



UFOP

Universidade Federal
de Ouro Preto

**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Departamento de Computação e Sistemas**

**Análise e caracterização de padrões de
escolha de disciplinas do curso de
Engenharia de Computação do ICEA
para recomendação de matrículas
semestrais**

Matheus Augusto Figueiredo Martins

**TRABALHO DE
CONCLUSÃO DE CURSO**

ORIENTAÇÃO:
Helen de Cassia Sousa da Costa Lima

**Março, 2023
João Monlevade–MG**

Matheus Augusto Figueiredo Martins

**Análise e caracterização de padrões de escolha
de disciplinas do curso de Engenharia de
Computação do ICEA para recomendação de
matrículas semestrais**

Orientador: Helen de Cassia Sousa da Costa Lima

Monografia apresentada ao curso de Engenharia de computação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Universidade Federal de Ouro Preto

João Monlevade

Março de 2023

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

M386a Martins, Matheus Augusto Figueiredo.

Análise e caracterização de padrões de escolha de disciplinas do curso de Engenharia de Computação do ICEA para recomendação de matrículas semestrais. [manuscrito] / Matheus Augusto Figueiredo Martins. - 2023.

61 f.: il.: color., gráf., tab..

Orientadora: Profa. Dra. Helen de Cassia Sousa da Costa Lima Lima. Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Aplicadas. Graduação em Engenharia de Computação .

1. Algoritmos. 2. Engenharia de Computação. 3. Evasão universitária. 4. Frequência universitária. 5. Mineração de dados (Computação). 6. Universidades e faculdades - Currículos. I. Lima, Helen de Cassia Sousa da Costa Lima. II. Universidade Federal de Ouro Preto. III. Título.

CDU 378.016

Bibliotecário(a) Responsável: Flavia Reis - CRB6-2431



FOLHA DE APROVAÇÃO

Matheus Augusto Figueiredo Martins

Análise e caracterização de padrões de escolha de disciplinas do curso de Engenharia de Computação do ICEA para recomendação de matrículas semestrais

Monografia apresentada ao Curso de Engenharia da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Engenharia da Computação

Aprovada em 28 de março de 2023.

Membros da banca

Doutora - Helen de Cassia Sousa da Costa Lima - Orientador(a) - Universidade Federal de Ouro Preto
Doutor - Filipe Nunes Ribeiro - Universidade Federal de Ouro Preto
Doutor - George Henrique Godim da Fonseca - Universidade Federal de Ouro Preto

Helen de Cassia Sousa da Costa Lima, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 30/04/2023.



Documento assinado eletronicamente por **Helen de Cassia Sousa da Costa Lima, PROFESSOR DE MAGISTERIO SUPERIOR**, em 30/04/2023, às 16:54, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0517053** e o código CRC **DB275757**.

Agradecimentos

Agradeço primeiramente a Deus, pois sem a Sua graça não seria capaz de alcançar a conclusão deste trabalho. Meu agradecimento a esta instituição por ter me proporcionado a estrutura necessária para que pudesse crescer academicamente e pessoalmente.

Toda a minha gratidão ao corpo docente e, em especial, a minha orientadora Helen por todos incentivos e apoio tão importantes. Sem sua ajuda e ensino nada disso seria possível.

Aos meus pais Amantino e Maria do Carmo, as minhas irmãs Milza e Maysa por me apoiarem e acreditarem em mim. Aos meus avós, Dico, Tercia e Milza por todo amor e dedicação empregados. A minha amiga e companheira Paola pelo apoio e cumplicidade. Aos tios, primos, familiares e amigos que sempre torceram e caminharam ao meu lado e me fizeram acreditar que tinha a força e as ferramentas necessárias para finalizar este trabalho.

E, por fim, agradeço todas as pessoas que, de alguma forma, foram essenciais para que alcançasse este objetivo com o qual sempre sonhei. É só o começo!

Resumo

Este trabalho apresenta uma metodologia capaz de identificar o perfil de disciplinas cursadas pelos discentes do curso de Engenharia da Computação (EC) do Instituto de Ciências Exatas e Aplicadas (ICEA), através da utilização de tarefas e técnicas de Mineração de Dados. O presente trabalho foi modelado baseado no histórico escolar dos alunos diplomados e evadidos de tal modo a diferir as estratégias de escolha e desempenho entre esses grupos de alunos. Para evidenciar os fatos mencionados, foi utilizado o Algoritmo *Apriori* para a Extração de Regras de Associação baseado nas disciplinas cursadas pelos discentes. Os resultados mostram-se bastante promissores ao identificar de forma sucinta regras capazes de levar os alunos ao sucesso acadêmico bem como a evasão. Para o grupo de diplomados, diferentes regras foram geradas envolvendo as disciplinas, como por exemplo o adiantamento de algumas disciplinas como Fundamentos de Ciência do Ambiente e Compiladores nos períodos finais. Essas regras resultam em uma diversidade maior de caminhos a serem percorridos para atingir o objetivo. Entretanto para os discentes que optaram pela evasão, esse comportamento pode ser observado a partir de quando a reprovação na disciplina Cálculo Diferencial e Integral I implica a reprovação em Geometria Analítica e Álgebra Linear com confiança superior à 50% para os quatro primeiros períodos do curso.

Palavras-chaves: Algoritmo *Apriori*, Mineração de Dados e Regras de Associação.

Abstract

This work presents a methodology capable of identifying the profile of disciplines taken by students of the Engenharia da Computação (EC) course at Instituto de Ciências Exatas e Aplicadas (ICEA), through the use of Data Mining tasks and techniques. The present work was modeled based on the academic history of graduated and dropped out students in such a way as to differ the choice and performance strategies between these groups of students. To highlight the mentioned facts, the *Apriori* Algorithm was used for the Extraction of Rules of Association based on the disciplines studied by the students. The results are very promising when they succinctly identify rules capable of leading students to academics as well as evasion. For the group of graduates, different rules were generated involving disciplines, such as the advancement of some disciplines such as Fundamentals of Environmental Science and Compilers in the final periods. These rules result in a greater diversity of paths to be taken to reach the goal. However, for students who choose to drop out, this behavior can be observed when failure in Differential and Integral Calculus implies failure in Analytical Geometry and Linear Algebra with confidence greater than 50% for the first four periods of the course.

Key-words: Apriori Algorithm, Data Mining and Association Rules.

Lista de ilustrações

Figura 1 – Fases do processo de KDD (adaptado de Fayyad, Piatetsky e Smyth (1996)	17
Figura 2 – Top 5 suportes 1º período diplomados.	30
Figura 3 – Top 5 suportes 1º período evadidos.	32
Figura 4 – Top 5 suportes 2º período diplomados.	34
Figura 5 – Top 5 suportes 2º período evadidos.	35
Figura 6 – Top 5 suportes 3º período diplomados.	37
Figura 7 – Top 5 suportes 3º período evadidos.	38
Figura 8 – Top 5 suportes 4º período diplomados.	40
Figura 9 – Top 5 suportes 4º período evadidos.	41
Figura 10 – Top 5 suportes 5º período diplomados.	43
Figura 11 – Top 5 suportes 6º período diplomados.	45
Figura 12 – Top 5 suportes 7º período diplomados.	47
Figura 13 – Top 5 suportes 8º período diplomados.	48
Figura 14 – Top 5 suportes 9º período diplomados.	49

Lista de tabelas

Tabela 1 – Características gerais da base de dados	23
Tabela 2 – Variáveis de registro	24
Tabela 3 – Variáveis de nota	24
Tabela 4 – Dados modelados para o problema.	25
Tabela 5 – Dados <i>input</i> do <i>Apriori</i>	26
Tabela 6 – Principais resultados do 1º período diplomados.	29
Tabela 7 – Principais resultados do 1º período evadidos.	31
Tabela 8 – Principais resultados do 2º período diplomados.	33
Tabela 9 – Principais resultados do 2º período evadidos.	34
Tabela 10 – Principais resultados do 3º período diplomados.	36
Tabela 11 – Principais resultados do 3º período evadidos.	37
Tabela 12 – Principais resultados do 4º período diplomados.	39
Tabela 13 – Principais resultados do 4º período evadidos.	40
Tabela 14 – Principais resultados do 5º período diplomados.	42
Tabela 15 – Principais resultados do 6º período diplomados.	44
Tabela 16 – Principais resultados do 7º período diplomados.	46
Tabela 17 – Principais resultados do 8º período diplomados.	47
Tabela 18 – Principais resultados do 9º período diplomados.	49

Lista de abreviaturas e siglas

ICEA Instituto de Ciências Exatas e Aplicadas

UFOP Universidade Federal de Ouro Preto

ML *Machine Learning*

KDD *Knowledge Discovery in Databases*

CSV *Comma-Separated Values*

EDM *Educational Data Mining*

EC Engenharia da Computação

LGPD Lei Geral de Proteção de Dados Pessoais

CPF Cadastro de Pessoa Física

UFRN Universidade Federal do Rio Grande do Norte

Sumário

1	INTRODUÇÃO	12
1.1	Objetivos	13
1.1.1	Justificativa	13
1.1.2	Organização do trabalho	14
2	REVISÃO BIBLIOGRÁFICA	15
2.1	Trabalhos Relacionados	15
2.2	Mineração de Dados	16
2.3	Regras de Associação	17
2.4	Algoritmo <i>Apriori</i>	18
2.5	Métricas de Avaliação	19
2.5.1	Suporte	19
2.5.2	Confiança	19
2.5.3	<i>Lift</i>	20
2.6	Considerações Finais	21
3	DESENVOLVIMENTO	22
3.1	Tecnologias Utilizadas	22
3.1.1	Linguagem <i>Python</i>	22
3.1.2	<i>Google Colaboratory</i>	22
3.2	Coleção dos Dados	23
3.2.1	Fonte dos Dados	23
3.2.2	Informações Gerais	23
3.2.3	Tipo das Variáveis	24
3.3	<i>Knowledge Discovery in Databases (KDD)</i>	24
3.3.1	Seleção e Pré-Processamento dos Dados	25
3.3.2	Transformação dos Dados	25
3.3.3	Extração das Regras de Associação	26
3.4	Considerações Finais	27
4	ANÁLISE DOS RESULTADOS	28
4.1	1º Período	28
4.1.1	1º Período Evadidos	30
4.2	2º Período	32
4.2.1	2º Período Evadidos	34
4.3	3º Período	35

4.3.1	3º Período Evadidos	37
4.4	4º Período	38
4.4.1	4º Período Evadidos	40
4.5	5º Período	41
4.6	6º Período	43
4.7	7º Período	45
4.8	8º Período	47
4.9	9º Período	48
4.10	10º Período	50
4.11	Considerações Finais	50
5	CONCLUSÃO	51
	REFERÊNCIAS	53
	ANEXOS	55

1 Introdução

A Mineração de Dados é uma área da Ciência da Computação consolidada pela diversidade de suas aplicações. No domínio acadêmico essa metodologia é aplicada desde 2005, após ser organizado o primeiro *Workshop on Educational Data Mining* para trazer este tema de forma aberta para discussão (COSTA et al., 2012). A partir de então, inúmeros trabalhos foram propostos utilizando as técnicas de Mineração de Dados para o contexto educacional, essa abordagem é nomeada de *Educational Data Mining* (EDM).

O volume de dados e as tecnologias de processamento e análise de dados estão em constante crescimento e evolução, segundo Romero Cristóbal e Ventura (2014) este fato facilita o processo de manipulação e transformação dos dados de forma ágil e com maior qualidade. Essa agilidade permite assegurar uma tomada de decisão mais assertiva sobre uma população e cenário específico. Por isso é importante avaliar não só os dados, mas também o contexto em que eles estão compreendidos, de modo a respeitar a particularidade de cada problema.

O Instituto de Ciências Exatas e Aplicadas (ICEA) é um campus da Universidade Federal de Ouro Preto (UFOP) que dentre os cursos ofertados estão o de Engenharia da Computação (EC). Este curso é ofertado no campus desde 2009 e contém 1.022 alunos que já foram atendidos pelo curso até 2020. Entretanto, apenas 8.5% conseguiram alcançar o sucesso nesse intervalo, ou seja, obter a diplomação, se comparar com a quantidade de evasão, essa estatística chega a ser assustadora, onde se passa pouco mais dos 60%. Inúmeros são os impactos desses números para a comunidade acadêmica, a sociedade e a economia. Dentre essas consequências, apresentada por Mussliner et al. (2021) estão a diminuição de investimento na área, a incapacitação de profissionais para o mercado de trabalho e conseqüentemente, o atraso no desenvolvimento da economia, além de aspectos emocionais para os discentes. O autor Hoed (2016) traz uma simples reflexão sobre este assunto que aborda tais aspectos listados:

“Assim, para o discente, começar e não terminar um curso de graduação pode acarretar uma frustração profissional que o acompanhará por toda a vida.”

Deste modo, para minimizar os impactos negativos de um insucesso acadêmico pode-se alinhar as técnicas de Mineração de Dados junto aos dados produzidos pelos próprios alunos do Instituto. Existem diferentes maneiras de unir esses dois domínios para se chegar às soluções capazes de beneficiar os discentes utilizando o processo de *Knowledge Discovery in Databases* (KDD) que através de uma análise profunda de uma respectiva

base de dados para extrair conhecimentos e padrões escondidos não observados por análises manuais. Essa inobservação pode se devido à complexidade da manipulação de dados devido à diversidade de atributos e ao volume de dados e informação. Para o contexto do Instituto de Ciências Exatas e Aplicadas (ICEA) já foi utilizado algumas tarefas de Mineração de Dados como, por exemplo, a Análise de Componentes Principais (CALDEIRA, 2021) e Classificação (GONÇALVES, 2022) e (RODRIGUES, 2022), entretanto, esse trabalho aborda a Extração de Regras de Associação.

1.1 Objetivos

O objetivo deste trabalho é apresentar para a comunidade acadêmica principalmente os envolvidos no curso de Engenharia da Computação (EC) do Instituto de Ciências Exatas e Aplicadas (ICEA) a escolha das disciplinas dos alunos de tal modo a analisar essa tomada de decisão em correspondência à matriz curricular atual, por meio de Extração de Regras de Associação. Essas regras são geradas a partir da análise das disciplinas cursadas pelos discentes do curso de EC. Esse estudo é fundamentado na análise dos dados históricos de alunos que obtiveram o sucesso acadêmico, isto é, os discentes diplomados. Dessa, forma têm-se objetivos específicos:

- Implementar o processo de *Knowledge Discovery in Databases* (KDD) nos dados históricos dos discentes.
- Modelar os dados de tal modo a ser padronizado para *input* da tarefa de Mineração de Dados.
- Utilizar uma técnica de Mineração para extrair regras e padrões do conjunto de dados.
- Avaliar os resultados com base no conhecimento prévio pelas observações da base de dados e do contexto do Instituto de Ciências Exatas e Aplicadas (ICEA).

1.1.1 Justificativa

A preocupação com o desempenho dos discentes é retratada na literatura por Filho et al. (2007) como sendo de suma importância para o desenvolvimento do Brasil. O autor explica que quanto antes for possível entrar com medidas de melhoria no desempenho do aluno na universidade, é plausível incrementar a qualidade do ensino e consequentemente reduzir as hipóteses do fracasso acadêmico.

Como pode se observar no início do Capítulo 1 somente para o curso de EC temos um número muito baixo de alunos conseguiram concluir a graduação. Este já é um problema conhecido e nos últimos anos ganhou espaço no Instituto com trabalhos que buscaram

entender e amenizar o impacto do insucesso para a sociedade e os discentes (CALDEIRA, 2021; GONÇALVES, 2022; RODRIGUES, 2022). Ainda assim, como complemento explorar outras áreas além das passadas para o contexto do curso no instituto, a fim de desenvolver análises relevantes das disciplinas tal modo a impulsionar o sucesso acadêmico explorando as inúmeras tarefas de Mineração de Dados.

