

Aplicando Mineração de Dados Educacionais para análise de Evasão das Alunas de Graduação em Curso de Computação da UnB

Alice da Silva de Lima¹, Maristela Terto de Holanda¹

¹Departamento de Ciência da Computação (CIC) – Universidade de Brasília (UnB)
Brasília, DF – Brasil

alice.lima@aluno.unb.br, mholanda@unb.br

Resumo. *A evasão feminina em cursos de computação tem como um dos fatores o desempenho em disciplinas iniciais desses cursos. Nesse contexto, a presente pesquisa tem por objetivo analisar o desempenho das alunas do Departamento de Ciência da Computação da Universidade de Brasília nas disciplinas Algoritmos e Programação de Computadores e Computação Básica a partir de uma mineração de dados educacionais utilizando um dataset.*

1. Introdução

A diversidade de gênero nos cursos do Departamento de Ciência da Computação na Universidade de Brasília ainda apresenta uma desigualdade notável [1]. Além do número de ingressantes do gênero feminino ser geralmente baixo em relação ao gênero masculino, um fator que intensifica ainda mais esse problema é a evasão. Dentre os diversos fatores que podem levar a evasão nesses cursos, tem-se a dificuldade que as alunas encontram em determinadas disciplinas, sobretudo aquelas do primeiro ano de curso. A disciplina Algoritmos e Programação de Computadores, ofertada pelo departamento, é uma das principais no período inicial desses cursos: esta apresenta conceitos iniciais de programação, conceito de algoritmos, e além disso é um pré-requisito para Estruturas de Dados, disciplina preponderante nos cursos de computação, que também acaba sendo um pré-requisito para diversas outras matérias do currículo.

A disciplina tem a finalidade introduzir os conceitos iniciais de programação e apresentar os elementos básicos para a criação de um programa (tipos de dados, condicionais, laços de repetição, funções e recursividade), bem como o conceito de algoritmo [2].

Nesse contexto, a presente pesquisa tem por objetivo analisar o desempenho das alunas do departamento nas disciplinas Algoritmos e Programação de Computadores e Computação Básica a partir de uma mineração de dados educacionais. A disciplina Computação Básica não é mais ofertada atualmente, mas uma vez que esta é equivalente a Algoritmos e Programação de Computadores, foi considerada nessa pesquisa. Foi utilizado um *dataset* com dados educacionais dos alunos do departamento para a realização do trabalho. O resto do artigo está dividido da seguinte maneira: Seção 2, metodologia utilizada; Seção 3, a análise dos resultados e por fim, Seção 4, que apresenta a conclusão da pesquisa.

2. Metodologia

O *dataset* utilizado para a análise dos dados tem como origem o SIGRA (Sistema Integrado de Gestão das Atividades Acadêmicas). O conjunto inicial apresentava em cada linha informações gerais de um aluno, período de ingresso na universidade, período de ingresso no curso, nome do curso, e informações sobre uma disciplina específica cursada pelo aluno. Uma vez que essa pesquisa tem por objetivo analisar o desempenho das alunas dos cursos Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia Mecatrônica nas disciplinas Algoritmos e Programação de Computadores e Computação Básica, foi necessário limpar os dados para filtrar as informações relevantes para a análise. As ferramentas utilizadas para a limpeza dos dados e a criação das visualizações foram linguagem Python e o Microsoft Power BI.

As questões de pesquisa definidas foram:

- QP1: Qual o quantitativo de alunas e alunos matriculadas nas duas disciplinas por ano?
- QP2: Como está a distribuição das menções das alunas por ano?

Os critérios de exclusão aplicados foram os seguintes:

- CE1: Informações sobre estudantes de cursos que não fossem Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia Mecatrônica;
- CE2: Informações que não fossem a respeito da disciplina Algoritmos e Programação de Computadores ou Computação Básica;
- CE3: Informações que não fossem dos anos de 2010 a 2019.

