

MARIA SÍLVIA RIBEIRO RUY

ANÁLISE DE DADOS ACADÊMICOS PARA PREVISÃO DA DEMANDA DE TURMAS

(versão pré-defesa, compilada em 8 de agosto de 2023)

Trabalho apresentado como requisito parcial à conclusão do Curso de Bacharelado em Ciência da Computação, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: Bruno Müller Junior.

CURITIBA PR

2023

RESUMO

Este trabalho se dedica à implementação e calibração de um algoritmo de previsão de demanda de turmas para o curso de Bacharelado em Ciência da Computação da Universidade Federal do Paraná. O objetivo é auxiliar a Coordenação do curso a determinar a quantidade necessária de turmas que devem ser solicitadas para cada disciplina. É descrito um algoritmo que estima a quantidade de alunos que vão querer cursar cada disciplina no próximo semestre. Em seguida, foram parametrizados pontos do algoritmo e experimentados diferentes valores, a fim de obter melhores resultados e analisar o comportamento do algoritmo diante das variações. Este estudo explora uma solução de gestão acadêmica para uma melhor previsão e planejamento das turmas, evitando a superlotação ou subutilização de recursos.

Palavras-chave: previsão. demanda. gestão acadêmica.

ABSTRACT

This paper is dedicated to the implementation and calibration of a class demand forecasting algorithm for the Bachelor in Computer Science degree program at the Federal University of Paraná. The objective is to assist the program coordinators in determining the number of classes necessary for each course. First it is described an algorithm that estimates the number of students who will want to take each course in the next semester. Then, algorithm parameters were set and different values were experimented in order to achieve better results and analyze the behavior of the algorithm under the variations. This study explores an academic management solution for improved class forecasting and planning, avoiding overcrowding or underutilization of resources.

Keywords: forecasting. demand. academic management.

LISTA DE FIGURAS

2.1	Modelo esquemático do ADEGA.	10
3.1	Exemplo dos limites da janela para aluno cursando o 5º período. Células verdes significam aprovado e e amarelas são matrículas em curso.	13
3.2	Exemplo dos limites da janela para aluno cursando o 4 e 5º período. Células verdes significam aprovado e e amarelas são matrículas em curso.	14
4.1	Gráficos de Dispersão. Dados da ocupação real estão como quadrados azuis para referência. As disciplinas estão dispostas no eixo X na mesma ordem da grade acadêmica.	20
4.2	Gráficos de Erro Absoluto. Disciplinas ordenadas de acordo com um dos Ns do experimento.	21

LISTA DE TABELAS

4.1	Métricas comparando os experimentos. 4 e 5 resultam no menor erro. (Experimentos variando o valor de MAXIMO MATRICULAS. Disciplinas não-ofertadas ignoradas na análise)	18
4.2	Métricas comparando os experimentos. (Experimentos variando o valor de MAXIMO MATRICULAS. Disciplinas não-ofertadas ignoradas na análise) . . .	18
4.3	Métricas para quando experimenta-se diferentes tamanhos de janela. (Experimentos variando o valor de EXTENSAO JANELA. Disciplinas não-ofertadas ignoradas na análise)	19
A.1	Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 3 e EXTENSAO JANELA = 1 (todas as disciplinas consideradas)	24
A.2	Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 4 e EXTENSAO JANELA = 1 (todas as disciplinas consideradas)	25
A.3	Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 5 e EXTENSAO JANELA = 1 (todas as disciplinas consideradas)	26
A.4	Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 6 e EXTENSAO JANELA = 1 (todas as disciplinas consideradas)	27
A.5	Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 3 e EXTENSAO JANELA = 1 (desconsidera disciplinas não-ofertadas)	28
A.6	Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 4 e EXTENSAO JANELA = 1 (desconsidera disciplinas não-ofertadas)	29
A.7	Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 5 e EXTENSAO JANELA = 1 (desconsidera disciplinas não-ofertadas)	30
A.8	Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 6 e EXTENSAO JANELA = 1 (desconsidera disciplinas não-ofertadas)	31
A.9	Tabela com os resultados do algoritmo para EXTENSAO JANELA = 0 e MAXIMO MATRICULAS = 5 (desconsidera disciplinas não-ofertadas)	32
A.10	Tabela com os resultados do algoritmo para EXTENSAO JANELA = 2 e MAXIMO MATRICULAS = 5 (desconsidera disciplinas não-ofertadas)	33
B.1	Tabela com as matrículas reais de 2023/1	34

LISTA DE ACRÔNIMOS

UFPR	Universidade Federal do Paraná
BCC	Bacharelado em Ciência da Computação
PET	Programa de Educação Tutorial
ADEGA	Análise de Dados Estatísticos de Grade Acadêmica
MVC	Modelo-Visão-Controlador
EMA	Erro Médio Absoluto
REQM	Raiz do Erro Quadrático Médio

SUMÁRIO

1	INTRODUÇÃO	8
2	ADEGA	10
3	ALGORITMO	11
3.1	DISCIPLINAS PENDENTES DO ALUNO	13
3.2	ELEGIBILIDADE NA DISCIPLINA	14
3.3	ASPECTOS DA IMPLEMENTAÇÃO	15
4	RESULTADOS E EXPERIMENTOS	16
4.1	PARÂMETROS.	16
4.2	BASE DE DADOS	16
4.3	MÉTRICAS	16
4.4	VARIANDO A QUANTIDADE MÁXIMA DE MATRÍCULAS POR ALUNO .	17
4.4.1	Experimentos com todas as disciplinas sendo consideradas	17
4.4.2	Experimentos ignorando as disciplinas não-ofertadas	18
4.5	VARIANDO O TAMANHO DA JANELA.	18
5	CONCLUSÃO	22
5.1	TRABALHOS FUTUROS	22
	APÊNDICE A – TABELAS COM OS RESULTADOS DOS EXPERIMENTOS	23
	APÊNDICE B – TABELA DA OCUPAÇÃO REAL	34

1 INTRODUÇÃO

Uma Coordenação de Curso de uma instituição federal de ensino superior é responsável por planejar e administrar as atividades acadêmicas de um curso específico. Uma de suas responsabilidades é garantir que os alunos do curso avancem na grade acadêmica de forma adequada. Existem várias ações que uma Coordenação pode executar para atingir este objetivo. Uma destas ações é ofertar disciplinas com uma quantidade de vagas suficiente para atender a demanda dos alunos. No entanto, recursos como professores e horários são limitados. Assim, é necessário encontrar um equilíbrio no número de vagas ofertadas, evitando turmas muito grandes ou muito pequenas.

Qualquer curso é composto por uma série de disciplinas que devem ser cursadas em uma sequência determinada pelo que se denomina "grade acadêmica do curso". Esta grade normalmente está organizada em colunas, onde cada coluna indica quais as disciplinas que cada aluno deve cursar em cada "período"(normalmente semestre ou ano).