1.1.2 Organização do trabalho

O restante deste trabalho está estruturado da seguinte forma, o [Capítulo 2](#) trás uma análise das principais bibliografias utilizadas para a realização do estudo. Seguido pelo [Capítulo 3](#) que contém as etapas do trabalho bem como a explicação das técnicas e ferramentas utilizadas. Já no [Capítulo 4](#) é feito o detalhamento dos padrões encontrados bem como a análise dos mesmos. Por fim, no [Capítulo 5](#) é feita a conclusão dos resultados alcançados.

2 Revisão Bibliográfica

Este capítulo faz a estruturação dos principais temas abordado no *Educational Data Mining* (EDM) e no *Knowledge Discovery in Databases* (KDD). Igualmente, discorre em detalhes cada etapa de ambos processos. Além disso, avaliação dos trabalhos para o contexto específico do Instituto de Ciências Exatas e Aplicadas (ICEA) e dos principais trabalhos correlatos na literatura.

2.1 Trabalhos Relacionados

A utilização do processo *Knowledge Discovery in Databases* (KDD) no âmbito educacional é relativamente recente conforme destaca Manhães et al. (2012). No entanto, a maioria dos trabalhos relacionados se fundamenta a utilizar tarefas preditivas de evasão ou desempenho escolar. Quando são utilizadas as tarefas de Mineração de Dados, os trabalhos se baseiam no agrupamento, segmentação dos grupos ou predição, principalmente com a visão relacionada ao estado de evasão. Deste modo, explorar outros aspectos presentes nas tarefas de Mineração de Dados pode gerar novos *insights* e complementar as soluções de alguns problemas.

A adoção do *Educational Data Mining* (EDM) ainda não é tão consolidado, mas essa realidade está mudando e cada vez mais novos trabalhos surgem neste contexto. Um fato importante de se destacar é a complexidade do assunto e a quantidade de atributos que podem ser utilizados neste âmbito, bem como as técnicas utilizadas para extração de conhecimento. Os autores Rodrigues et al. (2018) levantam quatro aspectos importantes a serem detalhados ao utilizar o processo de KDD.

- transformação dos dados (os dados colhidos nem sempre são diretamente tratados pelos algoritmos de mineração);
- identificar os atributos mais relevantes;
- identificar os algoritmos mais adequados;
- aplicar os algoritmos para identificar outros grupos de estudantes;

Foi realizado um estudo a respeito da evasão no curso de Sistemas de Informação na Universidade Federal do Rio Grande do Norte (UFRN) pelo autor Junior (2018), utilizando algumas técnicas de Mineração de Dados. O autor utilizou o algoritmo de agrupamento *k-means* e o algoritmo de classificação J48 para identificar a evasão dos discentes. A partir das análises realizadas, os autores identificaram alguns aspectos que levaram os alunos à

evasão, como, por exemplo, a reprovação nas disciplinas-base do curso, principalmente na de Fundamentos e Algoritmos.

No contexto do Instituto de Ciências Exatas e Aplicadas (ICEA) alguns trabalhos foram realizados utilizando os dados fornecidos pela própria universidade. O autor Caldeira (2021) utilizou as técnicas de clusterização de *Principal component analysis*, *K-means*, *Nearest Neighbors* para caracterizar o problema da evasão no ICEA. Ele extraiu características nestes grupo de alunos e a principal encontrada, foi o baixo rendimento escolar, ligado principalmente aos primeiros períodos dos cursos. Para o mesmo contexto, o autor Gonçalves (2022) e o autor Rodrigues (2022) abordaram o assunto de desempenho dos discentes do ICEA utilizando técnicas de classificação, e desenvolveram modelos de Aprendizado de Máquina capaz de prever a evasão, sendo o trabalho do Rodrigues (2022) especificamente para o Curso de Engenharia da Computação.

Percebe-se que as principais referências que abordam o assunto de Mineração de Dados Educacionais tem a preocupação com o desempenho do aluno para fins de previsão de rendimento escolar ou evasão. Entretanto, propor uma solução para facilitar ou induzir uma decisão assertiva para que os discentes obtenham sucesso, possui seus desafios, principalmente na criação de ações preventivas. Todavia, autores como Mondal et al. (2020) e Maphosa, Doorsamy e Paul (2020) já desenvolveram trabalhos de tal modo a propor sistemas capazes de gerar uma recomendação de disciplinas para os discentes, a partir da análise do histórico de discentes avaliando o desempenho acadêmico por meio da seleção de disciplinas que levaram ao sucesso acadêmico.

As técnicas de Extração de Regras de Associação são flexíveis e poderosas, elas podem traçar padrões que o ser humano não foi capaz de identificar anteriormente. Por isso para abordar esse aspecto pouco estudado no contexto do ICEA, este trabalho propõe o estudo das disciplinas que os discentes obtiveram sucesso acadêmico, busca responder qual é o caminho que leva o discente ao sucesso, não somente agrupar por afinidades ou prever o seu desempenho. Aplicar as técnicas de Extração de regras de associação através de um algoritmo de Aprendizado de Máquina, é a melhor opção para termos resultados com maior exatidão num tempo de computação viável.

2.2 Mineração de Dados

A Mineração de Dados é definida na literatura por (WITTEN et al., 2005) como o termo que se refere ao processo de descobrir padrões úteis em uma base de dados. Esse processo é conhecido também como *Knowledge Discovery in Databases* (KDD), com conceito semelhante, Fayyad et al. (1996) classifica o processo do KDD como interativo e iterativo que envolve várias etapas de transformação e análise de dados até se chegar a extração de conhecimento útil. Essas etapas são ilustradas pela Figura 1.

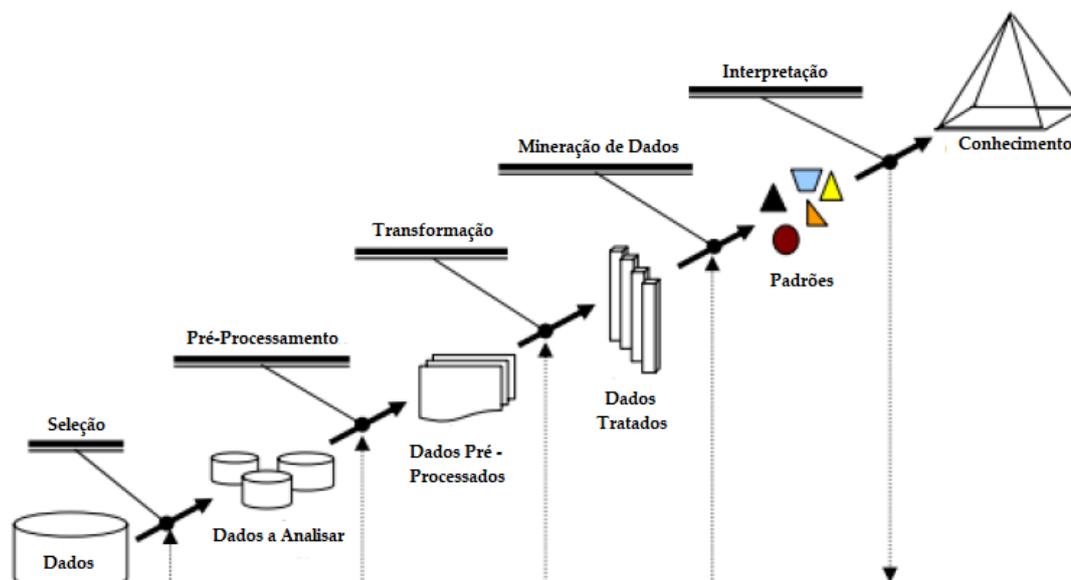


Figura 1 – Fases do processo de KDD (adaptado de Fayyad, Piatetsky e Smyth (1996))

A primeira fase do processo é composta pela Seleção dos atributos que serão utilizados para extração de informação e conhecimento da base de dados. Posteriormente na fase de Pré-Processamento é composta pela remoção de ruídos e dados inconsistentes, tais como *outliers* e dados ausentes. É também nesta etapa que as inconsistências da base de dados é avaliada isto é, a veracidade das informações. Em seguida é feita a etapa de transformação, na qual os dados são agregados por um atributo ou um conjunto de atributos. Desse modo os dados são padronizados para servir de *input* para a próxima etapa, a de Mineração de Dados.

A estimativa é que 80% do tempo desse processo do KDD é gasto entre as etapas de Seleção e de Transformação, por se tratar das fases fundamentais para a consistência de dados de *input* e gerar conhecimento preciso. Caso contrário, os resultados podem não ser uniformes, logo, enviesando-os (HAN; PEI; TONG, 2022). A posteriori na próxima etapa a de Mineração de Dados, finalmente é aplicado a técnica de Mineração de Dados, ou seja, os algoritmos que melhor se encaixam ao problema, para descobrir os padrões ocultos. Para finalizar, a etapa de Interpretação é composta pela análise e validação dos resultados, cujo objetivo central é a validação dos padrões encontrados. Esta etapa também é crítica, pois, envolve o domínio técnico do contexto da aplicação.

2.3 Regras de Associação

A identificação de Regras de Associação é uma tarefa da Mineração de Dados capaz de capturar a frequência de um item ou um conjunto de itens. É também conhecida e caracterizada por Larose (2005) como análise de afinidades, que envolve o estudo dos atributos ou categorias de itens que estão relacionados, ou possuem alguma associação entre

si. A motivação deste tema consiste em descobrir padrões para utilizá-los em recomendação de produtos, disposição de produtos em prateleiras, recomendação de séries e filmes entre outros.

A temática de Extração de Regras de Associação foi introduzida por [Agrawal, Srikant et al. \(1994\)](#), como uma maneira de encontrar padrões repetitivos e associativos entre um ou mais itens ou ações a partir de um conjunto de transações. Sua premissa básica é encontrar elementos que implicam na presença de outros em um universo específico ([SCHNEIDER, 2003](#)).

2.4 Algoritmo *Apriori*

O algoritmo *Apriori*, de acordo com ([ANSELMO, 2017](#)), é o pioneiro e a maior referência dos algoritmos que adotam a estratégia de Extração de Regras de Associação. Ele consegue identificar os padrões presentes em base de dados simples até às mais complexas. Ele baseia-se em um conjunto de transações que contém seus itens associados, cujo objetivo é calcular a frequência desses itens no universo em discussão. Desse modo, é possível realizar a associação de um item com os demais conjuntos de itens frequentes ([ROMÃO et al., 1999](#)).

Em termos gerais as regras de associação buscam padrões do tipo: se os clientes compram um produto X, eles também compram o produto Y? Por isso esse algoritmo foi escolhido para responder à seguinte questão: quais são as disciplinas mais frequentes no conjunto de alunos que obtiveram sucesso acadêmico? Dessa maneira é possível apresentar os resultados para apoiar a decisão dos alunos ao se matricular nas disciplinas.

Para entender melhor como o algoritmo funciona, é importante se ater ao conceito de item frequente. Esse conceito é definido por [Imielinski, Swami e Agarwal \(1993\)](#), como aquele item que aparece em um determinado número de vezes na base de dados. Um item é considerado frequente se satisfaz a seguinte condição: a frequência em que o item aparece no conjunto de dados é maior que um *threshold* especificado previamente, nomeado como Suporte Mínimo.

$$Suporte\{item\} \geq SuporteMínimo$$

No que diz respeito à descoberta de Regras de Associação, as medidas de avaliação detêm um papel fundamental, pois, é a partir dos resultados obtidos com estas medidas que conseguimos analisar qualitativamente as regras geradas pelos algoritmos. Como a técnica busca encontrar padrões escondidos na base de dados, algumas regras podem ser inviáveis para o cenário proposto ou sem sentido para o contexto de aplicação. Por isso é importante utilizar métricas para avaliar a geração dessas regras.

2.5 Métricas de Avaliação

De acordo com [Anselmo \(2017\)](#), as métricas mais utilizadas para auxiliar na avaliação das regras é o suporte mínimo, confiança e o *Lift*. Definir e compreender estes conceitos são importantes tanto na etapa criação das regras quanto a de avaliação dos resultados alcançados. Essas métricas são detalhadas na seguinte ordem: Suporte na [subseção 2.5.1](#), Confiança na [subseção 2.5.2](#) e o *Lift* na [subseção 2.5.3](#).

2.5.1 Suporte

Suporte é uma métrica que mede a frequência ou prevalência de um determinado conjunto de itens em um determinado conjunto de dados ([ANSELMO, 2017](#)). Formalmente, o suporte de um conjunto de itens X é definido como a proporção de transações no conjunto de dados que contém todos os itens de X .

Por exemplo, suponha que temos um conjunto de dados de transações de clientes em uma mercearia e estamos interessados em encontrar conjuntos de itens frequentes que contenham "leite". Se o limite de suporte for definido como 0,5 (50%), o suporte do conjunto de itens {leite} seria a proporção de transações que contém leite, e qualquer conjunto de itens com suporte maior ou igual a 0,5 seria considerado frequente conjunto de itens.

Em geral, a métrica de suporte é usada no algoritmo *Apriori* para identificar conjuntos de itens frequentes que atendem a um limite de suporte mínimo especificado, que é geralmente definido pelo usuário ou com base no conhecimento do domínio. Quanto maior o limite de suporte, mais seletivo será o algoritmo na identificação de conjuntos de itens frequentes e, quanto menor o limite, mais conjuntos de itens serão identificados ([GAMA et al., 2012](#)). Onde N é o número total de observações. Essa métrica é dada pela seguinte fórmula:

$$Suporte(X \Rightarrow Y) = \frac{Frequência(X, Y)}{N}$$

2.5.2 Confiança

No algoritmo *Apriori*, a confiança é uma métrica que mede a força de uma regra de associação entre dois conjuntos de itens, X e Y . É definida como a probabilidade condicional de encontrar o conjunto de itens B em uma transação, dado que o conjunto de itens A também está presente na mesma transação. Formalmente, a confiança de uma regra $X \rightarrow Y$ é definida como:

$$Confiança = \frac{Suporte(X \cup Y)}{Suporte(X)}$$

O numerador refere-se ao número de transações em que X e Y ocorrem simultaneamente. O denominador refere-se à quantidade de transações em que o item X ocorre. Em

termos gerais a confiança mede a probabilidade condicional de ocorrer Y dado que ocorreu X (ANSELMO, 2017).

Por exemplo, suponha que temos um conjunto de dados de transações de clientes em uma mercearia e queremos encontrar regras de associação entre os itens leite e pão. Se descobirmos que o suporte do conjunto de itens leite, pão é 0,2 e o suporte do conjunto de itens leite é 0,4, então a confiança da regra leite \rightarrow pão é 0,5 (ou seja, 0,2 / 0,4) . Isso significa que 50% das transações que contêm leite também contêm pão.

Em geral, a métrica de confiança é usada no algoritmo Apriori para identificar regras de associação fortes que atendem a um limite mínimo de confiança especificado, que é geralmente definido pelo usuário ou com base no conhecimento do domínio. Quanto maior o limite de confiança, mais seletivo será o algoritmo na identificação de regras fortes e, quanto menor o limite, mais regras serão identificadas (GAMA et al., 2012).

