados para análise e avaliação (Fialho, 2020).

A aplicação de novas tecnologias, especialmente sistemas de informação, é fundamental para a coleta, processamento, armazenamento e distribuição de dados, contribuindo para a tomada de decisão nas instituições (Fialho, 2020). Entre essas tecnologias, a mineração de dados destaca-se como uma ferramenta eficaz para explorar e analisar grandes volumes de dados, identificando padrões que subsidiam decisões estratégicas (Amaral, 2016).

Diante disso, a questão central do estudo é: como a mineração de dados pode ser utilizada para aprimorar a análise e a tomada de decisão nos programas de mobilidade internacional da UFPE? Assim, o objetivo geral da pesquisa é propor um modelo de tomada de decisão utilizando técnicas de mineração de dados, visando tratar e analisar os dados dos alunos de graduação participantes desses programas na UFPE.

2 Referencial Teórico

2.1 Processo Decisório

O processo decisório é essencialmente baseado em informações e, segundo Simon (1963), não se trata de um evento simples, mas de um processo social complexo que se desenvolve ao longo do tempo. Esse processo inclui etapas fundamentais, como a definição do problema, a avaliação das alternativas disponíveis e a escolha das soluções mais adequadas. De acordo com Fialho (2016), o processo decisório também leva em consideração a racionalidade do tomador de decisão, que é influenciada por suas crenças, pela forma como as informações são obtidas e pela análise dos custos e benefícios envolvidos, considerando, inclusive, a possibilidade de compensação entre os atributos das decisões.

As decisões podem ser classificadas em programadas e não programadas. As decisões programadas estão associadas ao nível operacional, caracterizando-se por serem rotineiras e estruturadas. Em contrapartida, as decisões não programadas, que envolvem problemas novos e de maior impacto, estão vinculadas aos níveis tático e estratégico da organização (Mintzberg, Raisinghani & Theoret, 1976; Simon, 1963).

Nesse contexto, o decisor, especialmente em posições institucionais estratégicas, enfrenta situações não estruturadas que demandam a adaptação de elementos estruturáveis e a análise de cenários complexos. Assim, o processo decisório é compreendido como "procedimentos de definição de problemas, avaliação de alternativas e escolha de uma diretriz de ações e/ou soluções".

2.2 Mineração de Dados

A mineração de dados utiliza diversos métodos de aprendizado de máquina, como classificação, predição, regressão, associação e agrupamento (Muller & Guido, 2016). Neste estudo, foram explorados especificamente os métodos de classificação, associação e agrupamento, com o objetivo de selecionar algoritmos adequados para identificar padrões relevantes nos dados analisados e propor soluções mais viáveis para o contexto da pesquisa (Fialho, 2020).

A mineração de dados permite a geração de regras de associação, que descrevem padrões de relacionamento entre os elementos de uma base de dados. Dado o volume e a complexidade dos dados, essas técnicas possibilitam identificar informações que vão além da capacidade de interpretação humana. Nesse contexto, a aplicação de etapas estruturadas auxilia na descoberta de conhecimento oculto, oferecendo *insights* valiosos para a tomada de decisão.

O processo de descoberta de conhecimento em bases de dados (KDD - *Knowledge Discovery in Databases*) é um método estruturado que visa a identificação de padrões válidos, novos, úteis e comprehensíveis em grandes volumes de dados (Fayyad et al. 1996). O KDD envolve várias etapas, desde a seleção e pré-processamento dos dados até a mineração e interpretação dos resultados, fornecendo uma abordagem sistemática para a extração de conhecimento (Witten & Frank, 2002). A mineração de dados é uma etapa de todo o processo, como demonstrado na Figura 1.