Assim, um aluno que ingressa no curso deve se matricular nas disciplinas da primeira coluna, ou primeiro período. Em um cenário utópico, o aluno é aprovado em todas estas disciplinas e então deve se matricular nas disciplinas do segundo período e assim por diante. Neste caso, todas as disciplinas da grade curricular terão sempre o mesmo número de alunos.

Porém no mundo real, nem todos os alunos conseguem se aprovar em todas as disciplinas em que se matricularam. Quando isto acontece, a quantidade de alunos em cada disciplina da grade no período seguinte pode sofrer grandes variações. Como exemplo, considere uma disciplina que em um semestre comporta n alunos. Se houver 30% de reprovação nesta disciplina, então no semestre seguinte a disciplina poderá conter n alunos oriundos do período anterior mais os $n * 0,30$ alunos que irão cursá-la novamente.

Quando a quantidade de alunos que se matriculam em uma disciplina é muito grande, isto pode prejudicar o aprendizado. Neste caso, a coordenação do curso muitas vezes divide os alunos no que se chama "turma". Por exemplo, se houver pedido de matrículas de m alunos em uma disciplina, a coordenação pode solicitar a abertura de t turmas, cada uma contendo m/t alunos. O número de alunos ideal por turma depende de fatores que vão desde o tamanho das salas (ou laboratórios) até a disponibilidade de professores do departamento que é responsável pela disciplina.

O ideal é que a coordenação saiba a priori quantos alunos irão se matricular em cada disciplina para que faça a solicitação de turmas aos departamentos, o que permite que os departamentos envolvidos se programem com antecedência.

Porém, na prática a solicitação de abertura de turmas de um período letivo ocorre ANTES do período letivo anterior terminar, quando não se sabe o número de aprovados e reprovados do período letivo anterior.

O que se pode fazer é estimar o número de alunos por disciplina baseado no histórico de aprovação e reprovação de cada disciplina. Assim, se m alunos estão matriculados em uma disciplina d pode-se examinar o número de alunos reprovados e aprovados nos semestres anteriores.

Uma vez estimadas as taxas de aprovação e reprovação, há ainda o problema de determinar em quais disciplinas o aluno irá efetivamente se matricular. Por exemplo, um aluno tem o direito de solicitar matrícula em qualquer disciplina em que ele estiver apto. Por exemplo, se ele estiver apto a se matricular em 20 disciplinas, o problema é como descobrir quais serão as escolhidas por ele, o que raramente é possível determinar.

O curso de Bacharelado em Ciência da Computação (BCC) da Universidade Federal do Paraná (UFPR) desenvolveu um conjunto de programas que estima a quantidade de alunos que irão se matricular em cada disciplina de um dado período letivo baseado no desempenho histórico dos alunos em cada disciplina.

Estes programas utilizam como entrada arquivos em formato csv disponibilizados pelo sistema informatizado de matrículas que contém as notas de cada aluno ao longo dos últimos períodos para fazer a estimativa do tamanho das disciplinas.

O algoritmo desenvolvido faz estimativas coerentes e tem sido utilizado com sucesso, porém vários parâmetros foram estipulados empiricamente sem um exame que permite dizer quais os parâmetros que resultam nas melhores estimativas.

Este trabalho tem como objetivo fazer uma análise dos valores de parâmetros que resultam nas melhores estimativas utilizando o sistema ADEGA. Este sistema está em desenvolvimento pelo PET-BCC-UFPR e utiliza como entrada os mesmos arquivos csv utilizados pela coordenação, porém reorganizados em tabelas no SGBD Postgres.

O trabalho está organizado da seguinte maneira: O Capítulo 2 apresenta o sistema ADEGA. O Capítulo 3 descreve o algoritmo desenvolvido pela coordenação. O Capítulo 4 apresenta os experimentos realizados e o Capítulo 5 resume os resultados obtidos assim como trabalhos futuros.

2 ADEGA

Os dados acadêmicos da UFPR estão armazenados em um sistema chamado Sistema Integrado de Gestão Acadêmica (SIGA). No entanto, devido a restrições de permissão de acesso e à disponibilidade desses dados apenas em planilhas extensas e de difícil manipulação, o SIGA acaba apresentando limitações à gestão do curso.

Visando solucionar essa questão, o PET Computação está desenvolvendo o ADEGA, um sistema que (1) recebe os dados do SIGA e os armazena em um banco de dados próprio e (2) disponibiliza esses dados para consulta e estabelece um ambiente para executar análises em uma aplicação web.

O ADEGA adota o padrão de arquitetura Modelo-Visão-Controlador (MVC) e é implementado com Ruby on Rails. Essa abordagem permite lidar de forma eficiente com as entidades presentes no sistema, como Aluno, Disciplina e Pré-requisitos, facilitando a consulta e análise dos dados.

A principal finalidade do sistema ADEGA é auxiliar a coordenação e os professores a examinar os dados dos alunos e do curso como um todo. O sistema permite o desenvolvimento de análises complexas, explorando as relações entre as entidades e fornecendo informações sobre o desempenho acadêmico dos alunos.

A Figura 2.1 apresenta o modelo esquemático do ADEGA. O Conversor de Entrada lê planilhas provenientes do SIGA, extrai os dados e popula o Banco de Dados através do Sistema Web.

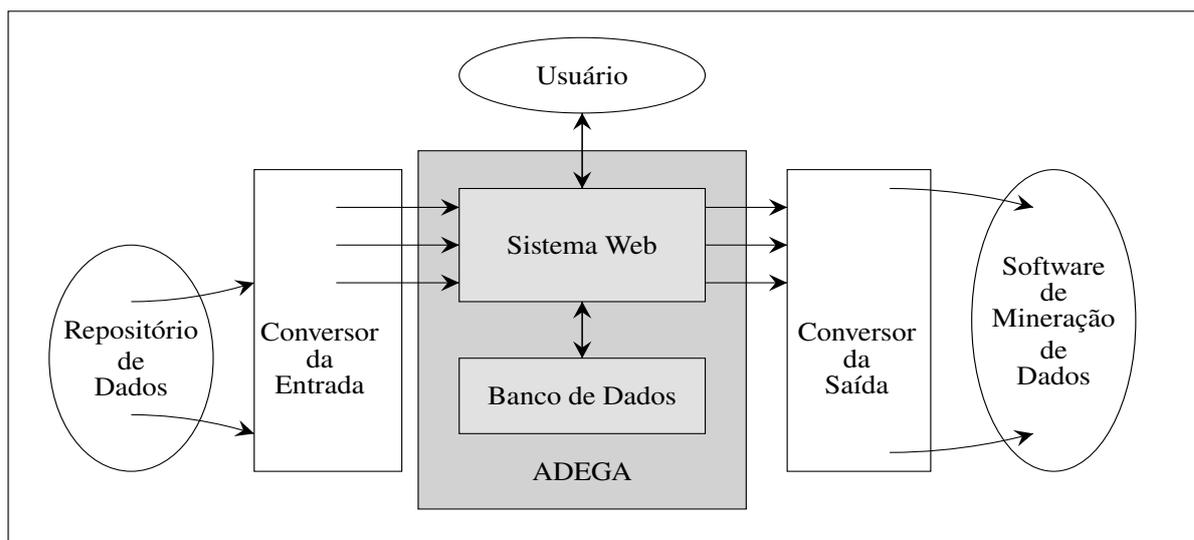


Figura 2.1: Modelo esquemático do ADEGA

As lógicas e análise contidas no ADEGA funcionam com base nos modelos internos. Dessa forma, seu núcleo é independente da fonte de dados (que hoje é o SIGA). Essa estrutura permite que o ADEGA seja flexível para receber dados de outras fontes.