2.5.3 Lift

O *Lift* é uma métrica que mede a força de uma regra de associação entre dois conjuntos de itens, X e Y, que leva em consideração a frequência de X e Y no conjunto de dados. O aumento geralmente é usado em combinação com as métricas de suporte e confiança para identificar regras de associação fortes (BRIN et al., 1997). O *Lift* de uma regra $X \rightarrow Y$ é definido como:

$$Lift(X \Rightarrow Y) = \frac{Suporte(X \cup Y)}{Suporte(X) * suporte(Y)}$$

Onde, $suporte(X \cup Y)$ é o suporte do conjunto de itens que contém X e Y, e $suporte(X)$ e $suporte(Y)$ são o suporte dos conjuntos de itens X e Y, respectivamente.

- Se $Lift(X \rightarrow Y) = 1$, então X e Y são independentes.
- Se $Lift(X \rightarrow Y) > 1$, então X e Y positivamente dependentes.
- Se $Lift(X \rightarrow Y) < 1$, então X e Y negativamente dependentes.

Em geral, o *Lift* é usado no algoritmo *Apriori* para identificar regras de associação forte que têm uma elevação maior que 1, indicando que os dois conjuntos de itens estão correlacionados positivamente. No entanto, é importante observar que o *Lift* sozinho pode não ser suficiente para identificar regras fortes e deve ser usado em combinação com outras métricas, como suporte e confiança.

2.6 Considerações Finais

Após realizar uma extensa análise bibliográfica para entender e compreender o que está por trás do processo [KDD](#) e do [EDM](#) é possível fundamentar e validar este trabalho. Assim, identificar o padrão frequente de disciplinas cursadas pelos discentes que obtiveram sucesso acadêmico apresenta uma decisão comum e eficiente entre esses discentes. Ao descobrir a relação dessas disciplinas por semestre, é possível desenvolver um layout onde pode-se traçar um caminho para o sucesso acadêmico, e por conseguinte contribuir para diminuir as taxas de evasão como já abordado por [Gonçalves \(2022\)](#).

3 Desenvolvimento

Neste capítulo são apresentadas as ferramentas utilizadas, os aspectos da base de dados e o detalhamento do processo de *Knowledge Discovery in Databases* (KDD) para o contexto do Instituto de Ciências Exatas e Aplicadas (ICEA) restrito ao curso de Engenharia da Computação (EC).

3.1 Tecnologias Utilizadas

Para realizar a Mineração de Dados, um recurso fundamental são as tecnologias, sejam elas de *hardware* e *software*, bem como domínio sobre o contexto do problema. As principais tecnologias utilizadas são descritas e detalhadas nas próximas subseções.

3.1.1 Linguagem *Python*

O *Python*⁰ é uma linguagem de programação amplamente usada em aplicações da *Web*, desenvolvimento de *software*, ciência de dados e *Machine Learning* (ML). Os desenvolvedores usam o *Python* porque é uma linguagem de alto nível, ou seja, fácil de aprender (FRANZ et al., 2021). Além disso, o *Python* possui uma vasta gama de bibliotecas e pacotes projetados especificamente para mineração e análise de dados, como *Pandas*, *NumPy*, *SciPy*, *Matplotlib* e *Scikit-learn*. Essas bibliotecas fornecem um rico conjunto de ferramentas e algoritmos que podem ser usados para executar tarefas complexas de mineração de dados com eficiência.

3.1.2 *Google Colaboratory*

O *Google Colaboratory*¹ é uma plataforma online que fornece um ambiente para escrever e executar código *Python*, bem como acesso aos recursos de computação em nuvem do *Google*. Essa aplicação funciona diretamente em um navegador, sem a necessidade de nenhum tipo de instalação de *software* em uma máquina. Além disso, é possível desenvolver e rodar códigos em documentos interativos que podem ser compartilhado com outros desenvolvedores, modificá-los a qualquer momento e mantê-los salvos totalmente *online*.

⁰ *Python*

¹ *Google Colaboratory*

3.2 Coleção dos Dados

Dados são uma coleção de fatos, estatísticas ou informações que podem ser analisadas para obter *insights*, tomar decisões informadas e resolver problemas. A análise de dados é crucial para a tomada de decisões informadas porque ajuda a identificar padrões, tendências e relacionamentos que podem não ser imediatamente aparentes. O processo de análise de dados pode revelar *insights* ocultos e fornecer uma compreensão mais profunda de questões complexas.

A seguir é apresentada a descrição da base de dados utilizadas e as suas características. É significativa a explicação detalhada dessa etapa, para entender a construção da lógica de aplicação do algoritmo e dos resultados obtidos por ele.

3.2.1 Fonte dos Dados

Este trabalho conta com duas bases de dados em formato *Comma-Separated Values (CSV)*, disponibilizadas pelo próprio Instituto. Essa base é composta por algumas informações sensíveis e particulares, como, por exemplo, Cadastro de Pessoa Física (CPF) e a matrícula do aluno, que impossibilita a identificação do aluno. Entretanto, conforme à Lei Geral de Proteção de Dados Pessoais (LGPD) para proteger os direitos fundamentais de liberdade e de privacidade de cada indivíduo, esses registros passaram por um processo de *encoding* em que não é possível realizar a identificação dos discentes.

3.2.2 Informações Gerais

As tabelas fornecidas pelo Instituto de Ciências Exatas e Aplicadas (ICEA) contêm os dados referentes aos discentes do Instituto nos âmbitos de diplomados, matriculados e evadidos. Possui dados referentes às notas, que em síntese representa o histórico escolar dos alunos, e outra que inclui os registros dos discentes, que contém elementos que remetem a admissão do aluno.

Para fins de estudo a categoria de alunos matriculados foi removida, assim manteve-se somente os discentes diplomados e evadidos. A Tabela 1 corresponde às características gerais das bases de dados, que contém todos os registros do curso de Engenharia de Computação.

Base de Dados	Quantidade de Variáveis	Nº Diplomados	Nº Evadidos
Registros	49	86	536
Notas	24	86	536

Tabela 1 – Características gerais da base de dados

Fonte: Autoria própria.

3.2.3 Tipo das Variáveis

As variáveis que constam na base são definidas entre categóricas e numéricas. As variáveis categóricas são do tipo *string*, as quais representam os valores de maneira qualitativa. Já as variáveis numéricas, pertencem ao tipo *float* ou *int* e possuem origem quantitativa. O tipo de dado *datetime*, é denominado para registros de valores referente a data e hora. A seguir, nas [Tabela 2](#) e [Tabela 3](#) são apresentadas as variáveis presentes na base conforme sua categoria, para as bases de registro e de notas, respectivamente.

Variáveis	Tipo
data ingresso, data da evasão, primeiro horário, ultimo horário	<i>datetime</i>
aluno, cidade nascimento, código curso, código curso admissão, código enfase, código habilitação admissão, código modo admissão, código situação aluno, código evasão tipo, curso, descrição da disciplina, descrição modo admissão situação aluno, destino, estado nascimento, feito por requerimento, modalidade concorrência, modalidade concorrência homologada, motivo, origem, pais nascimento, participou politica afirmativa, republica, sexo, situação aluno atual, tipo republica, turno, usou politica afirmativa	Catégorica
ano de admissão, ano diplomação, ano evasão, ano de nascimento, caixa arquivo, carga horaria cursada, carga horaria curso, classificação vestibular, código currículo, código currículo admissão, código habilitação, código polo, código polo admissão, CPF id, matrícula id, polo, pontuação vestibular, semestre de admissão, semestre diplomação, semestre evasão	Numérica

Tabela 2 – Variáveis de registro
Fonte: Autoria própria.

Variáveis	Tipo
gravação	<i>datetime</i>
caráter, código curso, código departamento, código disciplina, cor da pele, descrição da disciplina, descrição modo admissão, sexo, situação, tipo escola	Catégorica
ano, ano nascimento, código turma, exame especial, faltas, matrícula id, media final, semestre	Numérica

Tabela 3 – Variáveis de nota
Fonte: Autoria própria.

3.3 Knowledge Discovery in Databases (KDD)

O KDD é uma parte essencial da mineração de dados e envolve vários estágios, incluindo limpeza de dados, integração de dados, seleção de dados, transformação de dados, mineração de dados, avaliação de padrões e representação de conhecimento. Estas etapas são abordadas a seguir, de tal modo a delinear as decisões tomadas para o desenvolvimento deste trabalho.

3.3.1 Seleção e Pré-Processamento dos Dados

A seleção de dados é uma das etapas principais no processo de **KDD** e envolve a seleção de dados relevantes para análise com base no domínio do problema e nos objetivos da mineração de dados. O principal objetivo da seleção de dados é reduzir o tamanho do conjunto de dados removendo dados irrelevantes, isto é reduzir a dimensionalidade dos atributos e focar nos dados mais informativos para a tarefa de mineração de dados.

Como o objetivo é extrair regras baseadas nas disciplinas cursadas somente alguns atributos foram levados em consideração: **matricula id, descrição da disciplina, semestre, ano de ingresso, semestre de ingresso, situação do aluno e situação da disciplina.**

3.3.2 Transformação dos Dados

A entrada de dados para a fase de Mineração de Dados são os dados já agregados por algum atributo e transformados conforme o algoritmo pretendido. Para algoritmos de previsão, os dados têm que estar em uma escala coerente, e para isso são utilizadas técnicas de normalização e padronização. Já para os algoritmos de Extração de Regras de Associação, é necessário somente vetores de zeros e uns. Posto isso, como o objetivo é avaliar as disciplinas, foi utilizado a coluna código da disciplina como atributo fundamental dessa análise.

Para realizar a análise foi necessário criar um atributo, o período em que o aluno cursou a disciplina. Dessa maneira, foi possível descobrir as disciplinas que os alunos cursaram naquele determinado período e gerar as regras por período. Para a coluna que representa o código da disciplina, foi feita a caracterização da situação do aluno naquela disciplina, uma vez que temos duas situações: reprovado ou aprovado. As demais situações de cancelamento ou de trancamento não foram abordadas neste trabalho. Para manter a caracterização da coluna, o caractere A foi adicionado ao final do código da disciplina para referenciar a situação de aprovado e o caractere R foi utilizado para representar a situação de reprovado. Dessa forma, pode-se gerar tanto regras envolvendo as situações de aprovado quanto as de reprovado, aumentando a precisão dos resultados, como, por exemplo, a regra: se aprovado em uma disciplina X, então reprovado na disciplina Y. A [Tabela 4](#) é representada por um exemplo da modelagem desses dados.

Tabela 4 – Dados modelados para o problema.

mat_id	cod_disciplina	período_que_o_aluno_fez_a_disciplina
6288	CEA509_A	7
6550	CSI476_A	6
4614	CSI436_R	2

Fonte: Autoria Própria

Após essa modelagem foi necessário acrescentar uma etapa de transformação, pois, o algoritmo entende somente os vetores de zeros e uns. Então a base foi dividida em 10 períodos diferentes, contendo os dados somente do seu respectivo período, para o grupo de alunos diplomados. Já para o grupo de alunos evadidos, essa divisão foi feita em somente quatro períodos. Essa escolha se deu por a partir do quarto período ser apresentado poucos dados para análise, uma vez que o volume de registros contém uma quantidade insuficiente de disciplinas distintas para realizar o cálculo de suporte mínimo. Cada linha representa um discente único e cada coluna representa uma disciplina daquele conjunto de matérias cursadas no período em questão. Os valores das linhas são representados por zeros e uns, onde, quando o valor foi zero, significa que aquele discente em questão cursou aquela disciplina naquele período. Quando o valor for zero, significa que aquele discente não cursou aquela determinada disciplina naquele período. A [Tabela 5](#) apresenta os dados pós processados, que são a base de entrada para a tarefa de Mineração de Dados.

Tabela 5 – Dados *input* do *Apriori*.

mat_id	CSI466_A	CEA433_A	CEA341_R
2442	0	0	0
2448	0	0	0
2517	0	0	0
2566	0	0	0
2637	0	0	1

Fonte: Autoria Própria

3.3.3 Extração das Regras de Associação

O algoritmo *Apriori* é um método de aprendizado de máquina que trabalha com a ideia de destacar os itens mais frequentes em um conjunto de dados. A partir da contabilidade da frequência desses itens presentes na base de dados, eles são relacionados uns com os outros. A seguir, veja o funcionamento do algoritmo e as parametrizações conforme o contexto do curso se [EC](#).

1. Calcula-se a frequência dos elementos na base de dados que atendam a um suporte mínimo. Para esse conjunto de dados o suporte mínimo escolhido foi de 0.1, isto é, uma disciplina para ser relevante para a regra deverá aparecer em mais de 10% das vezes no conjunto à cada iteração.
2. Gera todas as combinações de itens e calcula a confiança. O *threshold* utilizado para a confiança foi de 0.5, isto significa que para uma regra ser válida ela tem que ocorrer pelo menos em 50% das vezes.
3. Repete o passo 1 até não conseguir gerar mais regras que satisfazem a condição de suporte mínimo e confiança mínima.

Após extração das regras, foi utilizado um ponto de corte para o *lift*, uma vez que somente para valores de $lift \geq 1$. Uma vez que esse valor ser superior à 1 vai representar uma correlação clara entre as disciplinas, caso contrário, não existe correlação clara entre os duas disciplinas.

3.4 Considerações Finais

A organização e estruturação dos dados para a Extração das Regras de Associação é a etapa crucial para obter assertividade nos resultados. Seguidamente pela definição do suporte mínimo e confiança, o contexto testado utilizou uma confiança mínima de 0.50 e o suporte mínimo como 0.10. Após essas definições, as regras foram geradas e apresentadas na próxima seção.

4 Análise dos Resultados

Este capítulo aborda e discute os experimentos com aplicação do algoritmo *Apriori*, o padrão de frequência de disciplinas por semestre dos alunos do curso de Engenharia da Computação (EC) do Instituto de Ciências Exatas e Aplicadas (ICEA). A ideia após extrair e analisar as regras descobertas é auxiliar a tomada de decisão dos discentes ao se matricular nas disciplinas. Além disso, conhecer as características das decisões passadas dos alunos que deram certo e também as que deram errado.

Nas próximas seções serão apresentados as *top 10* regras associadas à cada período do curso de EC ordenadas pela confiança seguida do suporte. Essa escolha se deu diante da complexidade de analisar todos os casos, então fazendo esse corte, em que se elegeu para análise somente as primeiras dez regras, e depois, foram ordenadas pelos seus respectivos níveis de confiança e suporte. Outrossim, para os casos do primeiro ao 4º período de evadidos, foram geradas regras baseadas somente nos casos de reprovação para fins de comparação entre casos de sucesso em comparação aos casos de fracasso. Ademais, os *insights* e possíveis hipóteses são discutidas período a período.

4.1 1º Período

O 1º período do curso de Engenharia da Computação é composto pelas disciplinas básicas do curso. Excepcionalmente neste período, os discentes não podem escolher quais disciplinas eles vão fazer e nem têm o poder de efetuar o trancamento de disciplinas. Por outro lado, eles podem abandonar a disciplina ou reprovar.

Após aplicar os dados do 1º período no algoritmo *Apriori*, foram mostradas nas 10 primeiras regras ordenadas pelas métricas suporte e confiança, uma vez que os resultados com $lift < 1$ já foram removidos como parâmetro do experimento. Veja os resultados na Tabela 6. Observa-se que as regras geradas são todas envolvendo a aprovação, que é representada pela letra A. Além disso, há uma combinação das disciplinas do 1º período, como Cálculo Diferencial e Integral I, Geometria Analítica e Álgebra Linear, Introdução a Engenharia de Computação, Metodologia de Pesquisa Aplicada à Computação e Programação de Computadores I. Ademais a disciplina Química Geral apareceu como antecedente na segunda regra e a disciplina de Introdução à Engenharia da Computação, sendo a mais comum nesse período.