Uma vez que a diferença do número de estudantes por gênero é consideravelmente desigual, é inviável considerar certas métricas para comparar o desempenho dos alunos por gênero diretamente.

O número de colunas do *dataset* também foi reduzido, a fim de manter apenas os campos de interesse para essa pesquisa, sendo estes:

- *id_aluno*: identificador de cada aluno;
- *gênero*: um caractere identificador do gênero do aluno;
- *curso*: nome do curso;
- *periodo_cursou_disciplina*: semestre em que o estudante cursou uma das duas disciplinas;
- *nome_disciplina*: identificador da disciplina (Algoritmos e Programação de Computadores ou Computação Básica);
- *menção_disciplina*: caracteres que indicam a menção do aluno na disciplina.
- *ano_cursou*: coluna criada após a limpeza dos dados que deriva da *periodo_cursou_disciplina*, a fim de facilitar a contagem dos dados por ano.

A próxima seção responde as questões de pesquisa.

3. Resultados

Esta seção tem por objetivo responder as questões de pesquisa, realizando a análise dos dados. Nesta seção, ambas as disciplinas Algoritmos e Programação de Computadores e Computação Básica serão referidas como "disciplina inicial de programação".

3.1. QP1: Qual o quantitativo de alunas e alunos matriculados nas duas disciplinas por ano?

A Figura 1 apresenta o número de estudantes de Ciência da Computação matriculados nas disciplinas de Algoritmos e Programação de Computadores ou Computação Básica de 2010 a 2019. É possível perceber que a diferença de gênero é alta durante todos os anos, sendo o número de estudantes do gênero feminino matriculadas inferior a metade do número de estudantes do gênero masculino, o que é um reflexo da baixa diversidade de gênero no curso. Com 20 matriculadas, o ano de 2018 teve o maior número de matriculadas nas disciplinas. Já o ano de 2016 registrou o menor número de matriculadas, com apenas 6 alunas.

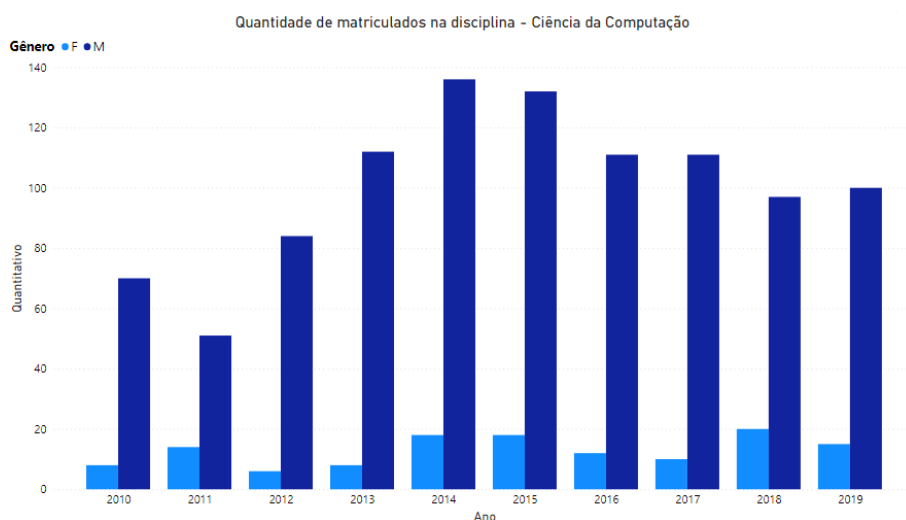


Figura 1. Relação de estudantes do curso Ciência da Computação matriculadas em disciplina inicial de programação por ano.

Para o curso de Licenciatura em Computação a presença feminina nessas disciplinas foi superior, como mostra a Figura 2. No ano de 2016 o quantitativo de alunas foi 25 (15,43%), enquanto o número de alunos do gênero masculino era 137. Já os anos de 2011 e 2017 registraram o menor número de meninas matriculadas, totalizando 7, o que representava 8,53% em 2011 e 5,26% em 2017.