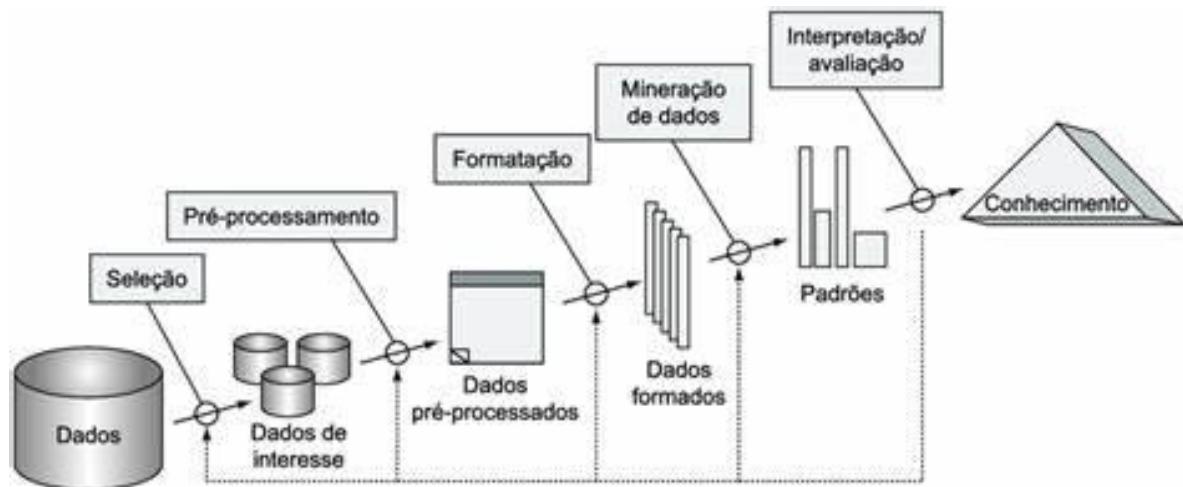


Figura 1 - Processo KDD. Fonte: Adaptado de Fayyad et al. (1996).

Outro *framework* amplamente utilizado para mineração de dados é o CRISP-DM (*Cross Industry Standard Process for Data Mining*), proposto por Wirth e Hipp (2000). O CRISP-DM é um processo padrão aplicado em diversos setores da indústria, oferecendo uma metodologia robusta e sistemática para transformar dados em conhecimento e gerar informações valiosas para o gerenciamento organizacional (Figura 2).

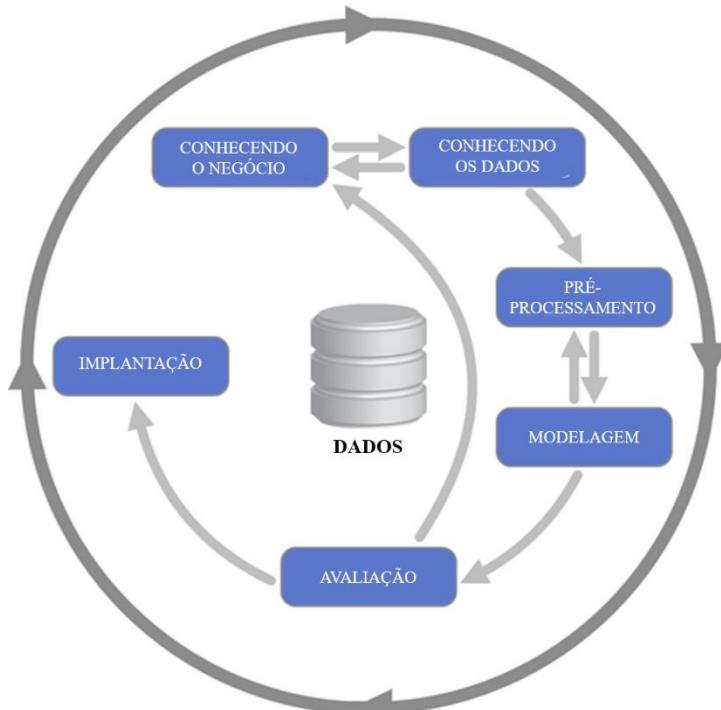


Figura 2 - *Framework* CRISP-DM. Fonte: Wirth e Hipp (2020).

No CRISP-DM, cada fase possui um propósito específico e interage com as demais, formando um ciclo contínuo que se assemelha aos processos utilizados na Administração. Com base nessa similaridade, o *framework* foi adaptado para o contexto administrativo, onde a fase "Conhecendo o Negócio" foi renomeada para "Conhecendo o Ambiente", e a fase de "Implantação" foi subdividida em duas etapas adicionais: "Relatório de Gestão" e "Retroalimentação", como apresentado na Figura 3.