3 ALGORITMO

O objetivo deste capítulo é apresentar os algoritmos considerados para a resolução do problema, será descrito em especial o algoritmo adotado para o estudo e depois comentado aspectos da implementação dele no ADEGA.

A entrada de dados consiste nos registros de todas as matrículas efetuadas pelos alunos ao longo dos anos. Cada registro contém o ano e semestre em que a matrícula foi efetuada, o resultado que o aluno teve naquela matrícula (aprovado, reprovado, cancelado, etc) e a nota obtida. Essas informações permitem fazer uma análise do que ocorreu no passado do curso e com isto estimar resultados futuros, como por exemplo o número de aprovações em uma determinada disciplina.

Ao longo do desenvolvimento deste trabalho, duas abordagens para a resolução do problema foram consideradas.

A primeira baseia-se na construção de um modelo usando o histórico dos alunos. A ideia é modelar cada disciplina do curso como nó de um grafo com arestas dirigidas. Cada aresta representa a porcentagem de alunos que historicamente se matriculam na disciplina de destino assim que são aprovados na disciplina de origem.

O grafo resultante assemelha-se a uma cadeia de Markov, mas apesar de ser um tema muito estudado, não encontramos bibliografia que trata do problema em questão e por isso não foi levado adiante.

A segunda abordagem baseia-se na análise das demandas individuais de cada aluno com base nas taxas de aprovação das disciplinas. A ideia é considerar as possíveis matrículas de cada aluno a partir das disciplinas que ele está cursando e falta cursar.

A Coordenação do curso de Bacharelado em Ciência da Computação (BCC) implementou uma versão deste segundo algoritmo e a utiliza para fazer previsões da quantidade de alunos em cada disciplina. O programa desenvolvido utiliza como entrada arquivos no formato csv obtidos diretamente do sistema da universidade.

Uma versão deste algoritmo foi implementado no ADEGA. Essa solução consiste em acumular as demandas individuais de cada aluno em um único mapa `previsao`, como mostra o pseudocódigo 1.

Algoritmo 1 CALCULADemandas

```

previsao ← {}
for A ∈ alunos_ativos do
    previsao ← previsao + demandaAluno(A)
end for

```

O algoritmo `DEMANDAALUNO` é dividido em quatro passos. O primeiro define um conjunto de disciplinas que o aluno deveria cursar. O segundo utiliza o resultado do primeiro para calcular quais disciplinas o aluno pode cursar. O terceiro calcula quais disciplinas o aluno pode reprovar. O quarto seleciona as disciplinas mais prováveis de virarem uma demanda real do aluno.

No final o resultado retornado é um mapa em que cada disciplina é associada a uma probabilidade entre 0 e 1, dada em função das taxas de aprovação e de reprovação das disciplinas.

Algoritmo 2 `DEMANDAALUNO`

```

1: demandas ← {}
2: disciplinas_pendentes ← disciplinasPendentesDoAluno(A)
3: for D ∈ disciplinas_pendentes do
4:   demandas[D] ← elegibilidade(aluno, D)
5: end for
6: for M ∈ disciplinas em que aluno esta matriculado do
7:   demandas[M] ← (1 - taxa_aprovacao(M))
8: end for
9: return top_demandas(demandas, n)

```

O algoritmo parte do princípio que as intenções de matrículas surgem de dois cenários: o aluno está pronto para cursar uma disciplina que está na sequência da grade acadêmica, ou o aluno irá reprovar na disciplina atualmente cursada, precisando repeti-la.

Desses dois cenários tem-se dois conjuntos de disciplinas em que é preciso analisar a situação do aluno: as disciplinas pendentes e as disciplinas matriculadas.

Nas disciplinas pendentes a probabilidade é definida pelo nível de elegibilidade do aluno na disciplina, enquanto nas disciplinas matriculadas a probabilidade é a taxa de reprovação.

Dentre todas as disciplinas que o aluno pode e precisa cursar, ele só vai se matricular em algumas. Então depois de explorar esses dois conjuntos restringe-se o número de disciplinas matriculadas em *N*, as com maior probabilidade. Em caso de probabilidades iguais não há um critério de desempate exato. No ADEGA a implementação usa o método `sort_by` do Ruby que possui comportamento imprevisível para valores iguais.

Esse número máximo *N* de disciplinas afetará o resultado. Por esta razão, é um parâmetro cuja variação será analisada no Capítulo 4.

No primeiro passo do algoritmo tem-se o desafio de definir quais são as disciplinas pendentes consideradas na análise. Essa questão é tratada na próxima seção e depois explica-se melhor como a elegibilidade na disciplina é calculada.

3.1 DISCIPLINAS PENDENTES DO ALUNO

O algoritmo `DISCIPLINASPENDENTESDoALUNO(A)` explora a grade acadêmica do aluno A , buscando por pendências: disciplinas que o aluno não possui aprovação nem está cursando.

Como as pendências podem ser muitas, para buscar um resultado mais próximo do real é preciso restringir quais serão consideradas. Assim, a busca não é na grade inteira, mas em um intervalo chamado aqui de janela. A janela inicia no primeiro período que não esta completo e acaba no primeiro período que não há nenhuma aprovação. A janela ainda é estendida em mais um período para que as disciplinas do próximo período entrem na análise também. Para evitar que um aluno fique sem um bom número de disciplinas para analisar, há um piso de 5 disciplinas que se não for atendido, estende-se a janela. As figuras 3.1 e 3.2 dão exemplos de como o intervalo analisado fica em relação a grade.

Como a janela é definida influencia o resultado da análise, portanto no Capítulo 4 também é analisado variações no tamanho da janela e como altera as estimativas.