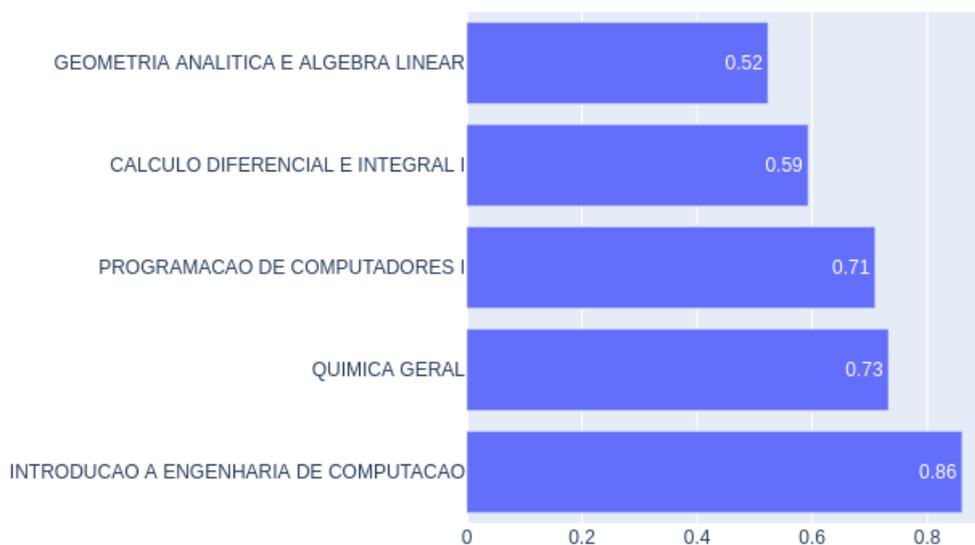
Tabela 6 – Principais resultados do 1º período diplomados.

antecedentes	consequentes	suporte	confiança
cálculo diferencial e integral I_A , metodologia de pesquisa aplicada a computação $_A$	introdução a engenharia de computação $_A$	0.2907	1
metodologia de pesquisa aplicada a computação $_A$, química geral $_A$	introdução a engenharia de computação $_A$	0.2674	1
geometria analítica e álgebra linear $_A$, cálculo diferencial e integral I_A , metodologia de pesquisa aplicada a computação $_A$	introdução a engenharia de computação $_A$	0.2674	1
programação de computadores I_A , cálculo diferencial e integral I_A , metodologia de pesquisa aplicada a computação $_A$	introdução a engenharia de computação $_A$	0.2674	1
geometria analítica e álgebra linear $_A$, metodologia de pesquisa aplicada a computação $_A$, química geral $_A$	introdução a engenharia de computação $_A$	0.2558	1
programação de computadores I_A , geometria analítica e álgebra linear $_A$, metodologia de pesquisa aplicada a computação $_A$, introdução a engenharia de computação $_A$	cálculo diferencial e integral I_A	0.2558	1
programação de computadores I_A , geometria analítica e álgebra linear $_A$, cálculo diferencial e integral I_A , metodologia de pesquisa aplicada a computação $_A$	introdução a engenharia de computação $_A$	0.2558	1
cálculo diferencial e integral I_A , metodologia de pesquisa aplicada a computação $_A$, química geral $_A$	geometria analítica e álgebra linear $_A$	0.2442	1
cálculo diferencial e integral I_A , metodologia de pesquisa aplicada a computação $_A$, química geral $_A$	introdução a engenharia de computação $_A$	0.2442	1
geometria analítica e álgebra linear $_A$, cálculo diferencial e integral I_A , metodologia de pesquisa aplicada a computação $_A$, química geral $_A$	introdução a engenharia de computação $_A$	0.2442	1

Fonte: Autoria própria.

A Figura 2 indica os cinco maiores suportes no conjunto de dados do 1º período diplomados. O gráfico está na cor azul pois representa as disciplinas que tiveram como resultado aprovação. Como as disciplinas do 1º período são pré-determinadas e base do curso, faz sentido ter um suporte mais alto e um maior índice de aprovação.

Figura 2 – Top 5 suportes 1º período diplomados.



Fonte: Autor do trabalho

4.1.1 1º Período Evadidos

Ao analisar os resultados dos alunos evadidos, foi identificado que a situação das regras geradas foram disciplinas que em sua maioria foi o estado de reprovação, ou seja, as disciplinas que está representada ao final com a letra R. São elas: Química Geral e Cálculo Diferencial e Integral I sendo maioria como consequência. Esse resultado nos alerta para o fato de a combinação dos antecedentes impactar com um peso maior negativamente em cima das reprovações nessas disciplinas, uma vez que se comparado com a [Tabela 6](#) a disciplina de Química por exemplo, aparece poucas vezes como antecedente de sucesso. Isso significa que, no grupo de alunos diplomados, a disciplina de Química, aparece poucas vezes nas regras e sempre aparece como antecedente. Já para o grupo de Evadidos, essa mesma disciplina aparece como consequência de reprovação em muitos casos. Observe a [Tabela 7](#) para maiores detalhes.

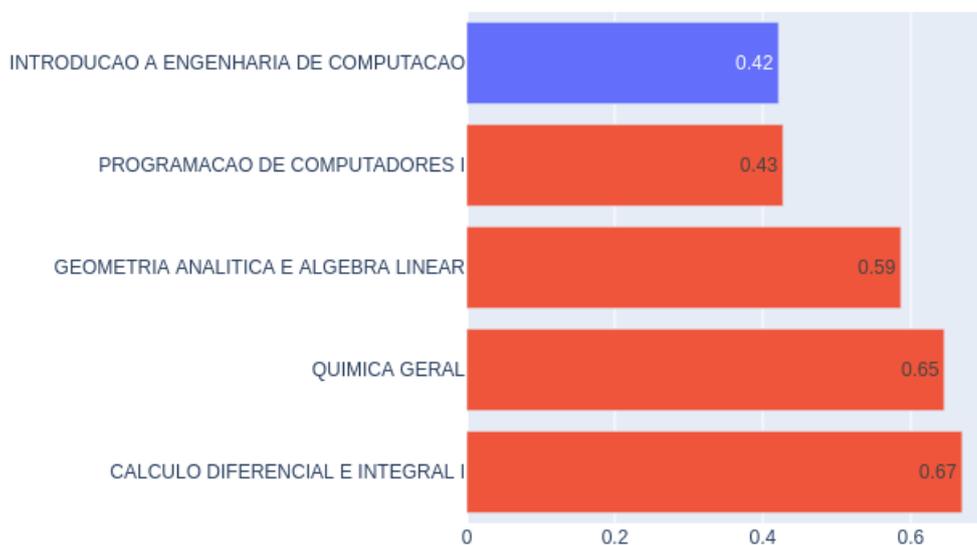
Tabela 7 – Principais resultados do 1º período evadidos.

antecedentes	consequentes	suporte	confiança
cálculo diferencial e integral I_R , geometria analítica e álgebra linear R , introdução a engenharia de computação R	química geral R	0.1532	0.9870
geometria analítica e álgebra linear R , programação de computadores I_R , introdução a engenharia de computação R	química geral R	0.1351	0.9853
cálculo diferencial e integral I_R , geometria analítica e álgebra linear R , programação de computadores I_R , introdução a engenharia de computação R	química geral R	0.1310	0.9848
geometria analítica e álgebra linear R , introdução a engenharia de computação R	química geral R	0.1573	0.9750
geometria analítica e álgebra linear R , introdução a engenharia de computação R , química geral R	cálculo diferencial e integral I_R	0.1532	0.9744
programação de computadores I_R , cálculo diferencial e integral I_R , química geral R	geometria analítica e álgebra linear R	0.1371	0.9714
geometria analítica e álgebra linear R , programação de computadores I_R , introdução a engenharia de computação R	cálculo diferencial e integral I_R	0.1331	0.9706
geometria analítica e álgebra linear R , introdução a engenharia de computação R , programação de computadores I_R , química geral R	cálculo diferencial e integral I_R	0.1310	0.9701
geometria analítica e álgebra linear R , introdução a engenharia de computação R	cálculo diferencial e integral I_R	0.1552	0.9625
metodologia da pesquisa R , introdução a engenharia de computação R	química geral R	0.1008	0.9615

Fonte: Autoria própria.

A análise do suporte das disciplinas para os discentes evadidos, observada na [Figura 3](#) aponta na cor vermelha o suporte de reprovação nas disciplinas. Essa análise indica apenas uma disciplina em comum nos top suportes, a Introdução a Engenharia de Computação. Esta análise identifica as principais disciplinas com reprovação no grupo de evadidos, embora os evadidos tenham obtido sucesso na disciplina de Introdução a Engenharia de Computação, o suporte de reprovação nas demais disciplinas se mostram mais evidentes neste grupo.

Figura 3 – Top 5 suportes 1º período evadidos.



Fonte: Autor do trabalho

4.2 2º Período

No 2º período as disciplinas que mais apareceram entre as mais frequentes no grupo de alunos diplomados foram: Física, Programação de Computadores II, Estatística e Probabilidade, Matemática Discreta e Cálculo Diferencial e Integral II. Contudo, como valor de consequentes têm-se uma frequência alta das disciplinas Cálculo Diferencial e Integral II e Estatística e Probabilidade. Isso significa que para quaisquer combinações entre as disciplinas antecedentes, sempre resultará em uma dessas sendo a consequentes, e o mais importante, todas de aprovação. As regras estão apresentadas na [Tabela 8](#). Porém, observe que o grau de confiança 1 é estabelecido apenas nas três primeiras regras, isto significa que 100% das vezes que as disciplinas antecedentes são cursadas e aprovadas, as consequentes também são.

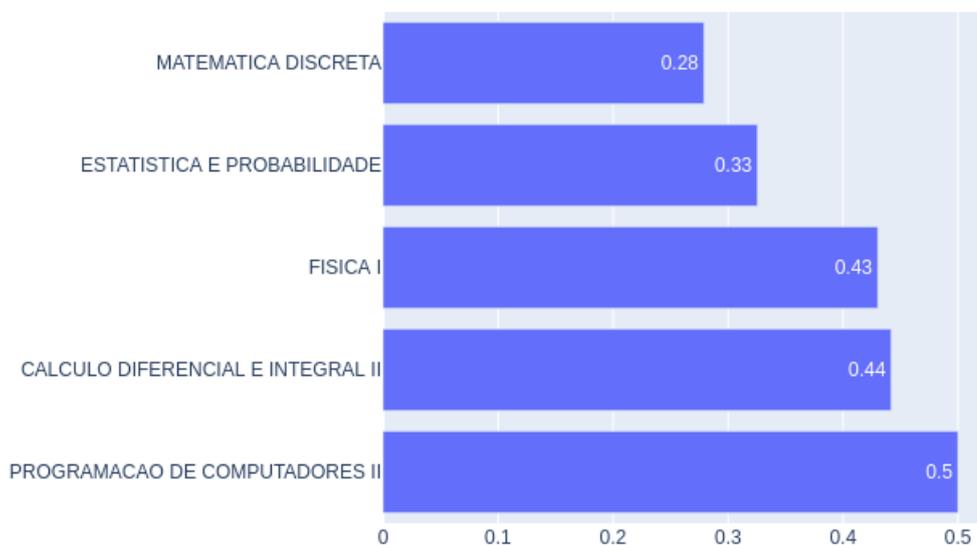
Tabela 8 – Principais resultados do 2º período diplomados.

antecedentes	consequentes	suporte	confiança
física I _A , programação de computadores II _A	cálculo diferencial e integral II _A	0.2209	1
física I _A , programação de computadores II _A , estatística e probabilidade _A	cálculo diferencial e integral II _A	0.1395	1
física I _A , programação de computadores II _A , matemática discreta _A	cálculo diferencial e integral II _A	0.1279	1
física I _A , estatística e probabilidade _A , programação de computadores II _A , matemática discreta _A	cálculo diferencial e integral II _A	0.1163	1
programação de computadores II _A , estatística e probabilidade _A	cálculo diferencial e integral II _A	0.1628	0.9333
física I _A , matemática discreta _A	cálculo diferencial e integral II _A	0.1512	0.9286
programação de computadores II _A , estatística e probabilidade _A , matemática discreta _A	cálculo diferencial e integral II _A	0.1395	0.9231
física I _A , estatística e probabilidade _A , matemática discreta _A	cálculo diferencial e integral II _A	0.1279	0.9167
física I _A , programação de computadores II _A , matemática discreta _A	estatística e probabilidade _A	0.1163	0.9091
física I _A , cálculo diferencial e integral II _A , programação de computadores II _A , matemática discreta _A	estatística e probabilidade _A	0.1163	0.9091

Fonte: Autoria própria.

A Figura 4 apresenta o *ranking* dos suportes para as cinco disciplinas do 2º período diplomados. Para os alunos diplomados, as disciplinas com maior frequência são Programação de Computadores II e Cálculo Diferencial e Integral II. É interessante observar que Matemática Discreta, por ser uma disciplina indicada para se fazer no 2º período, aparece com o menor suporte neste ranking. O curioso é que essas disciplinas fazem parte da grade atual como recomendadas para o segundo período, mas o suporte da disciplina Matemática Discreta e Estatística e Probabilidade chega a ser menor e mais distante do suporte das demais disciplinas.

Figura 4 – Top 5 suportes 2º período diplomados.



Fonte: Autor do trabalho

4.2.1 2º Período Evadidos

Após aplicar o *Apriori* para extrair as regras de associação entre as disciplinas do 2º período, foi identificado somente duas relações entre as disciplinas Geometria Analítica e Álgebra Linear e Cálculo Diferencial e Integral I, conforme apresentado na Tabela 9. Isso indica que aproximadamente 21% dos alunos cursam essas disciplinas no 2º período, e que aproximadamente 65% das vezes esses alunos reprovam neste conjunto de escolha. Diferentemente dos alunos que obtiveram sucesso na graduação e o sucesso no 1º período ou no 2º período e também cursaram essas disciplinas, para os alunos que chegam ao 2º período com a pendência dessas duas disciplinas, eles tendem ao fracasso acadêmico e por consequência, a evasão.

Tabela 9 – Principais resultados do 2º período evadidos.

antecedentes	consequentes	suporte	confiança
geometria analítica e álgebra linear _R	cálculo diferencial e integral I _R	0.2169	0.6792
cálculo diferencial e integral I _R	geometria analítica e álgebra linear _R	0.2169	0.6207

Fonte: Autoria própria.

O *ranking* de suporte apresentado na [Figura 5](#) apresenta as disciplinas com maiores suporte, entre elas está o Cálculo Diferencial e Integral I que contém disciplinas em situação de reprovado quanto aprovado. Entretanto a quantidade de disciplinas em reprovação continua sendo quase que unânime.