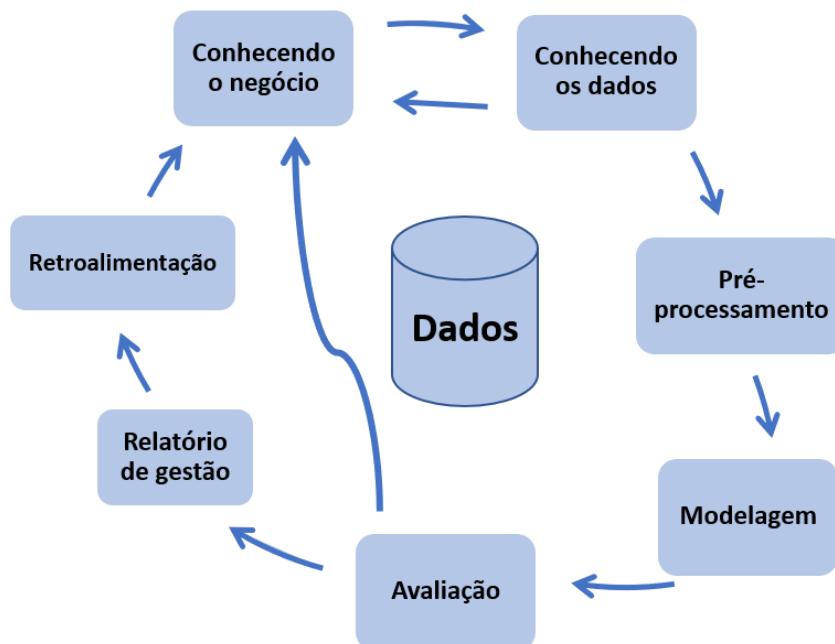


Figura 3 - Fases do *framework* CRISP-DM. Fonte: Wirth e Hipp (2020).

As definições resumidas de cada fase são: a) Conhecendo o Ambiente: coleta de informações sobre o contexto estudado, compreensão das necessidades, definição dos objetivos

e dos critérios de sucesso; b) Conhecendo os Dados: organização e estruturação dos dados disponíveis; c) Pré-processamento dos Dados: preparação das bases de dados e seleção das técnicas de mineração mais adequadas; d) Modelagem: aplicação das técnicas de mineração de dados de acordo com os objetivos estabelecidos; e) Avaliação dos Resultados: análise e interpretação dos resultados obtidos, preparando o processo para a tomada de decisão; f) Relatório de Gestão: organização e apresentação dos resultados em um relatório de gestão, servindo de base para o planejamento de ações futuras; g) Retroalimentação: elaboração de um plano de ação para solucionar fragilidades identificadas nos dados, acompanhando o progresso e promovendo melhorias contínuas.

A adaptação do *framework* CRISP-DM para o contexto da Administração Pública demonstrou ser uma abordagem eficaz para lidar com a complexidade e descentralização dos dados nas instituições de ensino superior. Ao modificar fases específicas e introduzir etapas como "Relatório de Gestão" e "Retroalimentação", torna-se possível uma análise mais detalhada e um acompanhamento contínuo dos resultados, facilitando o processo de tomada de decisão (Fialho, 2020).

2.3 Mineração de Dados Educacionais

Nesta seção serão abordados trabalhos que estudaram sobre mineração de dados educacionais e os tópicos importantes sobre avaliação de desempenho de alunos para efetuar comparativo com dados abertos governamentais disponíveis no Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e dados disponíveis na instituição estudada.

Os trabalhos acadêmicos na área mostram uma maior utilização da Mineração de Dados Educacionais (MDE) voltados para estudo na modalidade de ensino a distância como no estudo de Maschio et al. (2018), que buscou resultados de quais técnicas mais utilizadas; o que é mais investigado; e os dados mais relevantes para a pesquisa. As técnicas mais utilizadas foram de aprendizagem de máquina e agrupamento, a investigação mais realizada foi a de desempenho dos estudantes e os dados mais relevantes foram os números de interações entre alunos e professores que se dão nas avaliações, fóruns, chats, e outras formas de avaliação quantitativa. Este estudo colabora com a introdução sobre mineração de dados na seção anterior, pois reforça a utilização de algoritmos de aprendizagem de máquina e a hipótese de investigar o desempenho dos alunos como uma das aplicações da mineração de dados.

A MDE é considerada uma área recente, trata em desenvolver métodos para explorar conjuntos de dados coletados em ambientes educacionais e atualmente forte e consolidada linha de pesquisa que possui grande potencial para melhorar a qualidade do ensino (Bake et al., 2011). Os métodos utilizados na MDE são originalmente da área de mineração de dados, porém, Bake et al. (2011) afirma que muitas vezes eles precisam ser modificados, porque é preciso considerar a hierarquia dos níveis da informação. Isso é necessário devido à falta de independência estatística nos tipos de dados encontrados em ambientes educacionais.

Neste estudo os principais métodos e modelos de mineração de dados mais aplicados a este estudo de acordo com o que é proposto nos objetivos. Segundo Bake et al. (2011), citando categorias dos métodos de mineração de dados estudados por Moore (2009) os métodos: Predição, Agrupamento e Mineração de relações são de interesse tanto da área de MDE quanto da mineração de dados.