Algoritmo 3 `DISCIPLINASPENDENTESDoALUNO(A)`

```

1:  $disciplinas \leftarrow []$ 
2:  $p \leftarrow periodo\_min \leftarrow$  primeiro período incompleto
3:  $periodo\_max \leftarrow$  primeiro período sem nenhuma aprovação
4: while  $p \leq periodo\_max + 1$  ou  $disciplinas.size < 5$  do
5:    $disciplinas \leftarrow disciplinas +$  disciplinas pendentes do período  $p$ 
6:    $p \leftarrow p + 1$ 
7: end while
8: return  $disciplinas$ 

```

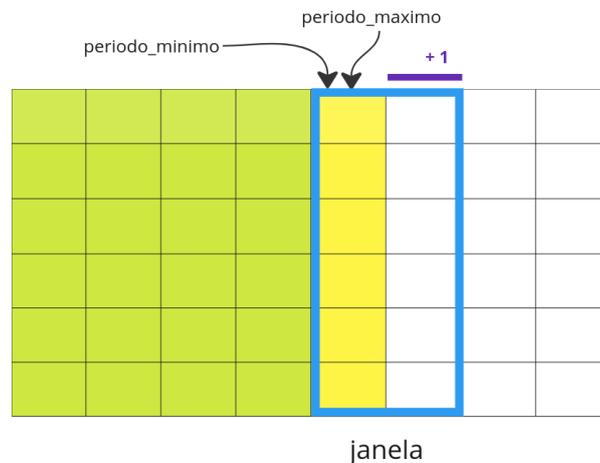


Figura 3.1: Exemplo dos limites da janela para aluno cursando o 5º período. Células verdes significam aprovado e amarelas são matrículas em curso.

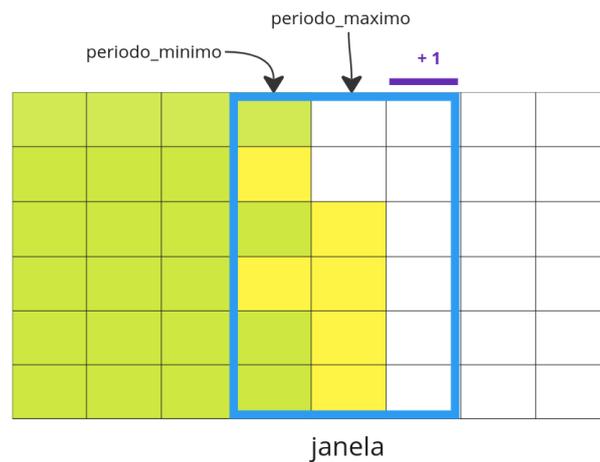


Figura 3.2: Exemplo dos limites da janela para aluno cursando o 4 e 5º período. Células verdes significam aprovado e amarelas são matrículas em curso.

3.2 ELEGIBILIDADE NA DISCIPLINA

O algoritmo $ELEGIBILIDADE(A, disciplina)$ retorna um número entre 0 e 1 que é usado como probabilidade do aluno A se matricular na $disciplina$.

Quando o aluno já foi aprovado em todos os pré-requisitos existentes admite-se que o aluno com certeza vai se matricular na disciplina, então a elegibilidade é 1.

Quando o aluno não tem todos os pré-requisitos aprovados, mas os que restam ele está cursando, admite-se que a chance de matrícula é a mesma de ele se tornar elegível no final do semestre, e esta por sua vez é a chance dele ser aprovado no pré-requisito mais difícil. Assim, a elegibilidade é a menor taxa de aprovação dentre os pré-requisitos sendo cursados.

Quando o aluno tem pré-requisitos pendentes e nem os esta cursando, então a elegibilidade é 0 pois teoricamente não há chance de ele poder se matricular. Na vida real o aluno pode requisitar uma quebra de barreira e de pré-requisitos, mas esses casos não são tratados aqui.

Algoritmo 4 $ELEGIBILIDADE(A, disciplina)$

- 1: $pr_pendentes \leftarrow$ pré-requisitos da disciplina que A ainda não cumpriu
 - 2: $probabilidade \leftarrow 0$
 - 3: **if** $pr_pendentes = \emptyset$ **then**
 - 4: $probabilidade \leftarrow 1$
 - 5: **else if** A está cursando todos os $pr_pendentes$ **then**
 - 6: $probabilidade \leftarrow menor_taxa_aprovacao(pr_pendentes)$
 - 7: **end if**
 - 8: **return** $probabilidade$
-

3.3 ASPECTOS DA IMPLEMENTAÇÃO

O algoritmo foi implementado em Ruby na forma de componentes MVC no sistema ADEGA. Os dados necessários foram o relatório de Histórico do curso retirado do SIGA, contendo todas as matrículas de todos os alunos acompanhadas de notas final e resultado na disciplina (Aprovado, Reprovado etc). Também foram usadas informações das versões da grade acadêmica: as disciplinas e modo como são ordenadas na grade, as relações de pré-requisitos e equivalências entre as versões da grade. A lista de alunos do curso também serviu como dado de entrada. A lista foi filtrada para considerar apenas os alunos ativos no curso, que constam no SIGA como "sem evasão", o que totalizou 483 alunos.

Há particularidades na implementação que são explicadas a seguir.

No cálculo das taxas de aprovação foram excluídas turmas dos anos da pandemia de COVID, 2020 e 2021. Essas turmas foram realizadas remotamente e tiveram performance irregular, podendo distorcer as taxas.

A solução trabalha apenas com disciplinas obrigatórias. Não são tratadas disciplinas optativas, de estágio e de trabalho de conclusão de curso. Essas disciplinas não tem problema com ofertas de vaga e não são uma demanda por parte da Coordenação.

O curso de BCC possui mais de um código por disciplina, como por exemplo CI055 e CI1055. Foi implementado para tratar essas equivalências como uma só, assim informações de dois códigos equivalentes são reunidos no código da grade de 2023.

Duas disciplinas não tem taxa de aprovação por falta de dados. Qualidade de Software (CI1005) não possui turmas no Histórico utilizado após a exclusão dos anos de pandemia. E a disciplina de pré-cálculo (CM310) é muito recente no currículo do curso.

4 RESULTADOS E EXPERIMENTOS

Este capítulo apresenta os experimentos feitos com a implementação do algoritmo no ADEGA. Será descrito como o programa foi parametrizado, a base de dados utilizada e as métricas que auxiliam na avaliação. Depois comenta-se os resultados dos 3 grupos de experimentos realizados.

4.1 PARÂMETROS

Para explorar como o algoritmo pode ser ajustado de forma que gere previsões mais acuradas, executou-se experimentos variando os valores de parâmetros como uma forma de calibrar a solução. Foram implementados três parâmetros:

- `MAXIMO_MATRICULAS`: número inteiro que regula quantas disciplinas assume-se que um aluno vai se matricular
- `EXTENSAO_JANELA`: número inteiro que regula o tamanho da janela da grade que será analisado
- `disciplinas_para_ignorar`: conjunto de disciplinas que são excluídas da análise

A performance de cada experimento é avaliada comparando as previsões com dados reais de matrículas. Nesses dados, algumas disciplinas não possuem turmas pois não foram ofertadas. Para que os resultados fiquem mais próximos da ocupação real, criou-se o parâmetro `disciplinas_para_ignorar`. Ele permite tirar da análise as disciplinas não-ofertadas.

4.2 BASE DE DADOS

A base de dados usada nos experimentos foi o histórico do curso de BCC coletado antes do término do semestre de 2022/2. E para a comparação de resultados, usou-se os registros de ocupação das vagas ofertadas na semana de matrículas para o semestre de 2023/1 (Tabela no Apêndice B.1).