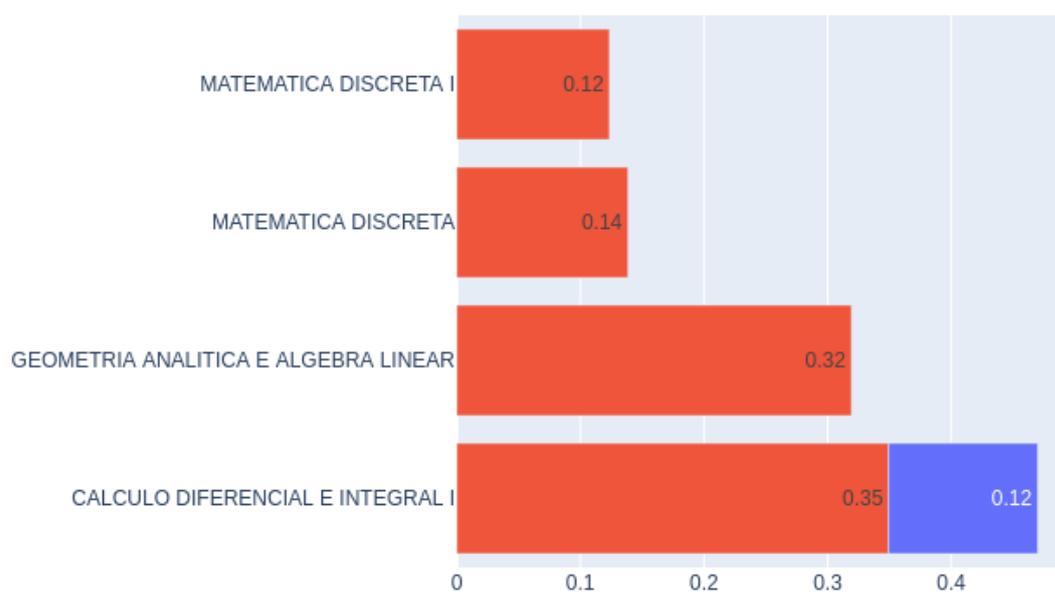


Figura 5 – Top 5 suportes 2º período evadidos.

4.3 3º Período

Com relação ao 3º período, para os alunos que obtiveram sucesso acadêmico, mais uma vez todas as disciplinas possuem caráter de aprovação. Sendo as mais comuns como Cálculo Diferencial e Integral III, Física II e Algoritmos e Estrutura de Dados I, disciplina de Introdução as Equações Diferenciais Ordinárias, também aparece entre as principais regras. Isto é, essas disciplinas além de serem mais comuns entres os alunos que obtiveram sucesso, elas garantem a aprovação. Veja a [Tabela 10](#).

Tabela 10 – Principais resultados do 3º período diplomados.

antecedentes	consequentes	suporte	confiança
física II _A , princípios de eletrônica digital _A , cálculo diferencial e integral III _A	introdução as equações diferenciais ordinárias _A	0.1977	1
cálculo diferencial e integral III _A , algoritmos e estrutura de dados I _A	introdução as equações diferenciais ordinárias _A	0.1860	1
física II _A , algoritmos e estrutura de dados I _A	cálculo diferencial e integral III _A	0.1628	1
física II _A , algoritmos e estrutura de dados I _A	introdução as equações diferenciais ordinárias _A	0.1628	1
física II _A , introdução as equações diferenciais ordinárias _A , algoritmos e estrutura de dados I _A	cálculo diferencial e integral III _A	0.1628	1
física II _A , algoritmos e estrutura de dados I _A , cálculo diferencial e integral III _A	introdução as equações diferenciais ordinárias _A	0.1628	1
física II _A , algoritmos e estrutura de dados I _A	cálculo diferencial e integral III _A , introdução as equações diferenciais ordinárias _A	0.1628	1
cálculo diferencial e integral III _A , princípios de eletrônica digital _A , algoritmos e estrutura de dados I _A	introdução as equações diferenciais ordinárias _A	0.1512	1
princípios de eletrônica digital _A , introdução as equações diferenciais ordinárias _A , algoritmos e estrutura de dados I _A	cálculo diferencial e integral III _A	0.1512	1
física II _A , princípios de eletrônica digital _A , algoritmos e estrutura de dados I _A	cálculo diferencial e integral III _A	0.1279	1

Fonte: Autoria própria.

A [Figura 6](#) apresenta o suporte dos discentes do 3º período diplomados. Aqui é possível observar a disciplina de Cálculo Diferencial e Integral III como o maior suporte. Essa disciplina encontra-se em quase todas as regras geadas na [Tabela 10](#), ressaltando a escolha e importância dessa disciplina no 3º período diplomados.

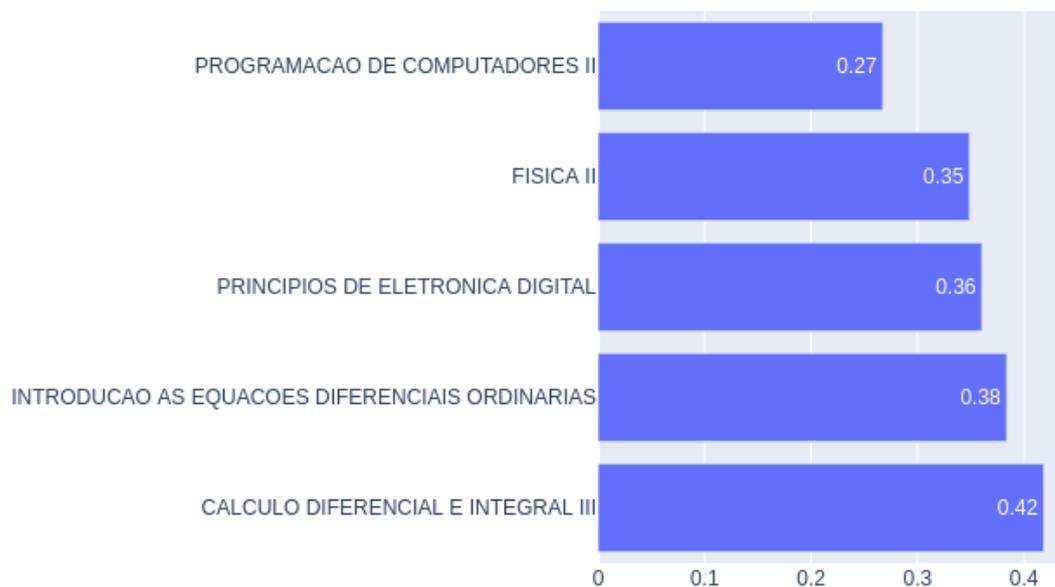


Figura 6 – Top 5 suportes 3º período diplomados.

4.3.1 3º Período Evadidos

A análise das regras para o 3º período dos alunos evadidos segue similar à análise dos alunos do 2º período diplomados. Desse modo as disciplinas mais presentes são Cálculo Diferencial e Integral III e Geometria Analítica e Álgebra Linear como apresentado na [Tabela 11](#). Embora o suporte e a confiança para essas regras tenham diminuído, são discrepante os resultados se comparados com a [Tabela 10](#) onde essas disciplinas já não fazem mais parte do conjunto de antecedentes ou consequentes. Nesse ponto, fica claro a diferença entre o caminho acadêmico dos alunos que concluíram a graduação dos que não concluíram.

Tabela 11 – Principais resultados do 3º período evadidos.

antecedentes	consequentes	suporte	confiança
cálculo diferencial e integral I_R	geometria analítica e álgebra linear $_R$	0.1723	0.6212
geometria analítica e álgebra linear $_R$	cálculo diferencial e integral I_R	0.1723	0.5942

Fonte: Autoria própria.

Observa-se na [Figura 7](#) que a no 3º período as disciplinas com suporte mais altos são apenas as que possuem reprovação. E todas essas disciplinas são da área de exatas. Isso pode confirmar as dificuldades que os discentes encontram ao cursar essas disciplinas.

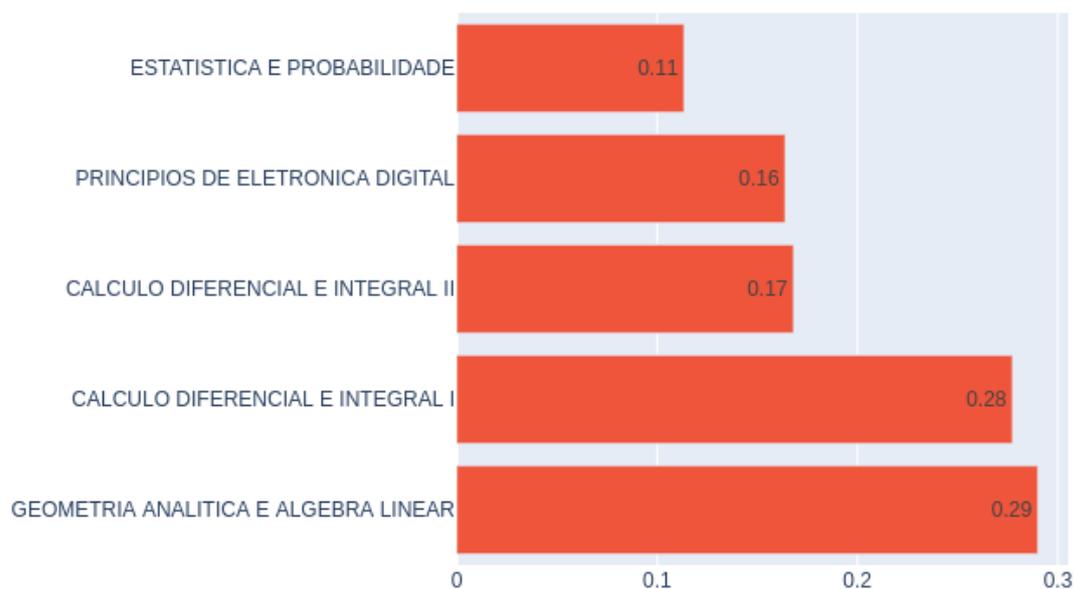


Figura 7 – Top 5 suportes 3º período evadidos.

4.4 4º Período

Após analisar as regras geradas para o 4º período, compreende-se que as matérias mais relevantes neste período são: Sinais e Sistemas, Algoritmos e Estrutura de Dados II, Teoria dos Grafos, Física III e Circuitos Elétricos I. Essas informações podem ser observadas na [Tabela 12](#), bem como as associações entre disciplinas antecedentes e consequentes.

Tabela 12 – Principais resultados do 4^o período diplomados.

antecedentes	consequentes	suporte	confiança
sinais e sistemas _A , algoritmos e estrutura de dados II _A	teoria dos grafos _A	0.1163	1
algoritmos e estrutura de dados II _A , circuitos elétricos I _A	sinais e sistemas _A	0.1047	1
algoritmos e estrutura de dados II _A , circuitos elétricos I _A	física III _A	0.1047	1
algoritmos e estrutura de dados II _A , circuitos elétricos I _A	teoria dos grafos _A	0.1047	1
teoria dos grafos _A , circuitos elétricos I _A	algoritmos e estrutura de dados II _A	0.1047	1
teoria dos grafos _A , circuitos elétricos I _A	sinais e sistemas _A	0.1047	1
teoria dos grafos _A , circuitos elétricos I _A	física III _A	0.1047	1
física III _A , sinais e sistemas _A , algoritmos e estrutura de dados II _A	circuitos elétricos I _A	0.1047	1
física III _A , algoritmos e estrutura de dados II _A , circuitos elétricos I _A	sinais e sistemas _A	0.1047	1
sinais e sistemas _A , algoritmos e estrutura de dados II _A , circuitos elétricos I _A	física III _A	0.1047	1

Fonte: Autoria própria.

A análise do suporte do 4^o período disponível na [Figura 8](#), indica uma congruência na decisão dos discentes, a distribuição do suporte está condensada entre as disciplinas que envolvem principalmente a área de Física, ou utilizam disciplinas desse âmbito como pré-requisito. Os valores de suportes juntos às disciplinas, indicam que geralmente 29% dos alunos cursam a disciplina Física III em período regular.

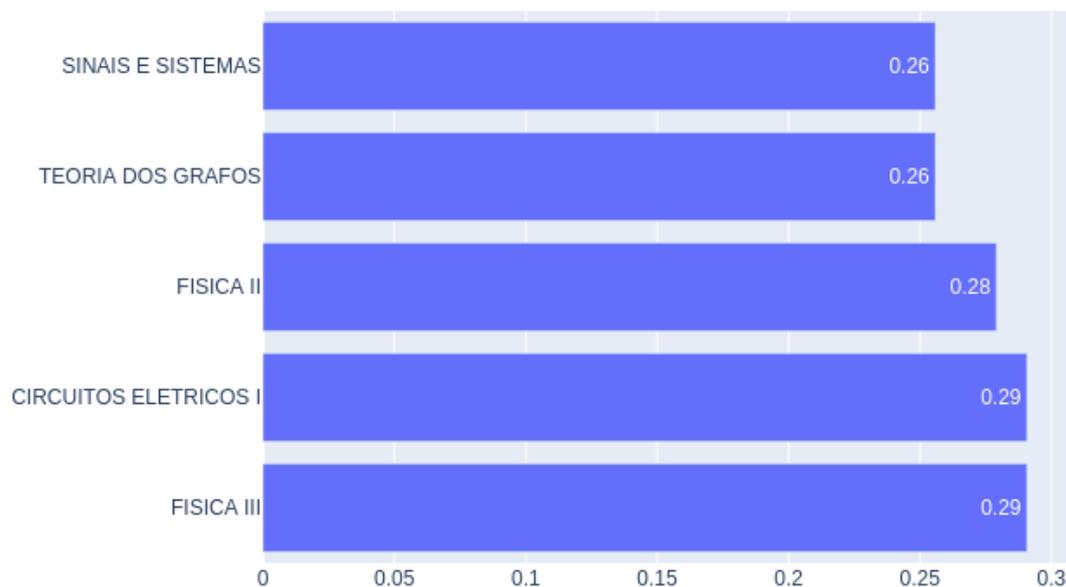


Figura 8 – Top 5 suportes 4º período diplomados.

Fonte: Autoria própria.

4.4.1 4º Período Evadidos

Analisando as regras para o 4º período para os alunos evadidos, chegou-se à mesma conclusão que para o segundo e 3º período, a única diferença é que o suporte mínimo e a confiança das regras seguem em decadência, como apresentado na Tabela 13. Por esse fato não foram realizadas posteriormente um comparativo com os discentes evadidos para os demais períodos.

Tabela 13 – Principais resultados do 4º período evadidos.

antecedentes	consequentes	suporte	confiança
cálculo diferencial e integral I_R	geometria analítica e álgebra linear $_R$	0.1032	0.6154
geometria analítica e álgebra linear $_R$	cálculo diferencial e integral I_R	0.1032	0.5517

Fonte: Autoria própria.

Já na análise de suporte para as disciplinas do 4º período dos alunos evadidos, temos ainda todas as disciplinas indicando a reprovação, como apresentado na Figura 9. Neste caso pode-se observar que as disciplinas do 1º período ainda possuem um peso muito grande para o sucesso acadêmico.

Para os demais períodos não foi gerado o *top 5* item sets com seu respectivo suporte, pois o volume de evadidos se faz presente em maior quantidade até o quartº período diplomados.

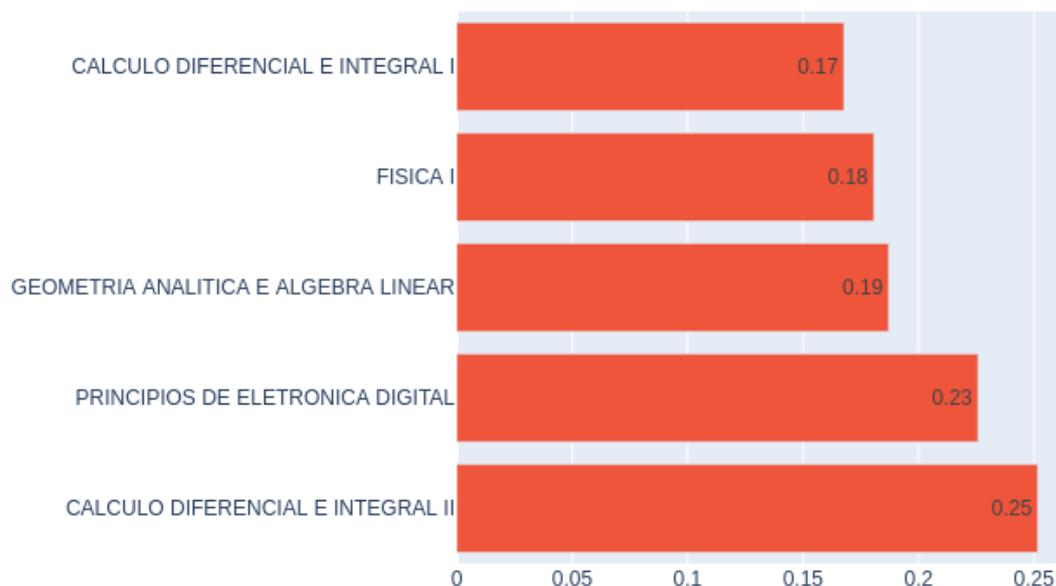


Figura 9 – Top 5 suportes 4º período evadidos.