O método de predição, utilizando classificação e regressão são algoritmos mais populares na MDE e geram informações sobre o construto examinado, como em curvas de aprendizagem, e podem predizer os benefícios educacionais para um conjunto de estudantes. Caso o número de dados seja muito extenso, o modelo pode ser construído por uma parte dos dados, auxiliar no desenvolvimento e estimar os benefícios educacionais previamente sendo aplicadas com os alunos.

No agrupamento, o princípio é encontrar dados que se agrupem naturalmente, desconhecidos inicialmente e classificando por grupos e/ou categorias diferentes. No caso de dados educacionais, é possível agrupar alunos por diferenças e similaridades de desempenho ou comportamentos. E em mineração, o objetivo é encontrar relações possíveis entre variáveis dos dados. Pode-se aprender quais variáveis se associam fortemente com uma determinada variável isolada. Estas identificações de relações podem ser de quatro tipos: regras de associação, correlações, sequências ou causas.

3 Metodologia

A metodologia adotada nesta pesquisa, buscou construir conhecimento para a compreensão da realidade, utilizando dados e análises de forma precisa e objetiva. A presente pesquisa adota uma posição epistemológica positivista, caracterizada pela busca de conhecimento objetivo e verificável, com base em evidências empíricas (Creswell, 2014). O positivismo, como corrente epistemológica, enfatiza a mensuração e a análise de dados quantitativos, visando identificar padrões e relações que possam ser generalizados (Guba & Lincoln, 1994).

Em relação aos fins, o estudo é caracterizado como descritivo, pois visa descrever as características do fenômeno em questão (Gil, 2017). De acordo com Cervo et al. (2007), este estudo pode ser classificado como: (i) de abordagem qualitativa e quantitativa, (ii) de natureza aplicada e (iii) de caráter descritivo, sendo realizado por meio de (iv) pesquisa bibliográfica e (v) estudo de caso (Moore, 2009; Brasil, 2006; Silveira & Córdova, 2009; Yin, 2015). A abordagem qualitativa busca descrever, compreender o objeto de pesquisa (Maschio et al., 2018). O caráter descritivo concerne às pesquisas cujo objetivo seja descrever fenômenos em um contexto (Moore, 2009; Brasil, 2006), neste trabalho é a internacionalização dos estudantes da UFPE.

Neste sentido, esta pesquisa obteve os dados utilizados para a mineração no setor de relações internacionais por meio de planilhas onde são registrados os dados dos alunos que participaram de mobilidade estudantil, dados solicitados ao NTI-UFPE (Núcleo de Tecnologia) de discentes matriculados na instituição constando dados de alunos com desempenho acadêmico e dados do Censo da educação superior, disponíveis no portal do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira). Foi feito para Centro da UFPE, a saber: CTG – Centro de Tecnologia e Geociências; CAC – Centro de Artes e Comunicação; CCB – Centro de Ciências Biológicas; CCJ - Centro de Ciências Jurídicas; CA - Centro de Artes; CCS - Centro de Ciências da Saúde; CFCH- Centro de Filosofia e Ciências Humanas; CIN - Centro de Informática; CCN- Centro de Ciências Naturais.

O estudo de caso teve início no setor de Diretoria de Relações Internacionais da instituição, no qual os primeiros dados foram coletados e unidos a dados coletados junto ao Núcleo de Tecnologia de Informações. As duas bases continham 2432 e 42.280 alunos, respectivamente. Houve a aplicação de regras no processo de mineração, são eles: Regra 1: Se aluno não teve reprovações, carga horária do curso menor ou igual; 3345 (provável conclusão do curso), pertencente ao CTG, com desempenho acadêmico ótimo ingressaram no mestrado e/ou doutorado. Regra 2: Se aluno não teve reprovações, carga horária do curso menor ou igual a 3345 (provável conclusão do curso), pertencente ao CTG e total de faltas menor ou igual a 44 então seu desempenho acadêmico foi bom. Regra 3: Se aluno não teve reprovações, carga horária do curso entre 525 e 3345 horas, pertencente ao CAC e teve aprovações menor ou igual a 45 então seu desempenho acadêmico foi ótimo. Regra 4: Se aluno não teve reprovações, carga horária do curso entre 525 e 3345 horas, pertencente ao CCB e teve aprovações menor ou igual a 39 então seu desempenho acadêmico foi bom. Regra 5: Se aluno não teve reprovações, carga horária do curso entre 525 e 3345 horas, pertencente ao CCJ e teve faltas menor ou igual a 26