Portanto, os experimentos a seguir simulam o desempenho que a solução teria caso fosse usada pela Coordenação durante o semestre de 2022/2 para estimar a demanda de 2023/1.

4.3 MÉTRICAS

Para medir o quão perto as estimativas estão dos valores reais calcula-se para cada conjunto de resultados o Erro Médio Absoluto (EMA) e a Raiz do Erro Médio Quadrático (REMQ).

$$EMA = \frac{1}{n} \sum_{i=1}^n |previsao_i - real_i| \quad (4.1)$$

$$REMQ = \sqrt{\frac{1}{n} \sum_{i=1}^n (previsao_i - real_i)^2} \quad (4.2)$$

Estas métricas indicam o tamanho do erro na previsão, portanto quanto menor elas forem, melhor os valores usados no experimento.

A REMQ é uma métrica mais sensível que a EMA pois eleva os erros ao quadrado. É uma métrica útil pois faz erros grandes pesarem mais. Uma previsão muito longe da realidade deve ser ressaltada a medida que não ajuda na decisão de abrir uma ou mais turmas para a disciplina. Portanto um erro na dimensão de até 10 alunos seria mais tolerável pois provavelmente não mudaria a tomada de decisão e alocação de recursos pela Coordenação.

4.4 VARIANDO A QUANTIDADE MÁXIMA DE MATRÍCULAS POR ALUNO

O algoritmo `DEMANDAALUNO` explora a situação do aluno em várias disciplinas, mas para a demanda final retorna apenas as N disciplinas com maior probabilidade. Pode se interpretar N como o número de disciplinas que em média os alunos escolhem se matricular no semestre. Segundo a grade acadêmica os alunos deveriam fazer 5 ou 6 disciplinas por semestre, mas é comum optarem por 3 ou 4 disciplinas. N é parametrizado por `MAXIMO_MATRICULAS`.

Neste experimento:

`MAXIMO_MATRICULAS` varia em 3, 4, 5 e 6.

`EXTENSAO_JANELA` é fixo em 1

Nas duas subseções a seguir são apresentadas duas variações deste experimento.

4.4.1 Experimentos com todas as disciplinas sendo consideradas

```
disciplinas_para_ignorar = []
```

Neste experimento não se ignora disciplinas não-ofertadas, calculando a demanda para todas as obrigatórias.

No gráfico da figura 4.1(a), observa-se primeiramente que, para a maioria das disciplinas, os dados dos experimentos não ficam muito distantes um dos outros. Os resultados para $N=3$ estão quase sempre logo abaixo dos resultados de $N=4$. Além disso, quanto maior o N , maior ficam as estimativas. Faz sentido pois ao aumentar o N assume-se que os alunos vão se matricular em mais disciplinas. Nos primeiros períodos os resultados ficam bem próximos e a partir do 4º e 5º período eles são mais dispersos.

No gráfico de erros da figura 4.2(a), muitas disciplinas estão com erro abaixo de 20, independente do N . Mas menos da metade tem erro menor que 10 alunos.

Por fim as métricas na tabela 4.1 permitem concluir que nesse experimento a melhor acurácia ocorreu quando MAXIMO MATRICULAS é igual a 5.

MAXIMO MATRICULAS	Erro Médio Absoluto	Raiz do Erro Quadrático Médio
3	21.991	28.27
4	19.980	26.31
5	18.520	23.67
6	20.995	31.92

Tabela 4.1: Métricas comparando os experimentos. 4 e 5 resultam no menor erro. (Experimentos variando o valor de MAXIMO MATRICULAS. Disciplinas não-ofertadas ignoradas na análise)

4.4.2 Experimentos ignorando as disciplinas não-ofertadas

Experimenta-se agora retirar as disciplinas que não serão ofertadas no período, a maioria sendo do 4º e 6º período.

```
disciplinas_para_ignorar = ["CI1059", "CI1061", "CI1064",
"CI1163", "CI1165", "CI1218", "CI1350", "CM303", "CM312"]
```

O efeito disso pode ser conferido nas figuras 4.1(b) e 4.2(b). No 5º período a ocupação real antes não era tão bem alcançada pelas previsões do que agora. Como disciplinas do 4º foram ignoradas, alunos foram deslocados para o 5º. Neste experimento tem-se mais disciplinas com o erro absoluto abaixo de 10. E pelas métricas na tabela 4.2 o melhor valor para N ainda é 5.

MAXIMO MATRICULAS	Erro Médio Absoluto	Raiz do Erro Quadrático Médio
3	14.967	23.93
4	13.021	21.93
5	11.513	18.09
6	12.820	19.47

Tabela 4.2: Métricas comparando os experimentos. (Experimentos variando o valor de MAXIMO MATRICULAS. Disciplinas não-ofertadas ignoradas na análise)

4.5 VARIANDO O TAMANHO DA JANELA

A janela é o intervalo de exploração na grade curricular do aluno. Ao aumentar a extensão da janela, trabalha-se com a hipótese que os alunos consideram se matricular disciplinas mais distantes de seu período atual.

Nesse experimento:

MAXIMO_MATRICULAS varia em 0, 1 e 2

EXTENSAO_JANELA é fixo em 5.

```
disciplinas_para_ignorar = ["CI1059", "CI1061", "CI1064",
"CI1163", "CI1165", "CI1218", "CI1350", "CM303", "CM312"]
```

Como no experimento anterior ficou evidenciado que resultados melhoram ignorando as disciplinas não-ofertadas, aqui não se experimenta prever todas as disciplinas.

No gráfico da figura 4.1(c) observa-se que o início do curso não é afetado pela variação pois uma vez que os pré-requisitos são mais rígidos, analisar disciplinas a frente do curso não muda o resultado. E a partir do 4º período é o $N=0$ que tem previsões mais acuradas.

Os resultados melhores de $N=0$ são especialmente importantes quando os outros experimentos tiveram resultados muito mais distantes. Por exemplo para CI1209, $N=2$ estimou o dobro da demanda real. Um erro nessa proporção seria a diferença entre abrir uma ou duas turmas para uma disciplina.

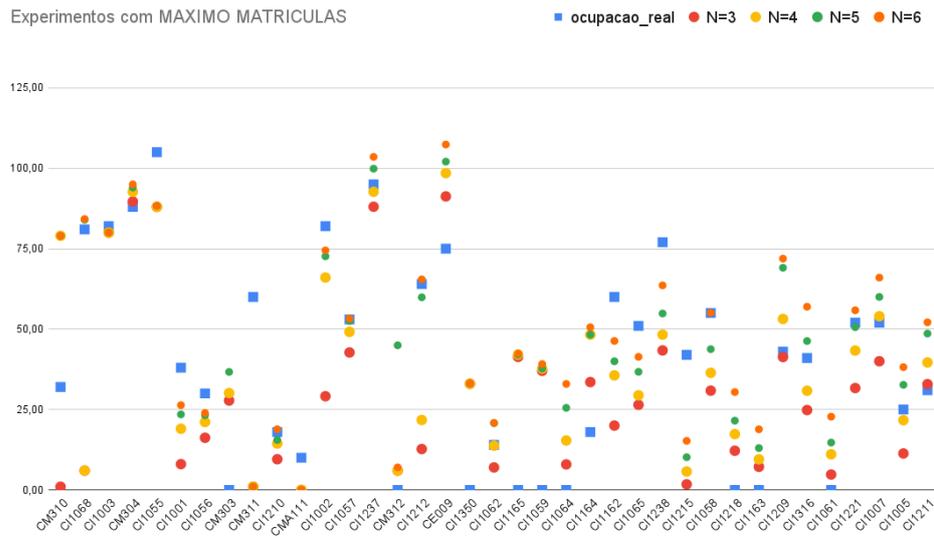
No gráfico erros na figura 4.2(c) nota-se que para todos os N s a curva de erro é bem similar, com exceção de 7 disciplinas.