Fonte: Autoria própria.

4.5 5º Período

Para o 5º período, o que observa-se é que o número de disciplinas cursadas já começa a diminuir, pois a quantidade de alunos que chegam ao 5º período também, com isso gera menos disciplinas distintas. As principais combinações são geradas envolvendo somente quatro disciplinas, são elas: Engenharia de Software, Organização e Arquitetura de Computadores, Análise Numérica e Física IV. A [Tabela 14](#) apresenta em mais detalhes o ajuste dessas disciplinas em conjunto.

Tabela 14 – Principais resultados do 5º período diplomados.

antecedentes	consequentes	suporte	confiança
engenharia de <i>software</i> I _A	análise numérica _A	0.1548	1
organização e arquitetura de computadores I _A	análise numérica _A	0.1548	1
organização e arquitetura de computadores I _A	engenharia de <i>software</i> I _A	0.1548	1
engenharia de <i>software</i> I _A	organização e arquitetura de computadores I _A	0.1548	1
análise numérica _A , organização e arquitetura de computadores I _A	engenharia de <i>software</i> I _A	0.1548	1
análise numérica _A , engenharia de <i>software</i> I _A	organização e arquitetura de computadores I _A	0.1548	1
organização e arquitetura de computadores I _A , engenharia de <i>software</i> I _A	análise numérica _A	0.1548	1
organização e arquitetura de computadores I _A	análise numérica _A , engenharia de <i>software</i> I _A	0.1548	1
engenharia de <i>software</i> I _A	análise numérica _A , organização e arquitetura de computadores I _A	0.1548	1
análise numérica _A , física iv _A	engenharia de <i>software</i> I _A	0.1310	1

Fonte: Autoria própria.

A análise do suporte se mantém distribuída igualmente entre as disciplinas da área da física entretanto a que mantém-se como maior destaque é a de Engenharia de Software.

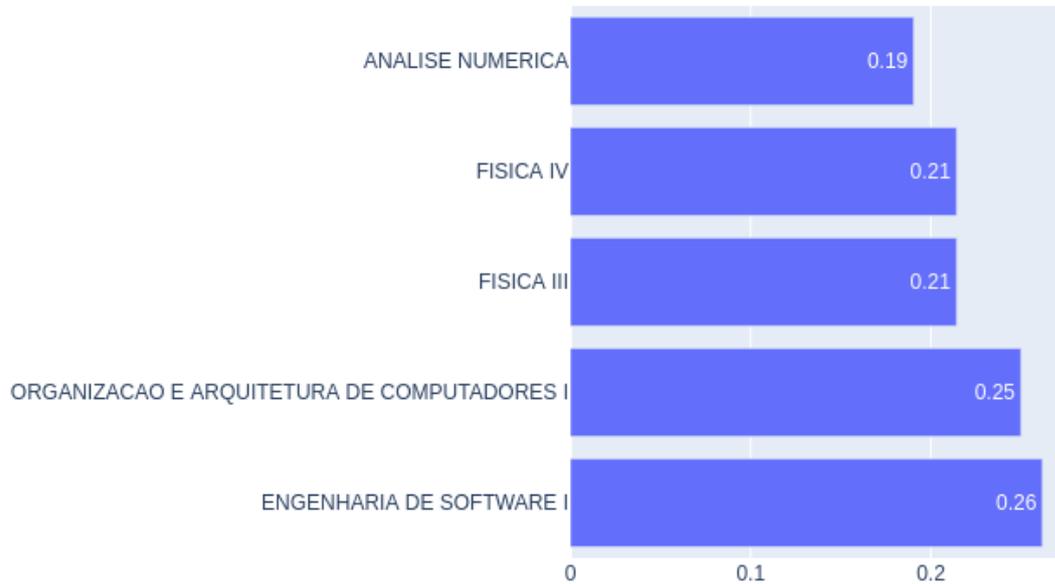


Figura 10 – Top 5 suportes 5º período diplomados.

Fonte: Autoria própria.

4.6 6º Período

No 6º período dos alunos diplomados aparecem regras envolvendo quatro disciplinas fundamentais: Eletrônica I, Fundamentos de Comunicações, Projeto e Análise de Algoritmos e Organização e Arquitetura de Computadores. Outra vez essas disciplinas encontram-se em situação de aprovação. Com um suporte relativamente alto para as primeiras disciplinas, esse suporte indica que aproximadamente 17% dos alunos que cursaram as disciplinas antecedentes e tiveram como consequentes as apresentadas na [Tabela 15](#) e obtiveram sucesso 100% das vezes.

Tabela 15 – Principais resultados do 6º período diplomados.

antecedentes	consequentes	suporte	confiança
eletrônica I _A , fundamentos de comunicações _A	projeto e análise de algoritmo _A	0.1667	1
eletrônica I _A , projeto e análise de algoritmo _A	fundamentos de comunicações _A	0.1667	1
organização e arquitetura de computadores II _A , fundamentos de comunicações _A	projeto e análise de algoritmo _A	0.1667	1
eletrônica I _A , organização e arquitetura de computadores II _A	fundamentos de comunicações _A	0.1667	1
eletrônica I _A , fundamentos de comunicações _A	organização e arquitetura de computadores II _A	0.1667	1
organização e arquitetura de computadores II _A , fundamentos de comunicações _A	eletrônica I _A	0.1667	1
eletrônica I _A , organização e arquitetura de computadores II _A	projeto e análise de algoritmo _A	0.1667	1
eletrônica I _A , projeto e análise de algoritmo _A	organização e arquitetura de computadores II _A	0.1667	1
eletrônica I _A , organização e arquitetura de computadores II _A , fundamentos de comunicações _A	projeto e análise de algoritmo _A	0.1667	1
eletrônica I _A , organização e arquitetura de computadores II _A , projeto e análise de algoritmo _A	fundamentos de comunicações _A	0.1667	1

Fonte: Autoria própria.

Para o 6º período o *item sets* das disciplinas também apresentam-se muito similares uns aos outros, o que trás evidências para o valor unânime de 1 na métrica de confiança.

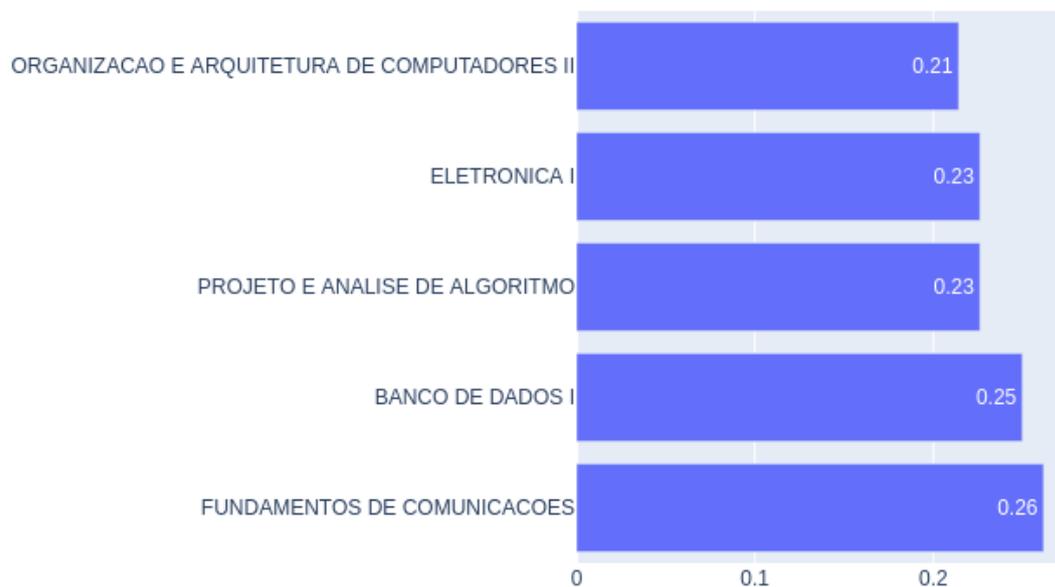


Figura 11 – Top 5 suportes 6º período diplomados.

Fonte: Autoria própria.

4.7 7º Período

A análise do 7º período também apresenta quatro disciplinas entre as mais comuns, são elas: Fundamentos Teóricos da Computação, Sistemas Operacionais, Computação Gráfica, Microprocessadores e Microcontroladores. Outra vez, as disciplinas encontram-se em situação de aprovado, entretanto o nível de confiança começa a cair a partir da terceira regra. Isto significa que essas regras geradas estão presentes na base de dados, entretanto, sem unanimidade nos casos. Esses resultados são apresentados na [Tabela 16](#).

Tabela 16 – Principais resultados do 7º período diplomados.

antecedentes	consequentes	suporte	confiança
fundamentos teóricos da computação _A , microprocessadores e microcontroladores _A	computação gráfica _A	0.1446	1
sistemas operacionais _A , microprocessadores e microcontroladores _A , fundamentos teóricos da computação _A	computação gráfica _A	0.1205	1
computação gráfica _A , microprocessadores e microcontroladores _A	fundamentos teóricos da computação _A	0.1446	0.9231
sistemas operacionais _A , computação gráfica _A	microprocessadores e microcontroladores _A	0.1325	0.9167
sistemas operacionais _A , computação gráfica _A	fundamentos teóricos da computação _A	0.1325	0.9167
sistemas operacionais _A , computação gráfica _A , fundamentos teóricos da computação _A	microprocessadores e microcontroladores _A	0.1205	0.9091
sistemas operacionais _A , computação gráfica _A , microprocessadores e microcontroladores _A	fundamentos teóricos da computação _A	0.1205	0.9091
computação gráfica _A	fundamentos teóricos da computação _A	0.1687	0.8750
fundamentos teóricos da computação _A , computação gráfica _A	microprocessadores e microcontroladores _A	0.1446	0.8571
sistemas operacionais _A , microprocessadores e microcontroladores _A	computação gráfica _A	0.1325	0.8462

Fonte: Autoria própria.

Após analisar o suporte das disciplinas, foi constatado que as disciplinas com suporte mais altos contém quatro estão de acordo com a grade curricular, são elas: Microprocessadores e Microcontroladores, Fundamentos Teóricos da Computação, Computação Gráfica e Sistemas Operacionais. Ademais a disciplina Fundamentos de Ciência do Ambiente aparece como um adiantamento de disciplina neste contexto, saindo do 10º período para o sétimo.

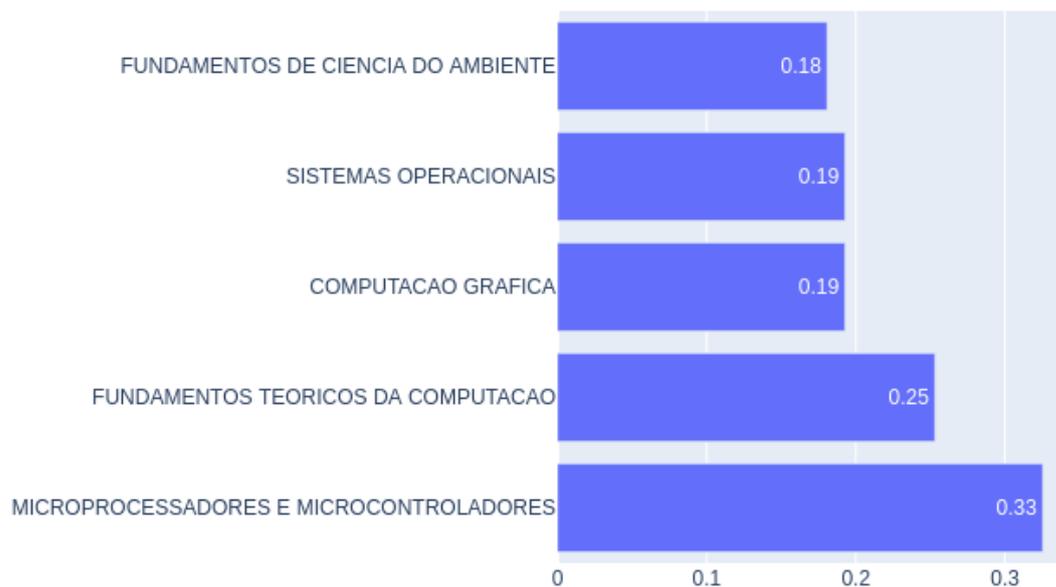


Figura 12 – Top 5 suportes 7º período diplomados.

Fonte: Autoria própria.

4.8 8º Período

A análise do 8º período não teve dados suficientes para a geração de no mínimo dez regras, esse fato pode ser explicado pela quantidade de alunos que chegam até ao final do curso, bem como a diversidade de escolha de disciplinas neste período, onde é possível pegar eletivas e muitos alunos se diferem nesse aspecto. Entretanto, as disciplinas frequentes foram: Redes de Computadores I, Inteligência Artificial, Linguagens de Programação, Interação Humano-Computador e Cálculo Diferencial e Integral II. Esses dados são apresentados a seguir na [Tabela 17](#).

Tabela 17 – Principais resultados do 8º período diplomados.

antecedentes	consequentes	suporte	confiança
redes de computadores I_A	cálculo diferencial e integral II_A	0.1084	0.9000
inteligência artificial A	linguagens de programação A	0.1446	0.8000
linguagens de programação A	inteligência artificial A	0.1446	0.7059
interação humano-computador A	inteligência artificial A	0.1084	0.6429
inteligência artificial A	interação humano-computador A	0.1084	0.6000
inteligência artificial A	redes de computadores I_A	0.1084	0.6000

Fonte: Autoria própria.

A análise do suporte para o 8º período é apresentada na [Figura 13](#) aparece em sua maioria as disciplinas sugeridas pela grade atual. Entretanto avalia-se também a popularidade da disciplina de Microprocessadores e Microcontroladores, neste caso pode indicar a repetência dos alunos, e a popularidade da disciplina de Compiladores I. Essa disciplina seria cursada no próximo semestre, entretanto os discentes optam por adiantá-la em um período.