então seu desempenho acadêmico foi ótimo. Regra 6: Se aluno não teve reprovações, carga horária do curso entre 2040 e 3345 horas, pertencente ao CCJ, teve mais de 26 faltas, não integralizou o curso entre 1 a 4 anos após a mobilidade e é do sexo feminino então seu desempenho acadêmico foi ótimo. Regra 7: Se aluno teve entre 5 e 7 reprovações, aprovações menor ou igual a 81, pertencente ao CAC e não integralizou o curso 7 anos após a mobilidade então seu desempenho acadêmico foi bom. Regra 8: Se aluno teve entre 5 e 7 reprovações, aprovações menor ou igual a 81, pertencente ao CA então seu desempenho acadêmico foi bom. Regra 9: Se aluno teve mais de 12 reprovações, aprovações menor ou igual a 81 e pertencente ao CTG então seu desempenho acadêmico foi regular. Regra 10: Se aluno teve mais de 12 reprovações, aprovações menor ou igual a 81 e pertencente ao CA então seu desempenho acadêmico foi regular. Regra 11: Se aluno teve carga horária do curso entre 3503 e 3922, aprovações entre 57 e 63 então pertencente ao CTG. Regra 12: Se a aluna é do sexo feminino, pertencente ao CAC então é natural da região metropolitana de Recife. Regra 13: Se aluno teve entre 57 e 63 aprovações e desempenho acadêmico bom então são do CTG. Regra 14: Se aluno é do CAC então é natural da região metropolitana de Recife (369 de ocorrência com 88% de confiança). Regra 15: Se aluna do sexo feminino tinha idade de 23 anos quando fez mobilidade então é natural da região metropolitana de Recife.

Por meio do uso do programa *SPSS Statistic* foi feito o cruzamento e agrupamento (*Clusters*) dos dados dos alunos que participam de mobilidade e os que não participam, agrupando em categorias: mobilidade e desempenho e os microdados: média geral, faltas, reprovações e percentual integralizado. O uso do *SPSS Statistics* e do Weka foi escolhido devido à sua robustez e flexibilidade na análise de grandes volumes de dados educacionais.

O *SPSS Statistics* permite realizar análises estatísticas detalhadas e aplicação de técnicas de clusterização, facilitando a segmentação dos dados e a identificação de padrões ocultos. Por sua vez, o Weka é amplamente utilizado para mineração de dados e aprendizado de máquina, oferecendo uma variedade de algoritmos e ferramentas gráficas para visualização, o que facilita a interpretação dos resultados. A combinação dessas ferramentas proporcionou uma análise mais precisa e eficiente, adequada às necessidades específicas deste estudo.

A aplicação das técnicas de mineração de dados, integradas ao *framework* adaptado CRISP-DM, possibilita uma análise aprofundada dos padrões e relacionamentos ocultos nos dados dos alunos de mobilidade internacional. Ao identificar perfis de desempenho acadêmico e padrões de participação, os resultados obtidos fornecem uma base sólida para apoiar decisões estratégicas na gestão da mobilidade.

Isso inclui, por exemplo, a otimização dos critérios de seleção dos alunos e a alocação de recursos para programas de internacionalização, resultando em políticas mais eficazes e alinhadas às necessidades institucionais. Dessa forma, a metodologia aplicada não apenas contribui para uma compreensão detalhada dos dados, mas também oferece insights práticos para a melhoria contínua dos processos de tomada de decisão. Nos dados conseguidos pelo NTI foi feita uma “filtragem” dos alunos e considerar apenas os alunos com percentual de integração acima de 30% por ser um dos critérios utilizados para os estudantes participarem da mobilidade e não gerar dados inconsistentes.

Outro critério utilizado foram quais alunos que participaram do programa Ciência sem Fronteiras (CSF) em separado, por ter sido um programa com influência relevante podendo gerar dados que não condizem com a realidade. Este critério é necessário porque o programa foi interrompido em 2016, tendo a participação de aproximadamente 1900 alunos da instituição no período de 2012 a 2016 e o programa chegou a elevar a média de participação de alunos em mobilidade acadêmica em mais de 4 vezes (dados calculados referente ao ano de 2014): Participante mobilidade de 2014 = 853 e a média de participantes de mobilidade dos cinco anos antes do CSF foi de 200.