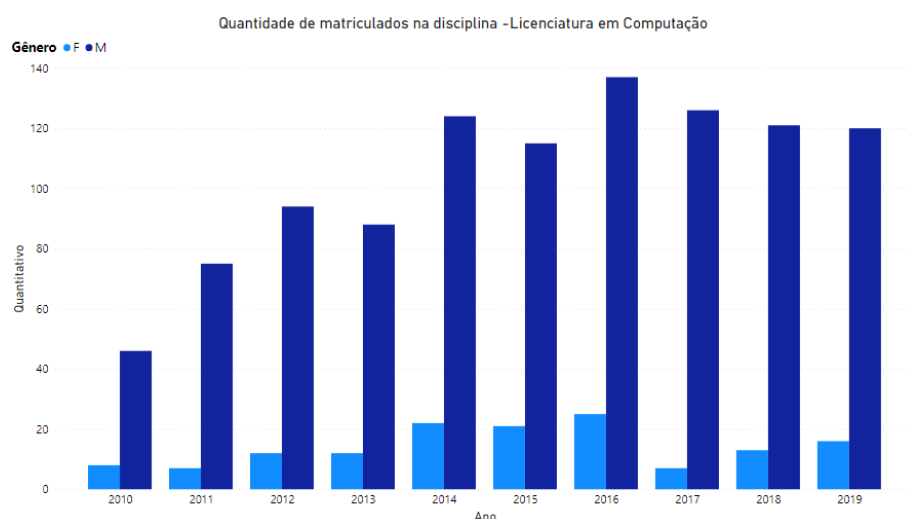


Figura 2. Relação de estudantes do curso Licenciatura em Computação matriculadas em disciplina inicial de programação por ano.

É possível observar que no contexto do curso de Engenharia de Computação houve um aumento no número de alunas cursando as disciplinas do ano de 2015 (6 alunas) até o ano de 2019, sendo que este último registrou o maior número de matriculadas com 26 alunas, 21,67% do total. Observando por outra perspectiva, considerando o número de matriculados dos dois gêneros, percebe-se que nos dois últimos anos o número de alunos cursando as disciplinas diminuiu se comparado aos quatro anos anteriores.

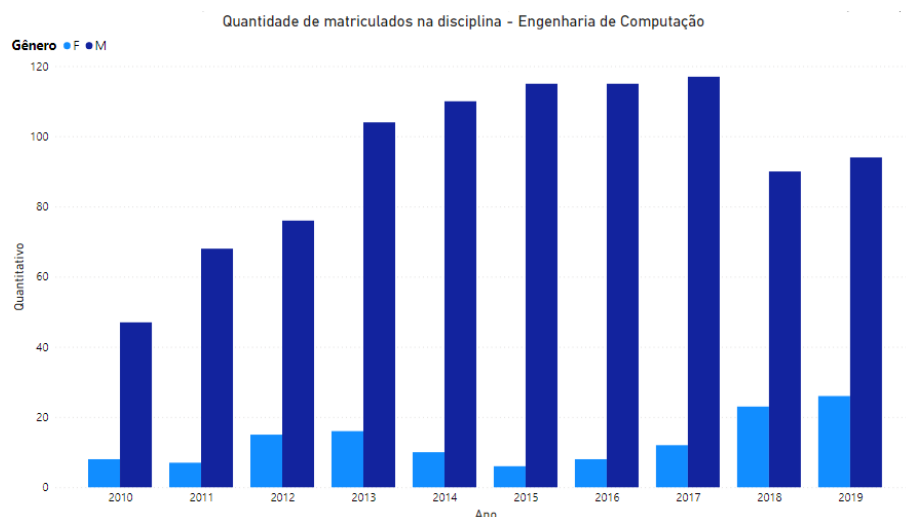


Figura 3. Relação de estudantes do curso Engenharia de Computação matriculadas em disciplina inicial de programação por ano.