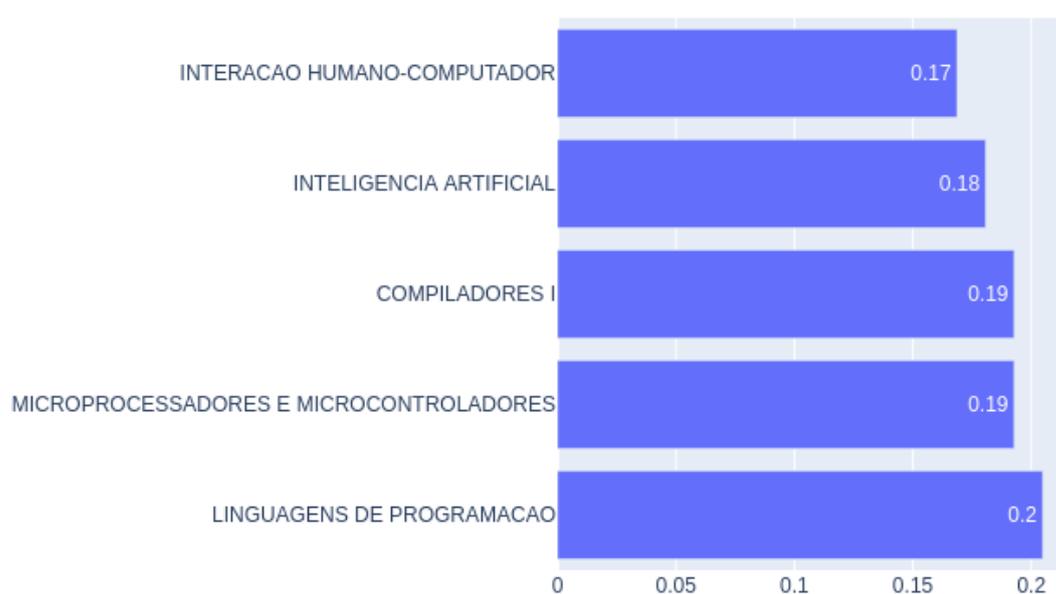


Figura 13 – Top 5 suportes 8º período diplomados.

Fonte: Autoria própria.

4.9 9º Período

Para o 9º período é observado que as disciplinas mais frequentes são: Redes de Computadores II, Sistemas Distribuídos, Trabalho de Conclusão de Curso I e Compiladores I. Note que o valor do suporte cai se comparado com os períodos anteriores, bem como a confiança, disponível na [Tabela 18](#). Isso se dá ao fato da redução da amostra e da variabilidade de disciplinas escolhidas por esses alunos. Entretanto, vale salientar que estes conjuntos de regra ainda trabalham somente com as situações de aprovação, isto é, essas disciplinas contem a certeza de aprovação em aproximadamente 10% dos casos, seguindo a análise pelo suporte.

Tabela 18 – Principais resultados do 9º período diplomados.

antecedentes	consequentes	suporte	confiança
redes de computadores II _A , trabalho de conclusão de curso I _A	sistemas distribuídos _A	0.1013	1
sistemas distribuídos _A , trabalho de conclusão de curso I _A	redes de computadores II _A	0.1013	0.8000
trabalho de conclusão de curso I _A	sistemas distribuídos _A	0.1266	0.7692
redes de computadores II _A , sistemas distribuídos _A	trabalho de conclusão de curso I _A	0.1013	0.7273
redes de computadores II _A	sistemas distribuídos _A	0.1392	0.6875
compiladores I _A	redes de computadores II _A	0.1266	0.6667
redes de computadores II _A	compiladores I _A	0.1266	0.6250
trabalho de conclusão de curso I _A	redes de computadores II _A	0.1013	0.6154
trabalho de conclusão de curso I _A	redes de computadores II _A , sistemas distribuídos _A	0.1013	0.6154
sistemas distribuídos _A	redes de computadores II _A	0.1392	0.6111

Fonte: Autoria própria.

O suporte das disciplinas do 10º período estão coerentes com a grade atual. Contudo a disciplina Fundamentos de Ciência do Ambiente é mais uma vez uma opção de antecipação.

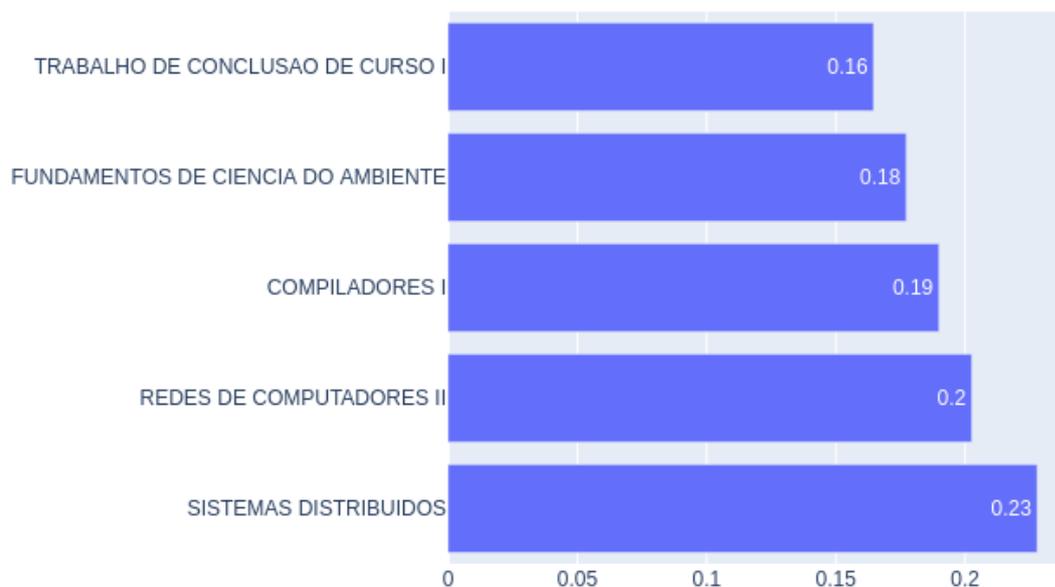


Figura 14 – Top 5 suportes 9º período diplomados.

Fonte: Autoria própria.

4.10 10º Período

Para os alunos do 10º período não foi possível gerar nenhuma regra de associação que respeitasse os parâmetros de suporte mínimo de 10% e $lift \geq 1$. Esse fato também pode ser explicado pela relação das disciplinas e a quantidade de alunos por período. Note que a partir do 10º período, a curva de discentes e a curva de disciplinas tendem a diminuir a distância conforme apresentado na ???. Esse fato emprega maior variabilidade na geração das regras, com isso as possibilidades de conjuntos *item sets* são tantas e a quantidade de instâncias são cada vez menores, assim não é possível obter o suporte mínimo exigido no experimento.

4.11 Considerações Finais

Nesta seção, aplicou-se técnicas de Mineração de Dados para Extração de Regras de Associação nas disciplinas cursadas pelos discentes do curso de Engenharia da Computação (EC) para identificar quais são as principais escolhas de sucesso dos alunos.

Foram descobertos alguns padrões interessantes entre os alunos diplomados, e um padrão fixo nos alunos evadidos. Para os alunos diplomados fica evidente a importância da aprovação nas disciplinas básicas do curso, que são suporte para as disciplinas posteriores. Além disso, é importante destacar que para o grupo de diplomados, gerou um maior número de possibilidades entre o conjunto de regras com suporte e confiança mais altos que o demais, indicando de certo modo a aplicabilidade dessas possibilidades gerarem sucesso. Em contrapartida, também neste período, para os alunos reprovados nas disciplinas de Cálculo Diferencial e Integral I e Geometria Analítica e Álgebra Linear podem ter grande correlação com o insucesso acadêmico no grupo de evadidos.

5 Conclusão

O presente trabalho mostrou que as técnicas de Mineração de Dados permitem explorar padrões e retirar conhecimento das bases de diversas bases de dados para apoiar a tomada de decisão no âmbito educacional. Esse pensamento é importante para a melhoria contínua de processos, principalmente quando envolve assuntos delicados como o desenvolvimento acadêmico. Foram apresentados os padrões de disciplinas cursadas pelos discentes do curso de Engenharia da Computação (EC) do Instituto de Ciências Exatas e Aplicadas (ICEA). Esses padrões permitiram realizar uma análise minuciosa período à período para entender o comportamento dos discentes que obtiveram sucesso acadêmico, ou seja, conseguiram diplomar-se, em relação aos discentes que evadiram.

Esses padrões foram extraídos através de uma modelagem de problema capaz de diferir a aprovação ou reprovação desses alunos nas disciplinas, assim foi possível encontrar relações de dependência entre a trajetória acadêmica dos discentes, descrevendo o comportamento dos alunos, retirando informações sobre padrões de aprovações e reprovações. Assim é possível recomendar as disciplinas adequadas para cada condição.

Após a aplicação do algoritmo *Apriori*, foi possível observar uma maior separação entre o grupo de alunos diplomados e evadidos, onde nos períodos iniciais a aprovação dos alunos nas disciplinas primárias do curso, levou a decisões mais assertivas em resultados de aprovação, com suporte e confiança alto entre os discentes. Entretanto, a reprovação nas disciplinas de Cálculo Diferencial e Integral I e Geometria Analítica e Álgebra Linear pode ser evidenciada no grupo de evadidos. Essa dificuldade desse grupo foi observada a partir do segundo período, sendo recorrente até o quarto período.

Ademais, o conhecimento adquirido após esse estudo, pode ser útil para a implementação de políticas de intervenção para as disciplinas Cálculo Diferencial e Integral I e Geometria Analítica e Álgebra Linear, oferecendo apoio e suporte aos alunos durante esse período por meio de flexibilização e aumento os horários de monitoria, ter uma tutoria com os professores, bem como aulas extras para essas disciplinas. Para as regras extraídas das bases de alunos diplomados, disponibilizar para os discente terem conhecimento das disciplinas que levaram os discentes anteriores ao sucesso acadêmico.

Ademais, como trabalhos futuros, expandir essa análise não só para o curso de Engenharia da Computação (EC), como também para os demais cursos e realizar experimentos incluindo as disciplinas comuns aos cursos, visando entender se esse padrão é de cunho individual ao curso de Engenharia da Computação (EC) ou se pode ser inferido para o Instituto. Além disso revisar o processamento de dados, onde é possível identificar disciplinas com o mesmo nome, entretanto com o código diferente, unificar essas disciplinas

em somente um código é interessante.

Referências

- AGRAWAL, R.; SRIKANT, R. et al. Fast algorithms for mining association rules. In: CITESEER. *Proc. 20th int. conf. very large data bases, VLDB*. [S.l.], 1994. v. 1215, p. 487–499. Citado na página 18.
- ANSELMO, F. C. G. Regras de associação-market basket analysis-itens frequentes e itens raros. 2017. Citado 3 vezes nas páginas 18, 19 e 20.
- BRIN, S. et al. Dynamic itemset counting and implication rules for market basket data. In: *Proceedings of the 1997 ACM SIGMOD international conference on Management of data*. [S.l.: s.n.], 1997. p. 255–264. Citado na página 20.
- CALDEIRA, D. M. *Caracterização do problema de evasão de discentes nos cursos do ICEA mediante técnicas de mineração de dados*. 2021. Monografia (Graduação em Sistema de Informação) - Instituto de Ciências Exatas e Aplicadas, Universidade Federal de Ouro Preto, João Monlevade, Brasil. Citado 3 vezes nas páginas 13, 14 e 16.
- COSTA, E. et al. Mineração de dados educacionais: Conceitos, técnicas, ferramentas e aplicações. In: . [S.l.: s.n.], 2012. Citado na página 12.
- FAYYAD, U. M. et al. Knowledge discovery and data mining: Towards a unifying framework. In: *KDD*. [S.l.: s.n.], 1996. v. 96, p. 82–88. Citado na página 16.
- FILHO, R. S. et al. A evasão no ensino superior brasileiro. *Cadernos De Pesquisa*, v. 37, 12 2007. Citado na página 13.
- FRANZ, A. C. M. et al. Desenvolvimento de uma ferramenta visual de classificação de imagens para o ensino de machine learning no ensino médio. Florianópolis, SC, 2021. Citado na página 22.
- GAMA, J. et al. Extração de conhecimento de dados: data mining. 2012. Citado 2 vezes nas páginas 19 e 20.
- GONÇALVES, A. P. F. *Definição de um modelo de inteligência artificial para a identificação do padrão curricular dos alunos do ICEA*. 2022. Monografia (Graduação em Sistemas de Informação) - Instituto de Ciências Exatas e Aplicadas, Universidade Federal de Ouro Preto, João Monlevade, Brasil. Citado 4 vezes nas páginas 13, 14, 16 e 21.
- HAN, J.; PEI, J.; TONG, H. *Data mining: concepts and techniques*. [S.l.]: Morgan kaufmann, 2022. Citado na página 17.
- HOED, R. M. Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de computação. 2016. Citado na página 12.
- IMIELINSKI, T.; SWAMI, A.; AGARWAL, R. Mining association rules between sets of items in large databases. In: *Proc. ACM SIGMOD Conf. Management of Data*. [S.l.: s.n.], 1993. v. 10, n. 170035.170072. Citado na página 18.

JUNIOR, I. B. *Uso de mineração de dados educacionais para a classificação e identificação de perfis de evasão de graduandos em Sistemas de Informação*. Dissertação (B.S. thesis) — Universidade Federal do Rio Grande do Norte, 2018. Citado na página 15.

LAROSE, D. T. An introduction to data mining. *Traduction et adaptation de Thierry Vallaud*, 2005. Citado na página 17.

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.], 2012. v. 1, n. 1. Citado na página 15.

MAPHOSA, M.; DOORSAMY, W.; PAUL, B. A review of recommender systems for choosing elective courses. *International Journal of Advanced Computer Science and Applications*, v. 11, 01 2020. Citado na página 16.

MONDAL, B. et al. A course recommendation system based on grades. In: *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*. [S.l.: s.n.], 2020. p. 1–5. Citado na página 16.

MUSSLINER, B. O. et al. O problema da evasão universitária no sistema público de ensino superior: uma proposta de ação com base na atuação de uma equipe multidisciplinar. *Brazilian Journal of Development*, v. 7, n. 4, p. 42674–42692, 2021. Citado na página 12.

RODRIGUES, E. H. A. *Análise de caracterização quantitativa e predição da evasão escolar nos cursos da área de computação do ICEA por meio de técnicas de Data Science*. 2022. Monografia (Graduação em Engenharia da Computação) - Instituto de Ciências Exatas e Aplicadas, Universidade Federal de Ouro Preto, João Monlevade, Brasil. Citado 3 vezes nas páginas 13, 14 e 16.

RODRIGUES, G. D. et al. Uma avaliação empírica de transformações aleatórias aplicadas a ensembles de agrupamento de dados. Universidade Federal de Uberlândia, 2018. Citado na página 15.

ROMÃO, W. et al. Extração de regras de associação em c&t: O algoritmo apriori. *XIX Encontro Nacional em Engenharia de Produção*, sn, v. 34, p. 37–39, 1999. Citado na página 18.

ROMERO CRISTÓBAL, J. R.; VENTURA, S. A survey on pre-processing educational data. In: _____. [S.l.: s.n.], 2014. v. 524, p. 29–64. ISBN 978-3-319-02737-1. Citado na página 12.

SCHNEIDER, L. F. Aplicação do processo de descoberta de conhecimento em dados do poder judiciário do estado do rio grande do sul. 2003. Citado na página 18.

WITTEN, I. H. et al. Practical machine learning tools and techniques. In: *Data Mining*. [S.l.: s.n.], 2005. v. 2, n. 4. Citado na página 16.