4 Análise dos resultados e Discussões

Após análise descritiva das duas bases obteve as seguintes informações comparando-as. A média geral é maior ($7,99 > 7,42$), 7,7% maior, para os alunos que fizeram mobilidade e tem valor mínimo de seis. E, a média de faltas dos alunos (75,68). Segue abaixo, o Quadro 1 quadro com a compilação da interpretação das regras e padrões encontrados na mineração de dados separados por centros de ensinos, experimentos, regras e seus pontos positivos e negativos por cada centro da UFPE.

Centros	Regras	Pontos mais positivos	Pontos menos negativos
CTG	1, 2, 9, 11 e 13	Alunos com carga horária ≤ 3345 e com desempenho acadêmico ótimo ingressaram no mestrado e/ou doutorado.	Alunos que tiveram mais de 12 reprovações tiveram desempenho acadêmico regular.
	1, 2, 3 e 4		
CAC	3, 12 e 14	90% são do sexo feminino e natural da Região Metropolitana do Recife.	Alunos com desempenho acadêmico bom, com reprovações entre 5 e 7 componentes, mas ainda não integralizaram o curso após a mobilidade.
	5, 6 e 7		
CCB	4	Alunos que não tiveram reprovações tem desempenho acadêmico bom.	95,5% dos alunos que fizeram mobilidade neste centro foram pelo CSF. Apenas 4,5% fizeram mobilidade por outros programas.
	13		
CCJ	5 e 6	Alunos que não tiveram reprovações tiveram desempenho acadêmico ótimo.	-
CA	8 e 10	Alunos que tiveram entre 5 e 7 reprovações obtiveram desempenho acadêmico bom.	Alunos que tiveram mais de 12 reprovações obtiveram desempenho acadêmico regular.
	11		
CCS	8	94,33% dos alunos que fizeram mobilidade foram pelo programa CSF.	Apenas 5,66% dos alunos fizeram mobilidade por outros programas de mobilidade.
CFCH	9 e 10	Nenhum aluno que fez mobilidade na Espanha teve desempenho acadêmico regular ou insuficiente.	Nenhum aluno deste centro fez mobilidade nos países: Inglaterra, Austrália e Irlanda.
CIn	12	93,3% dos alunos que fizeram mobilidade foram pelo programa CSF.	Apenas 6,7% dos alunos fizeram mobilidade por outros programas de mobilidade



Centros	Regras	Pontos mais positivos	Pontos menos negativos
CCEN	4	91,7% dos alunos que fizeram mobilidade foram pelo programa CSF.	Apenas 8,3% dos alunos fizeram mobilidade por outros programas de mobilidade.

Quadro 1 - Compilação dos dados.

O último experimento foi aplicado um algoritmo de agrupamento, dividindo a base de dados em três grupos, a fim de encontrar características distintas entre os alunos de forma automática. Através de análise descritiva com ajuda das ferramentas gráficas e 14 estatísticas do *software* Weka (ferramenta para mineração de dados), foi possível identificar características demonstradas na Figura 4 abaixo.

Na Figura 4 pode-se observar que, houve uma clara intercessão entre os grupos 01 e 02 com os atributos “Naturalidade” e “Programa”. E o grupo 03 ficou separado por alunos com naturalidade do interior de Pernambuco, que participaram do Programa Ciência sem Fronteiras (CSF) e do sexo masculino.