A respeito dos dados da Engenharia Mecatrônica, observa-se na Figura 4 que o período com mais alunas cursando as disciplinas foi ano de 2016 com 24 alunas (20%), e o período com o menor número de alunas o ano de 2011, contendo apenas 4 alunas matriculadas junto a 82 do gênero masculino. Pode ser observado que no período de 2015 a 2019 o número de matriculadas manteve-se igual ou superior a vinte.

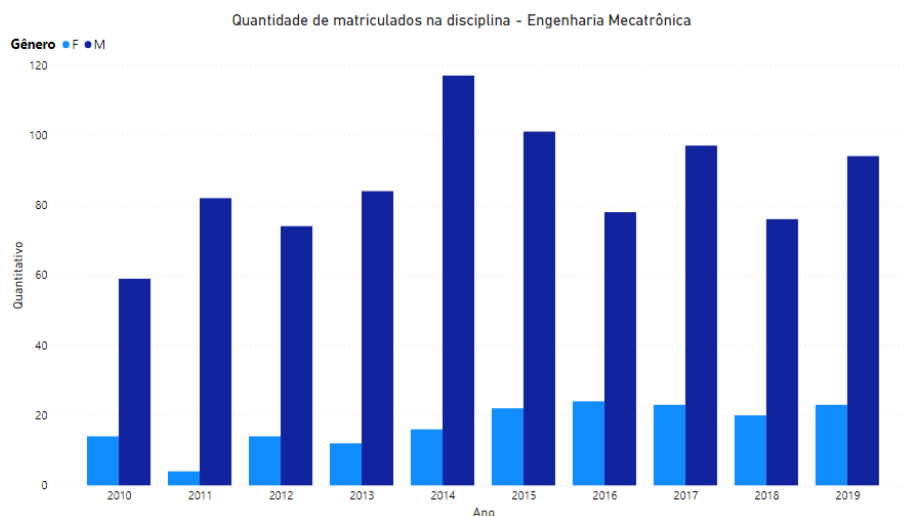


Figura 4. Relação de estudantes do curso Engenharia Mecatrônica em Computação matriculadas em disciplina inicial de programação por ano.

3.2. QP2: Como está a distribuição das menções das alunas por ano?

As figuras a seguir ilustram o desempenho das alunas de cada curso nas disciplinas. Uma vez que o *dataset* não apresentou a média final, não foi possível apresentar essa métrica especificamente. A análise foi realizada considerando a distribuição das menções.

Uma menção é atribuída ao aluno de acordo com um intervalo de notas, sendo N a nota final:

Tabela 1. Definição das menções

Nota	Menção
$9.0 \leq N \leq 10.0$	SS
$7.0 \leq N < 9.0$	MS
$5.0 \leq N < 7.0$	MM
$3.0 \leq N < 5.0$	MI
$0.0 < N < 3.0$	II
Acima de 25% de faltas (ou $N = 0$)	SR

Além disso, a menção também pode se configurar em três casos especiais:

- CC – Crédito Concedido, ocorre quando o estudante pede o aproveitamento da matéria, pois já cursou posteriormente em outra instituição;
- TR – Caso onde há trancamento da matéria;
- TJ – Caso onde há trancamento justificado da matéria.

É importante ressaltar que menções inferiores a MM acarretam em reprovação.

A Figura 5 apresenta a distribuição de menções das alunas de Ciência da Computação. De maneira geral, considerando todos os anos a menção mais recorrente é MS. Os anos 2010, 2015 e 2019 representam bem esse cenário. Nos anos 2012, 2016 e 2017 nenhuma estudante conseguiu menção SS na disciplina, isto é, teve a média final igual ou superior a nove.