Anexos



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
PRÓ-REITORIA DE GRADUAÇÃO



CURSO DE ENGENHARIA DE COMPUTAÇÃO (BACHARELADO) - 2022/2(currículo 1)
CAMPUS JOAO MONLEVADE

CÓDIGO	DISCIPLINAS OBRIGATÓRIAS	PRÉ-REQUISITO	CHS/CHE	CHA	AULAS		PER
					T	P	
CEA038	QUIMICA GERAL		30/0	36	2	0	1
CEA039	QUIMICA GERAL EXPERIMENTAL		30/0	36	0	2	1
CEA049	GEOMETRIA ANALITICA E ALGEBRA LINEAR		60/0	72	4	0	1
CEA050	CALCULO DIFERENCIAL E INTEGRAL I		60/0	72	4	0	1
CSI030	PROGRAMACAO DE COMPUTADORES I		60/0	72	2	2	1
CSI201	INTRODUCAO A ENGENHARIA DE COMPUTACAO		30/0	36	2	0	1
CSI427	METODOLOGIA DE PESQUISA APLICADA A COMPUTACAO		30/0	36	2	0	1
EAD707	COMUNICACAO E EXPRESSAO		30/0	36	2	0	1
CEA051	CALCULO DIFERENCIAL E INTEGRAL II	CEA049 CEA050	60/0	72	4	0	2
CEA055	ESTADISTICA E PROBABILIDADE	CEA050	60/0	72	4	0	2
CEA061	FISICA I	CEA049 CEA050	60/0	72	3	1	2
CSI032	PROGRAMACAO DE COMPUTADORES II	CSI030	60/0	72	2	2	2
CSI443	MATEMATICA DISCRETA		60/0	72	4	0	2
CEA052	CALCULO DIFERENCIAL E INTEGRAL III	CEA051	60/0	72	4	0	3
CEA056	INTRODUCAO AS EQUACOES DIFERENCIAIS ORDINARIAS	CEA051	60/0	72	4	0	3
CEA062	FISICA II	CEA038 CEA039 CEA051 CEA061	60/0	72	3	1	3
CEA341	PRINCIPIOS DE ELETRONICA DIGITAL	CSI030	60/0	72	2	2	3
CSI488	ALGORITMOS E ESTRUTURA DE DADOS I	CSI030	60/0	72	3	1	3
CEA063	FISICA III	CEA038 CEA039 CEA051 CEA055 CEA061	60/0	72	3	1	4
CEA552	CIRCUITOS ELETRICOS I	CEA056 CEA062	60/0	72	2	2	4
CEA562	SINAIS E SISTEMAS	CEA056	60/0	72	2	2	4
CSI429	ALGORITMOS E ESTRUTURA DE DADOS II	CSI032 CSI488	60/0	72	4	0	4
CSI466	TEORIA DOS GRAFOS	CSI443 CSI488	60/0	72	4	0	4
CEA064	FISICA IV	CEA052 CEA056 CEA062 CEA063	60/0	72	3	1	5
CEA502	ELETROMAGNETISMO	CEA052 CEA063 CEA552	60/0	72	2	2	5
CSI148	ANALISE NUMERICA	CEA056 CSI030	60/0	72	2	2	5
CSI203	ORGANIZACAO E ARQUITETURA DE COMPUTADORES I	CEA341 CSI030	60/0	72	2	2	5
CSI485	ENGENHARIA DE SOFTWARE I	CSI032 CSI488	60/0	72	4	0	5
CEA561	ELETRONICA I	CEA552	60/0	72	2	2	6
CEA582	FUNDAMENTOS DE COMUNICACOES	CEA562	60/0	72	2	2	6
CSI440	BANCO DE DADOS I	CSI488	60/0	72	4	0	6
CSI509	ORGANIZACAO E ARQUITETURA DE COMPUTADORES II	CSI203	60/0	72	4	0	6
CSI546	PROJETO E ANALISE DE ALGORITMO	CSI466	60/0	72	4	0	6
CEA551	MODELAGEM E ANALISE DE SISTEMAS LINEARES	CEA064 CEA552 CSI148	60/0	72	2	2	7
CEA580	MICROPROCESSADORES E MICROCONTROLADORES	CSI203	60/0	72	2	2	7
CSI437	SISTEMAS OPERACIONAIS	CSI203 CSI488	60/0	72	4	0	7
CSI476	FUNDAMENTOS TEORICOS DA COMPUTACAO	CSI466	60/0	72	4	0	7
CSI508	COMPUTACAO GRAFICA	CSI488	60/0	72	4	0	7
CSI419	LINGUAGENS DE PROGRAMACAO	CSI476	60/0	72	4	0	8
CSI450	INTERACAO HUMANO-COMPUTADOR	CSI485	60/0	72	4	0	8
CSI457	INTELIGENCIA ARTIFICIAL	CSI466	60/0	72	4	0	8
CSI548	REDES DE COMPUTADORES I	CSI488	60/0	72	2	2	8
EAD344	TEORIA DA ADMINISTRACAO I		60/0	72	4	0	8
CSI433	SISTEMAS DISTRIBUIDOS	CSI548	60/0	72	4	0	9
CSI495	TRABALHO DE CONCLUSAO DE CURSO I	2200 horas	90/0	108	0	6	9
CSI506	COMPILADORES I	CSI476	60/0	72	4	0	9
CSI510	REDES DE COMPUTADORES II	CSI548	60/0	72	3	1	9
EAD351	FILOSOFIA E ETICA		60/0	72	4	0	9



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
PRÓ-REITORIA DE GRADUAÇÃO



CÓDIGO	DISCIPLINAS OBRIGATÓRIAS	PRÉ-REQUISITO	CHS/CHE	CHA	AULAS		PER
					T	P	
CEA066	FUNDAMENTOS DE CIENCIA DO AMBIENTE	CEA038 CEA039 CEA062 CEA063	30/0	36	2	0	10
CSI491	INFORMATICA E SOCIEDADE		30/0	36	2	0	10
CSI496	TRABALHO DE CONCLUSAO DE CURSO II	CSI495	60/0	72	0	4	10
CSI693	AVALIACAO DE DESEMPENHO DE SISTEMAS	CSI433	60/0	72	4	0	10

CÓDIGO	DISCIPLINAS ELETIVAS	PRÉ-REQUISITO	CHS/CHE	CHA	AULAS		PER
					T	P	
CEA026	DESENHO COMPUTACIONAL		60/0	72	2	2	
CEA057	METODOS MATEMATICOS APLICADOS A ENGENHARIA ELETRICA	CEA052 CEA056	60/0	72	4	0	
CEA147	ELETRONICA EMBARCADA	CEA580	60/0	72	2	2	
CEA453	SISTEMAS MULTIMIDIA	CEA562	60/0	72	3	1	
CEA563	CIRCUITOS ELETRICOS II	CEA552	60/0	72	2	2	
CEA570	SISTEMAS DE CONTROLE I	CEA562	60/0	72	2	2	
CEA571	ELETRONICA II	CEA551 CEA561	60/0	72	2	2	
CEA572	INSTRUMENTACAO	CEA561 CEA570	60/0	72	2	2	
CEA671	PROCESSAMENTO DIGITAL DE SINAIS	CEA562	60/0	72	3	1	
CEA708	PRINCIPIOS DAS COMUNICACOES DIGITAIS	CEA562	60/0	72	3	1	
CEA709	PROPAGACAO DE ONDAS DE RADIO	CEA502 CEA562	60/0	72	3	1	
CEA710	PROBABILIDADE E PROCESSOS ESTOCASTICOS PARA ENGENHARIA		60/0	72	3	1	
CEA711	REDES E SISTEMAS DE COMUNICACOES MOVEIS	CEA341 CSI488	60/0	72	4	0	
CEA712	TOPICOS ESPECIAIS EM TELECOMUNICACOES	CEA562	60/0	72	3	1	
CSI001	ANALISE DE MIDIAS SOCIAIS	CSI466	60/0	72	4	0	
CSI401	PROGRAMACAO PARA DISPOSITIVOS MOVEIS		60/0	72	1	3	
CSI402	COMPUTACAO MOVEI		60/0	72	4	0	
CSI434	TOPICOS EM REDES DE COMPUTADORES	CSI548	60/0	72	4	0	
CSI436	COMPUTACAO MOVEI	CSI440	60/0	72	4	0	
CSI442	BANCO DE DADOS II	CSI429 CSI440	60/0	72	4	0	
CSI472	INTELIGENCIA COMPUTACIONAL PARA OTIMIZACAO	CSI488	60/0	72	4	0	
CSI477	SISTEMAS WEB I	CSI440 CSI485 CSI548	60/0	72	2	2	
CSI486	ENGENHARIA DE SOFTWARE II	CSI485	60/0	72	4	0	
CSI501	INTRODUCAO A ROBOTICA	CEA341 CSI030	60/0	72	2	2	
CSI502	LINGUAGEM DE DESCRICAO DE HARDWARE	CEA341 CSI030	60/0	72	2	2	
CSI503	DESENVOLVIMENTO AGIL DE SOFTWARE	CSI485	60/0	72	2	2	
CSI512	TOPICOS EM ENGENHARIA DE SOFTWARE I	CSI485	60/0	72	4	0	
CSI513	TOPICOS EM DESENVOLVIMENTO DE SISTEMAS I	CSI440 CSI485	60/0	72	4	0	
CSI515	PROCESSAMENTO DIGITAL DE IMAGENS	CEA341 CEA562 CSI488	60/0	72	3	1	
CSI516	TOPICOS EM DESENVOLVIMENTO DE SISTEMAS II	CSI486	60/0	72	4	0	
CSI518	TOPICOS EM ENGENHARIA DE SOFTWARE II	CSI486	60/0	72	4	0	
CSI520	TOPICOS EM INTELIGENCIA ARTIFICIAL	CSI032 CSI488	60/0	72	4	0	
CSI549	AVALIACAO DE SISTEMAS INTERATIVOS	CSI450	60/0	72	4	0	
CSI557	COMPUTACAO EVOLUCIONARIA	CSI488	60/0	72	2	2	
CSI558	DESENVOLVIMENTO DIRIGIDO POR MODELOS	CSI485	60/0	72	4	0	
CSI559	FUNDAMENTOS DE LINGUAGENS DE PROGRAMACAO	CSI419	60/0	72	4	0	
CSI564	GEOPROCESSAMENTO E SISTEMA DE INFORMACAO GEOGRAFICA	900 horas CSI440	60/0	72	4	0	
CSI566	MODELAGEM E SIMULACAO DE SISTEMAS TERRESTRES	900 horas CSI032 CSI488	60/0	72	4	0	
CSI567	PADROES DE PROJETOS	CSI485	60/0	72	3	1	
CSI568	QUALIDADE DE SOFTWARE	CSI485	60/0	72	4	0	
CSI569	SEMANTICA FORMAL	CSI419	60/0	72	4	0	
CSI574	SISTEMAS DE TIPOS	CSI419	60/0	72	4	0	
CSI575	SISTEMAS WEB II	CSI477 CSI486	60/0	72	2	2	
CSI576	TESTE DE SOFTWARE	CSI485	60/0	72	4	0	
CSI577	LOGICA APLICADA A COMPUTACAO	CSI443 CSI488	60/0	72	4	0	
CSI578	TOPICOS EM LINGUAGEM DE PROGRAMACAO	CSI419	60/0	72	4	0	



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
PRÓ-REITORIA DE GRADUAÇÃO



CÓDIGO	DISCIPLINAS ELETIVAS	PRÉ-REQUISITO	CHS/CHE	CHA	AULAS		PER
					T	P	
CSI604	COMPILADORES II	CSI506	60/0	72	4	0	
CSI730	TOPICOS EM SISTEMAS DE COMPUTACAO	CSI032 CSI488	60/0	72	4	0	
CSI731	BANCO DE DADOS GEOGRAFICOS	CSI440	60/0	72	4	0	
CSI732	ENGENHARIA ONTOLOGICA E LOGICA FUZZY	CSI419 CSI457	60/0	72	4	0	
CSI733	APRENDIZAGEM DE MAQUINA	CSI457	60/0	72	4	0	
CSI734	INTERNET DAS COISAS	CSI548	60/0	72	4	0	
CSI735	GERENCIA DE PROJETOS DE SOFTWARE	CSI440 CSI485	60/0	72	4	0	
ENP151	CIENCIA TECNOLOGIA E SOCIEDADE		60/0	72	4	0	
ENP152	ETICA E RESPONSABILIDADE SOCIOAMBIENTAL		60/0	72	4	0	
ENP153	PROGRAMACAO LINEAR	CSI030 CSI488	60/0	72	3	1	
ENP160	OTIMIZACAO COMBINATORIA	ENP153	60/0	72	3	1	
ENP493	EMPREENDEDORISMO	1800 horas	60/0	72	4	0	
LET966	INTRODUCAO A LIBRAS		60/0	72	2	2	
TELECOMUNICAÇÕES							
CEA453	SISTEMAS MULTIMIDIA	CEA562	60/0	72	3	1	
CEA708	PRINCIPIOS DAS COMUNICACOES DIGITAIS	CEA562	60/0	72	3	1	
CEA709	PROPAGACAO DE ONDAS DE RADIO	CEA502 CEA562	60/0	72	3	1	
CEA710	PROBABILIDADE E PROCESSOS ESTOCASTICOS PARA ENGENHARIA		60/0	72	3	1	
CEA711	REDES E SISTEMAS DE COMUNICACOES MOVEIS	CEA341 CSI488	60/0	72	4	0	
CEA712	TOPICOS ESPECIAIS EM TELECOMUNICACOES	CEA562	60/0	72	3	1	
CSI515	PROCESSAMENTO DIGITAL DE IMAGENS	CEA341 CEA562 CSI488	60/0	72	3	1	
ENGENHARIA DE SOFTWARE							
CSI436	COMPUTACAO MOVEL	CSI440	60/0	72	4	0	
CSI442	BANCO DE DADOS II	CSI429 CSI440	60/0	72	4	0	
CSI472	INTELIGENCIA COMPUTACIONAL PARA OTIMIZACAO	CSI488	60/0	72	4	0	
CSI477	SISTEMAS WEB I	CSI440 CSI485 CSI548	60/0	72	2	2	
CSI486	ENGENHARIA DE SOFTWARE II	CSI485	60/0	72	4	0	
CSI512	TOPICOS EM ENGENHARIA DE SOFTWARE I	CSI485	60/0	72	4	0	
CSI513	TOPICOS EM DESENVOLVIMENTO DE SISTEMAS I	CSI440 CSI485	60/0	72	4	0	
CSI516	TOPICOS EM DESENVOLVIMENTO DE SISTEMAS II	CSI486	60/0	72	4	0	
CSI518	TOPICOS EM ENGENHARIA DE SOFTWARE II	CSI486	60/0	72	4	0	
CSI549	AVALIACAO DE SISTEMAS INTERATIVOS	CSI450	60/0	72	4	0	
CSI557	COMPUTACAO EVOLUCIONARIA	CSI488	60/0	72	2	2	
CSI558	DESENVOLVIMENTO DIRIGIDO POR MODELOS	CSI485	60/0	72	4	0	
CSI564	GEOPROCESSAMENTO E SISTEMA DE INFORMACAO GEOGRAFICA	900 horas CSI440	60/0	72	4	0	
CSI567	PADROES DE PROJETOS	CSI485	60/0	72	3	1	
CSI568	QUALIDADE DE SOFTWARE	CSI485	60/0	72	4	0	
CSI575	SISTEMAS WEB II	CSI477 CSI486	60/0	72	2	2	
CSI576	TESTE DE SOFTWARE	CSI485	60/0	72	4	0	

CÓDIGO	ATIVIDADES	PRÉ-REQUISITO	CARÁTER	CHS	PER
ATV100	ATIVIDADE ACADEMICO CIENTIFICO-CULTURAL		OBRIGATORIA	190	
ATV500	ESTAGIO CURRICULAR OBRIGATORIO	1500	OBRIGATORIA	200	



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
PRÓ-REITORIA DE GRADUAÇÃO



Componentes Curriculares Exigidos para Integralização no Curso	Carga Horária
Disciplinas Obrigatórias	2940
Disciplinas Eletivas	300
Disciplinas Optativas	0
Disciplinas Facultativas	0
Atividades	390
Extensão	0
Total	3630