Neste experimento obtém-se os menores EMA e REQM de todos para quando a janela tem extensão de 0 como mostra a tabela 4.3

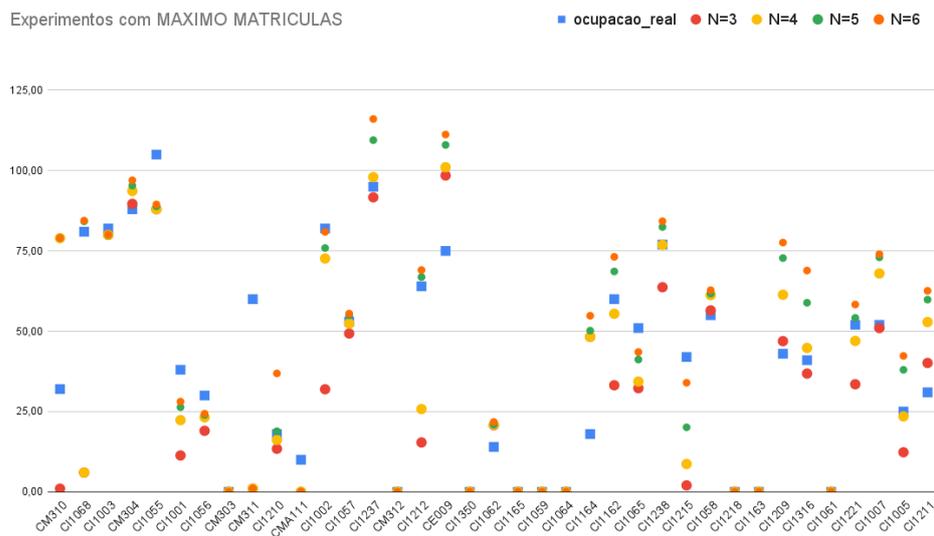
EXTENSAO JANELA	Erro Médio Absoluto	Raiz do Erro Quadrático Médio
0	10.079	16.51
1	11.513	18.09
2	13.075	20.72

Tabela 4.3: Métricas para quando experimenta-se diferentes tamanhos de janela. (Experimentos variando o valor de EXTENSAO JANELA. Disciplinas não-ofertadas ignoradas na análise)

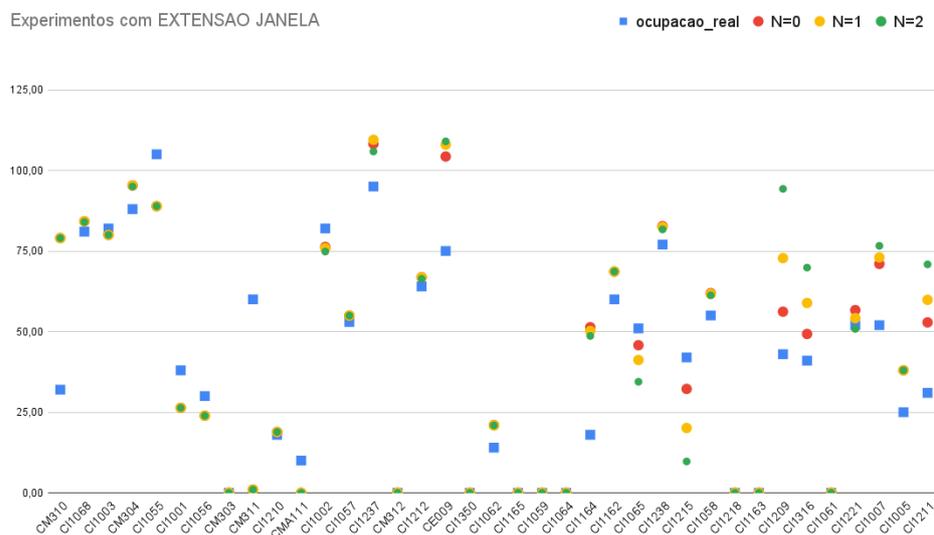
Os resultados completos dos experimentos podem ser conferidos no Apêndice A.



(a) Dispersão variando o valor de MAXIMO MATRICULAS. Todas as disciplinas consideradas.

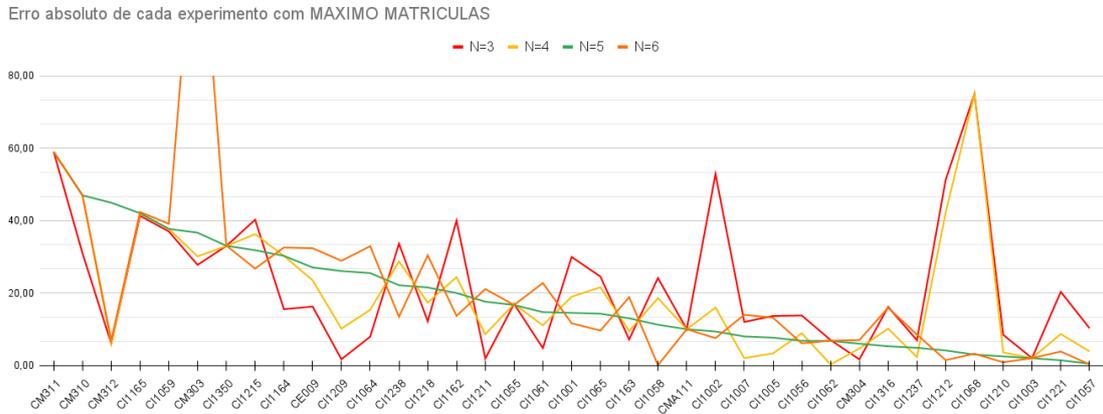


(b) Dispersão variando o valor de MAXIMO MATRICULAS. Disciplinas não-ofertadas ignoradas.