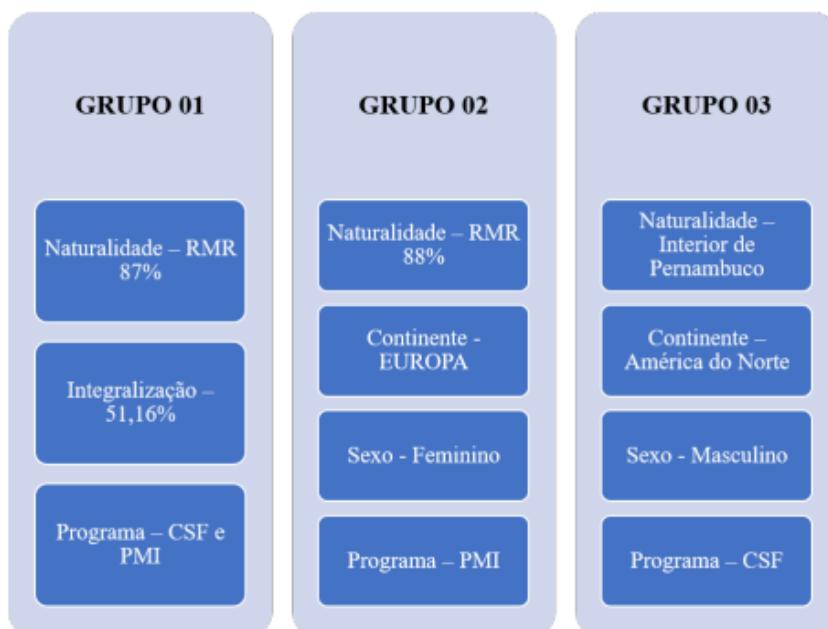


Figura 4 - Cluster os dados. Legenda: PMI – Programa de Mobilidade internacional. CSF - Programa Ciência sem fronteiras.

Análise descritiva do grupo 01, verificando a fundo os dados do grupo 01 foi possível separar suas principais características nos atributos: naturalidade, integralização após a mobilidade e programas. A naturalidade é de 87,41% da região metropolitana de Recife, ou seja, 1090 alunos do grupo. 51,16% dos alunos do grupo integralizaram o curso, enquanto os grupos dois e três tiveram um percentual bem menor, 21,72% e 21% respectivamente. O programa CSF representou 75,7% dos alunos deste grupo e o programa de Mobilidade Internacional (PMI) 12,91%.

Analizando melhor os dados do grupo 02 foi possível separar suas principais características nos atributos: continente, naturalidade, sexo, total de reprovações, programa e desempenho. O continente europeu representa 83,51% dos alunos neste grupo, enquanto os grupos um e três tiveram resultados 64,88% e 8,33% respectivamente. A naturalidade dos alunos foi 88,27% para região metropolitana de Recife. Predominou o sexo feminino com 83,62% dos alunos neste grupo. Apesar de apresentar o valor máximo do total de reprovações

maior que os outros grupos, 51 reprovações, obteve a menor média de reprovações em 2,582 reprovações. Neste grupo foi o maior percentual de alunos no programa PMI com 29,62%, bem maior que os grupos 01 e 03, 12,91% e 3,09% respectivamente. O desempenho dos alunos foi de 33% neste grupo, sendo o maior percentual para desempenho ótimo. Grupo 01 foi 24,3% e o grupo 03 14,2%.

Já a análise descritiva do grupo 03, o continente América do Norte representou 79,63% dos alunos deste grupo, sendo bem mais alto que os grupos um e dois que apresentaram 29,35% e 8,94% respectivamente. O grupo ficou representado por 83,33% de homens. Grupo com maior percentual de alunos, 48,46%, com a naturalidade interior de Pernambuco. Grupo 01 obteve 6,33% e grupo 02 4,53%. 63,27% dos alunos deste grupo não integralizaram o curso 5 anos ou mais após a mobilidade internacional. Sendo 52,2% destes não integralizaram após 6 anos. Este grupo obteve menor valor máximo, 676, para total de faltas, 853 para grupo 01 e 864 o grupo 02. A maior média de faltas 86,728, sendo 74,804 do grupo 01 e 72,798 do grupo 02.

5 Considerações finais

Esta pesquisa teve como objetivo propor um modelo de suporte à tomada de decisão, utilizando técnicas de mineração de dados para analisar informações de alunos de graduação participantes de programas de mobilidade internacional da Universidade Federal de Pernambuco (UFPE). Foram adotados como referenciais o processo KDD (Frawley et al., 1992; Fayyad et al., 1996) e o framework CRISP-DM (Wirth & Hipp, 2000), adaptado ao contexto da instituição.

A aplicação das técnicas permitiu identificar padrões relacionados ao desempenho acadêmico e à participação nos programas, bem como agrupar perfis de alunos com características comuns. Essas informações podem ser utilizadas para revisar critérios de seleção, planejar recursos e acompanhar a execução das políticas de mobilidade. O uso conjunto do SPSS Statistics e do Weka possibilitou a organização, o processamento e a análise das bases de dados, com geração de regras e agrupamentos. A metodologia pode ser replicada em outros contextos institucionais para o tratamento de dados descentralizados.

Entre as limitações do estudo, destacam-se a dependência da consistência das bases analisadas e a exclusão de registros vinculados ao programa Ciência sem Fronteiras. Como continuidade, recomenda-se a ampliação das variáveis analisadas e a avaliação dos impactos das medidas adotadas a partir dos padrões identificados.