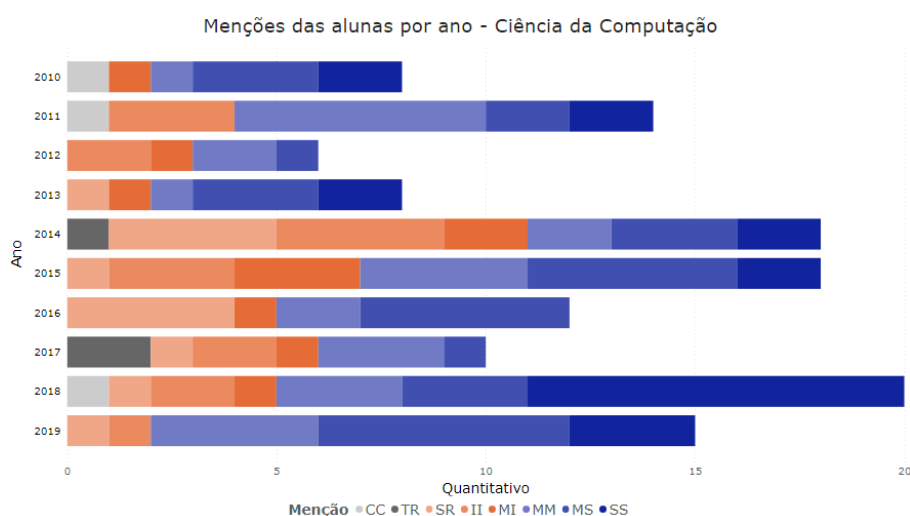


Figura 5. Menções das alunas de Ciência da Computação na matéria inicial de programação por ano.

Também é importante se atentar ao ano de 2014, em que apesar da distribuição de menções ser homogênea se comparado a outros anos, esse período registrou o maior número de menções II, com 4 alunas, sendo este também o quantitativo da menção no ano de 2016. Já o ano de 2018 apresenta o melhor desempenho, onde 9 alunas receberam a menção SS, além de que foram 15 aprovadas no total, e 4 reprovações. Esses resultados considerando um ano inteiro são positivos, a tabela abaixo mostra os resultados por semestre deste ano:

Tabela 2. Desempenho das alunas (Ciência da Computação) na disciplina Algoritmo e Programação de Computadores no ano de 2018 por semestre.

Semestre	Porcentagem de aprovadas	Porcentagem de reprovadas
2018/0	100%	0%
2018/1	75%	25%
2018/2	77.8%	22.2%

Observações: 2018/0 refere-se ao período das aulas de verão, 2018/1 o primeiro semestre letivo, 2018/2 o segundo.

A respeito do curso Licenciatura em Computação, primeiramente observa-se que em 4 anos não se tem registro da menção SS em nenhum dos semestres. Especificamente no ano de 2018 também não há registros de menções MS, o que implica que todas as alunas aprovadas na disciplina obtiveram nota mediana. A menção mais recorrente considerando os dez anos é MM, sendo essa a menção mínima para aprovação. Desconsiderando os casos de trancamento e créditos concedidos, pode-se afirmar que o ano com maior aprovação foi o de 2017, com 71,43% aprovadas. O ano seguinte, 2018, apresentou o cenário inverso: foram 69,23% foram reprovadas e 30,77% foram aprovadas.