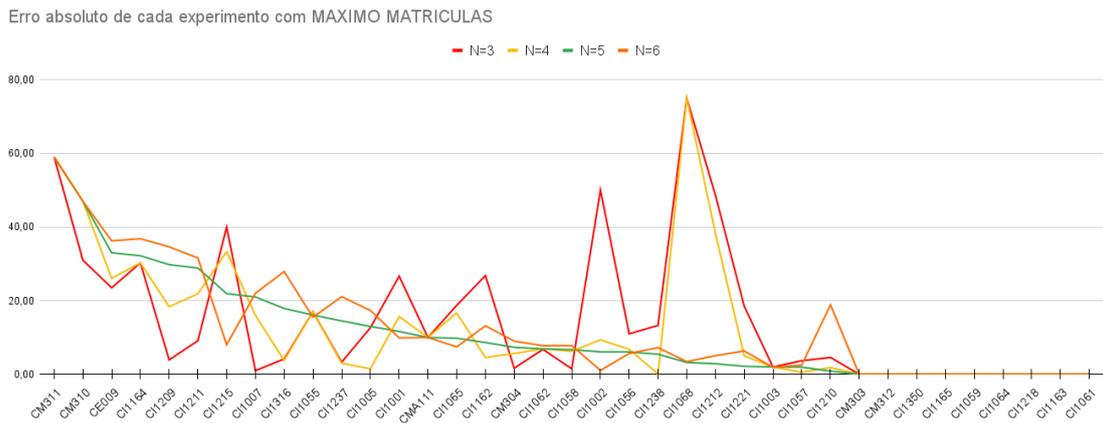


(c) Dispersão variando o valor de EXTENSAO JANELA. Disciplinas não-ofertadas ignoradas.

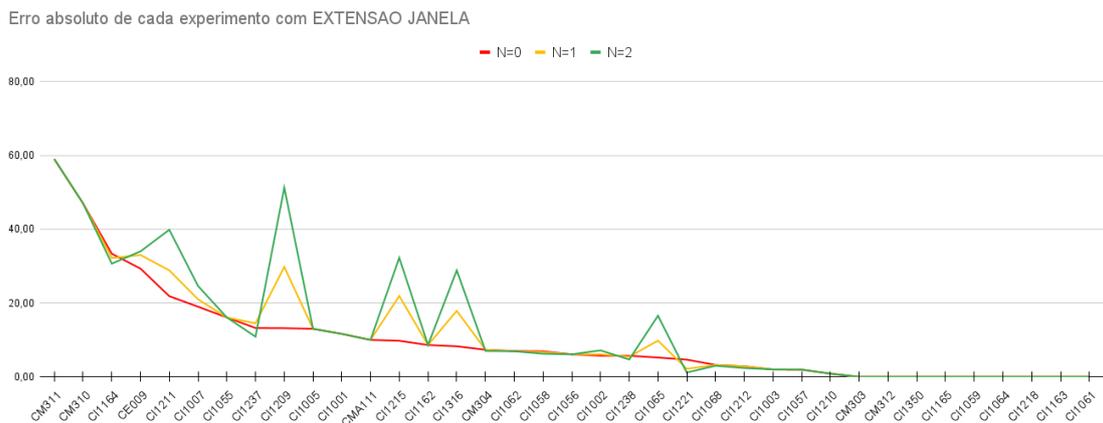
Figura 4.1: Gráficos de Dispersão. Dados da ocupação real estão como quadrados azuis para referência. As disciplinas estão dispostas no eixo X na mesma ordem da grade acadêmica.



(a) Erro Absoluto variando o valor de MAXIMO MATRICULAS. Todas as disciplinas consideradas.



(b) Erro Absoluto variando o valor de MAXIMO MATRICULAS. Disciplinas não-ofertadas ignoradas.



(c) Erro Absoluto variando o valor de EXTENSAO JANELA. Disciplinas não-ofertadas ignoradas.

Figura 4.2: Gráficos de Erro Absoluto. Disciplinas ordenadas de acordo com um dos Ns do experimento.

5 CONCLUSÃO

Neste trabalho, foram explorados valores para os parâmetros que resultam nas melhores estimativas para a demanda de turmas em um período letivo do curso de BCC. O algoritmo desenvolvido pela coordenação foi implementado no sistema ADEGA e executou-se experimentos nele que permitiram analisar como a variação dos parâmetros influenciam os resultados.

É comum a todos os experimentos que, visivelmente, os dados ficam mais esparsos após o 4º período, indicando que as disciplinas da segunda metade do curso são mais difíceis de prever a demanda. Isso pode ser devido à maior liberdade que a grade oferece em relação a pré-requisitos e assim prever comportamento e escolhas dos alunos é mais delicado. Além disso, o que influencia este problema é o algoritmo em questão não levar em consideração optativas, estágio e TCC e a demanda que estes consomem dos alunos.

Observou-se também o efeito de ignorar disciplinas na análise. As diferenças nas previsões mostrou que há um redirecionamento de alunos para outras disciplinas indicando que a Coordenação de curso ao não ofertar uma disciplina, deve prever aumento da demanda nas outras.

Foi notado ainda que nenhum experimento escapou de apresentar erros significativos em algumas disciplinas, o que sugere que a previsão precisa da demanda de turmas ainda é um desafio.

Por fim, pelos resultados conclui-se que, na previsão de matrículas do período letivo de 2023/1, o algoritmo teve maior acurácia quando considerado que alunos se matriculam em 5 disciplinas e a janela de exploração não é estendida por parâmetro.

5.1 TRABALHOS FUTUROS

Para trabalhos futuros sugere-se explorar métodos de prever a demanda que levem em consideração aspectos não incluídos por este trabalho.

Um aspecto é o desempenho individual do aluno. Este trabalho não leva em conta o índice de rendimento acadêmico do aluno por exemplo, tratando todos como iguais. Outro fator que influencia as matrículas são os professores alocados nas turmas. Poderia ser explorado como a previsão pode ser afetada pela preferência/reserva dos alunos por determinados professores.

Como extensões deste trabalho pode-se buscar personalizar o limite de matrículas por aluno em vez de usar um valor único para todos e também incorporar a solução no SIGA.

Essas melhorias contribuirão para uma coordenação mais eficiente e efetiva, promovendo uma melhor experiência acadêmica para os alunos do curso de BCC da UFPR.

APÊNDICE A – TABELAS COM OS RESULTADOS DOS EXPERIMENTOS

As tabelas a seguir foram extraídas do ADEGA no formato CSV e possuem os dados plotados nos gráficos de dispersão do capítulo 4.

Cada disciplina é exibida pelo seu código na grade acadêmica. A coluna DEMANDA PREVISTA mostra a quantidade estimada de alunos que vão se matricular na disciplina. As demais colunas são detalhes de como o algoritmo chegou na demanda prevista:

- PRS COMPLETOS: alunos que possuem todos os pré-requisitos completos, portanto estão elegíveis para cursar a disciplina
- PRS SEMICOMPLETOS: alunos que estão com os pré-requisitos semicompletos, falta aprovação em alguns e caso obtenham aprovação neles, vão ser elegíveis para a disciplina
- MATRICULADOS: alunos cursando a disciplina

Os alunos nessas colunas são os que estão sendo usados para calcular a demanda. Não consta os alunos que, por exemplo, estão cursando a disciplina mas esta não entrou para as mais prováveis de serem uma demanda para eles.