O estudo contribui para a gestão institucional ao transformar dados brutos e descentralizados em informações acionáveis, apoiando decisões sobre seleção de alunos, alocação de recursos e acompanhamento de resultados. Do ponto de vista metodológico, demonstra a viabilidade da aplicação da mineração de dados educacionais na avaliação de políticas de internacionalização, oferecendo um modelo adaptável a outras instituições de ensino superior. Assim, a pesquisa reforça a importância de integrar análise de dados e gestão acadêmica para subsidiar decisões baseadas em evidências.

Referências

Altbach, P., & Knight, J. (2007). The Internationalization of Higher Education: motivations and Realities. *Journal of Studies in International Education*, n.3, v. 4, p. 290–305.

Amaral, F. (2016). *Aprenda mineração de dados: teoria e prática*. 1^a ed. Rio de Janeiro, Atlas Books.

Baker, R., Isotani, S., & Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o brasil. *Revista Brasileira de informática na educação*, v.19, n.02 p.3–13.

Brasil. (2006). *Decreto nº 5.8000, de 08 de junho de 2006*. Dispõe sobre o Sistema Universidade Aberta do Brasil - UAB. Retrieved March, 10, Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/decreto/d5800.htm. Acesso em: 12 jan. 2025.

Cervo, A.L., Bervian, P.A., & Silva, R. (2007). *Metodologia científica*. São Paulo, Pearson Prentice Hall.

Creswell, J.W. (2014). *Investigação qualitativa e projeto de pesquisa: escolhendo entre cinco abordagens*. Porto Alegre, Penso Editora.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, v.17, n.3, p.37–37. DOI: <https://doi.org/10.1609/aimag.v17i3.1230>

Frawley, W.J., Piatetsky-Shapiro, G., & Matheus, C.J. (1992). Knowledge discovery in databases: An overview. *AI magazine*, v.13, n.3, p.57–57. DOI: <https://doi.org/10.1609/aimag.v13i3.1011>

Fialho, A.M.G. (2020). *Mineração de dados educacionais dos alunos de graduação da UFPE: um estudo de caso da mobilidade e internacionalização*. Universidade Federal de Pernambuco, Recife.

Gacel-Avila, J. (2005). The Internationalisation of Higher Education: A Paradigm for Global Citizenship. *Higher Education*, v.9, n.2, p.121–136.

Gil, A.C. (2017). *Como elaborar projetos de pesquisa*. 6. ed. São Paulo, Atlas.

Guba, E.G., & Lincoln, Y.S. (1994). Competing paradigms in qualitative research. In Norman K. Denzin and Yvonna S. Lincoln (Eds.). *Handbook of qualitative research*. Thousand Oaks, Sage Publications, p.105–117.

Luce, M.B., Faundes, C.V., & Mediel, O.G. (2016). Internacionalização da educação superior: a dimensão intercultural e o suporte institucional na avaliação da mobilidade acadêmica. *Avaliação*, v.21, n.2, p. 317–340. DOI: <https://doi.org/10.1590/S1414-40772016000200002>

Maschio, P., Vieira, M.A., Costa, N., Melo, S., & Pereira Júnior, C. (2018). Um Panorama acerca da Mineração de Dados Educacionais no Brasil. In *SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO – SBIE/ Congresso Brasileiro de Informática na Educação – CBIE*, p.1936–1940.

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of "unstructured" decision processes. *Administrative science quarterly*, v.21, p. 246–275.

Moore, A. (2009). *Statistical Data Mining Tutorials*.

Müller, A.C., & Guido, S. (2016). *Introduction to machine learning with Python*: a guide for data scientists. O'Reilly Media Inc., Sebastopol.

Oliveira, A.L., & Freitas, M.E. (2016). Motivações para Mobilidade Acadêmica Internacional: a Visão de Alunos e Professores Universitários. *Educação em Revista*, v.32, n.3, p. 217–246.

Oliveira, E.C.S. (2023). *Internacionalização do ensino superior e o impacto do intercâmbio na graduação: um estudo de caso da FGV EBAPE*. Fundação Getúlio Vargas, Rio de Janeiro.

Silveira, D.T., & Córdova, F.P. (2009). A pesquisa científica. In Tatiana Engel Gerhardt and Denise Tolfo Silveira. *Métodos de pesquisa*. Porto Alegre, Editora da UFRGS, p.33–44.

Simon, H.A. (1963). *A capacidade de decisão e liderança*. Rio de Janeiro, Fundo de Cultura.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*. p.29–39.

Witten, I.H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record*, v.31, n.1, p.76–77.

Yin, R.K. (2015). *Estudo de Caso: Planejamento e métodos*. Porto Alegre, Bookman Editora.