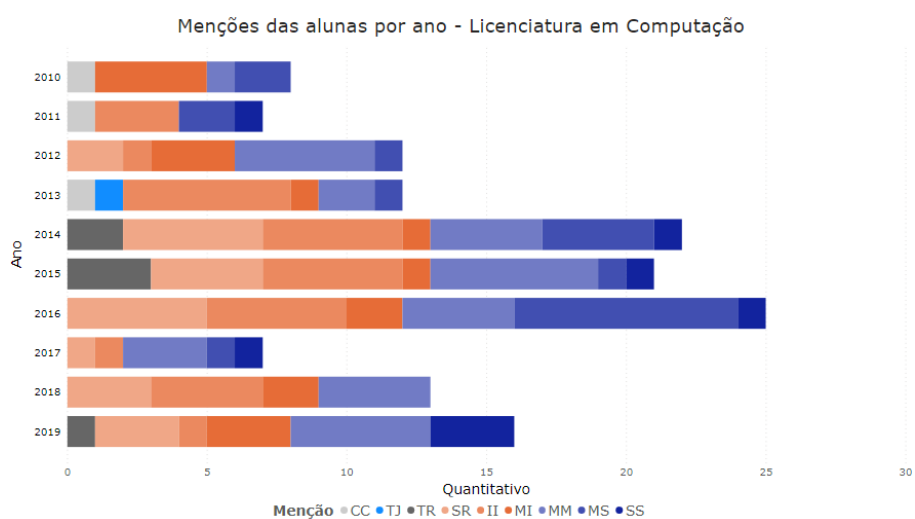


Figura 6. Menções das alunas de Licenciatura em Computação na matéria inicial de programação por ano.

O curso de Engenharia de Computação passou a ser ofertado pela universidade no segundo semestre de 2009, o que pode ser uma das justificativas para o baixo número de matrículas nas disciplinas nos dois anos iniciais, isso também é observado no ano de 2015 (6 alunas). A ausência da menção SS se dá nos anos 2011 e 2017. O ano de 2010 registrou 100% de aprovação feminina na disciplina.

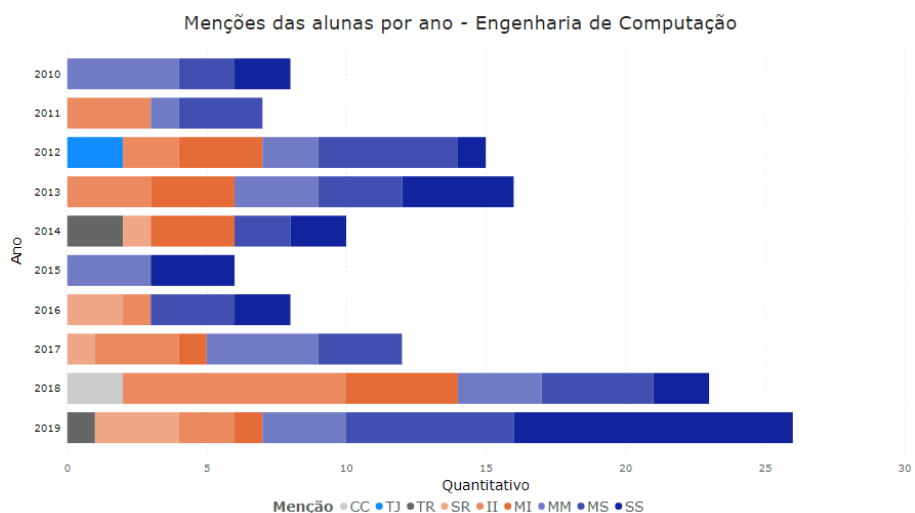


Figura 7. Menções das alunas de Engenharia de Computação na matéria inicial de programação por ano.

A Figura 8 apresenta os dados de Engenharia Mecatrônica. É notável que os últimos cinco anos além de apresentarem os maiores números de matriculadas, também apresentam mais variedade de menções. O único ano sem menções SS é o de 2018 e é possível notar um número notável de trancamentos neste mesmo ano e em 2019. As menções MS e SS foram as mais recorrentes nesse curso, totalizando 40 menções cada. Analisando as alunas que não trancaram a disciplina e nem pediram o aproveitamento de créditos, o ano de melhor desempenho é o de 2010, com 92,86% aprovadas.

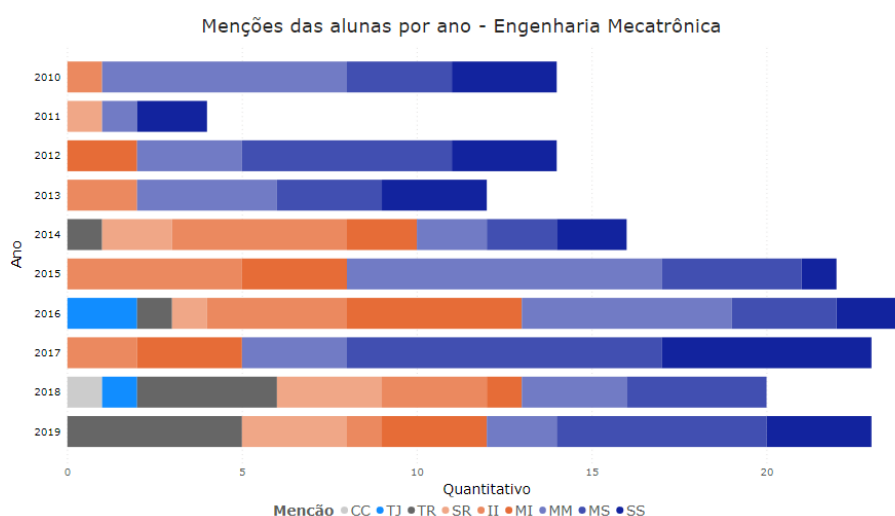


Figura 8. Menções das alunas de Engenharia Mecatrônica na matéria inicial de programação por ano.

A Tabela abaixo apresenta o quantitativo de semestres com 100% de aprovação nos quatro cursos:

Tabela 3. Períodos com 100% de aprovação na disciplina inicial de programação por curso.

Curso	Quantitativo	Períodos
Ciência da Computação	6	2010/1 - 2011/2 - 2017/1 - 2018/0 - 2019/0 - 2019/2
Licenciatura em Computação	2	2017/1 - 2018/0
Engenharia de Computação	6	2010/1 - 2010/2 - 2015/1 - 2017/1 - 2018/0 - 2019/0
Engenharia Mecatrônica	4	2010/1 - 2011/2 - 2012/1 - 2013/1

Embora os dados apresentem em alguns períodos apenas menções de aprovação e trancamentos, foram considerados períodos com 100% de aprovação aqueles que apresentaram menções de aprovação (MM, MS, SS) e créditos concedidos (CC).

4. Conclusão

A partir da análise realizada, é possível concluir que o número de matrículas por ano nas disciplinas iniciais de programação tem um número desproporcional de estudantes do gênero feminino em relação do gênero masculino, o que está ligado à baixa diversidade de gênero presente nos cursos de computação. Esse fator influencia a porcentagem de reprovação das alunas, por ter um número pequeno de matriculadas. Apesar disso, pode ser observado que a maioria das menções foram MM e MS, considerando a contagem geral dos dez anos. Todavia, também houve um número notável de reprovações durante os dez anos, além de alguns trancamentos.

Para cada curso há períodos distintos de melhor desempenho, todavia foi observado que nos períodos de aulas de verão o desempenho foi bastante positivo em todos os cursos, o que poderia ser justificado pelo fato de que um número limitado de disciplinas são ofertadas nessa época, limitando a quantidade de matérias em que um aluno pode se matricular, o que acarreta em mais tempo de dedicação para a disciplina.

Uma limitação da pesquisa foi a ausência dos dados do ano de 2020, que seria bastante importante para entender como está o cenário mais recente. O *dataset* também não apresentou as notas dos discentes, apenas as menções, impossibilitando analisar o desempenho das alunas utilizando outras métricas estatísticas.

O departamento de Ciência da Computação já vem tomando iniciativas para diminuir a taxa de reprovação de Algoritmos e Programação de Computadores através da organização de monitorias e tutorias como pode ser observado em [3], além de melhoras nos materiais didáticos implantados na disciplina. É importante persistir com essas iniciativas para aumentar as aprovações e menções máximas, contribuindo também para diminuir a evasão.

Referências

- [1] Holanda, Maristela, et al. "Iniciativas Educacionais para Permanência das Mulheres em Cursos de Graduação em Computação no Brasil."(2020).
- [2] ALGORITMOS E PROGRAMAÇÃO DE COMPUTADORES. Disponível em: <https://matriculaweb.unb.br/graduacao/disciplina.aspx?cod=113476>
- [3] SSaturday - Programa de Monitoria de APC. Disponível em: <https://cic.unb.br/ssaturday-programa-de-monitoria-de-apc/>