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	89.67	89	0	2	66.6
CE009	91.25	90	0	4	68.75
CI1210	9.54	3	8	1	67.84
CI1237	88.04	79	3	18	60.9
CI1057	42.73	1	53	0	90.58
CM303	27.78	27	0	2	61.02
CI1059	37.02	15	30	10	64.71
CI1350	33.0	6	43	0	94.44
CI1165	41.35	16	39	3	65.43
CI1316	24.83	23	0	5	63.41
CI1065	26.45	23	3	4	62.05
CI1007	40.0	40	0	0	95.56
CI1003	80.0	80	0	0	89.84
CI1055	88.0	88	0	0	70.04
CI1164	33.53	18	21	5	61.44
CI1058	30.86	30	1	1	78.16
CI1238	43.37	42	1	3	74.47
CI1068	6.0	6	0	0	77.71
CI1211	32.89	26	0	28	75.41
CI1209	41.31	38	0	9	63.2
CI1163	7.18	5	2	2	83.33
CI1064	7.94	4	4	5	74.07
CI1218	12.18	12	0	1	82.27
CI1221	31.67	20	14	0	74.71
CI1056	16.21	5	16	0	78.74
CI1162	20.0	20	0	0	85.19
CI1002	29.12	17	12	9	64.11
CI1005	11.33	8	4	0	
CI1001	8.02	4	5	2	74.07
CI1212	12.71	10	4	0	75.0
CI1062	6.99	5	3	1	92.34
CI1061	4.78	4	1	0	89.81
CI220	3.0	3	0	0	92.71
CI1215	1.74	1	1	0	72.27
CM310	1.0	1	0	0	
CM311	1.0	1	0	0	65.88
CM312	6.0	6	0	0	63.22

Tabela A.1: Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 3 e EXTENSAO JANELA = 1 (todas as disciplinas consideradas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	92.67	92	0	2	66.6
CE009	98.5	96	0	8	68.75
CI1210	14.42	4	13	1	67.84
CI1056	21.11	5	23	0	78.74
CI1237	92.7	83	4	18	60.9
CI1057	49.13	1	61	1	90.58
CI1002	66.03	18	60	10	64.11
CM303	30.12	27	0	8	61.02
CI1059	37.73	15	30	12	64.71
CI1350	33.0	6	43	0	94.44
CI1165	42.05	16	39	5	65.43
CI1316	30.83	29	0	5	63.41
CI1007	54.0	54	0	0	95.56
CI1065	29.4	23	7	5	62.05
CI1221	43.33	20	28	0	74.71
CI1003	80.0	80	0	0	89.84
CI1055	88.0	88	0	0	70.04
CM310	79.0	79	0	0	
CI1164	48.27	18	45	5	61.44
CI1058	36.41	32	5	5	78.16
CI1238	48.27	46	2	4	74.47
CI1162	35.61	35	1	0	85.19
CI1068	6.0	6	0	0	77.71
CI1211	39.62	32	0	31	75.41
CI1062	13.75	6	12	2	92.34
CI1209	53.15	48	0	14	63.2
CI1163	9.51	7	2	4	83.33
CI1064	15.37	7	10	7	74.07
CI1218	17.35	17	0	2	82.27
CI1001	19.02	6	16	7	74.07
CI1212	21.75	10	14	9	75.0
CI1215	5.72	1	6	1	72.27
CI1005	21.67	10	14	0	
CI1061	11.07	9	2	5	89.81
CI220	5.0	5	0	0	92.71
CM311	1.0	1	0	0	65.88
CM312	6.0	6	0	0	63.22

Tabela A.2: Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 4 e EXTENSAO JANELA = 1 (todas as disciplinas consideradas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	94.0	93	0	3	66.6
CE009	102.06	98	0	13	68.75
CI1210	15.52	4	14	2	67.84
CI1056	23.21	5	26	0	78.74
CI1001	23.48	6	22	8	74.07
CI1237	99.87	85	10	21	60.9
CI1057	52.52	1	64	12	90.58
CI1002	72.63	18	68	12	64.11
CI1212	59.88	10	68	15	75.0
CM303	36.68	32	0	12	61.02
CI1059	37.73	15	30	12	64.71
CI1350	33.06	6	43	1	94.44
CI1165	42.05	16	39	5	65.43
CI1316	46.29	43	0	9	63.41
CI1007	60.0	60	0	0	95.56
CI1211	48.61	40	0	35	75.41
CI1065	36.71	25	14	7	62.05
CI1221	50.59	22	34	1	74.71
CI1061	14.72	10	5	8	89.81
CI1003	80.0	80	0	0	89.84
CI1055	88.3	88	0	1	70.04
CM310	79.0	79	0	0	
CI1068	84.0	84	0	0	77.71
CI1164	48.27	18	45	5	61.44
CI1058	43.76	32	16	7	78.16
CI1238	54.84	49	7	5	74.47
CI1162	40.0	38	3	0	85.19
CI1209	69.07	60	1	23	63.2
CI1062	20.77	6	23	5	92.34
CI1064	25.51	7	26	8	74.07
CI1005	32.67	11	26	0	
CI1163	13.01	10	2	7	83.33
CI1218	21.53	21	0	3	82.27
CI1215	10.17	1	12	1	72.27
CI220	5.0	5	0	0	92.71
CM311	1.0	1	0	0	65.88
CM312	44.96	36	3	19	63.22

Tabela A.3: Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 5 e EXTENSAO JANELA = 1 (todas as disciplinas consideradas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	95.0	93	0	6	66.6
CE009	107.38	103	0	14	68.75
CI1210	18.87	4	15	10	67.84
CI1056	23.91	5	27	0	78.74
CI1001	26.36	6	25	11	74.07
CI1237	103.53	86	14	21	60.9
CI1057	53.28	1	64	20	90.58
CI1002	74.47	18	70	13	64.11
CI1212	65.45	10	74	21	75.0
CM303	136.56	112	0	63	61.02
CI1059	39.14	15	30	16	64.71
CI1350	33.12	6	43	2	94.44
CI1165	42.39	16	39	6	65.43
CI1316	56.95	50	0	19	63.41
CI1007	66.0	66	0	0	95.56
CI1211	52.1	43	0	37	75.41
CI1221	55.84	22	40	2	74.71
CI1065	41.37	25	20	9	62.05
CI1061	22.77	10	15	12	89.81
CI1209	71.91	61	1	28	63.2
CI1003	80.0	80	0	0	89.84
CI1055	88.3	88	0	1	70.04
CM310	79.0	79	0	0	
CI1068	84.22	84	0	1	77.71
CI1164	50.58	18	45	11	61.44
CI1058	55.06	34	31	7	78.16
CI1238	63.58	49	21	5	74.47
CI1162	46.31	41	8	0	85.19
CI1218	30.42	29	0	8	82.27
CI1062	20.84	6	23	6	92.34
CI1064	32.95	8	36	9	74.07
CI1005	38.17	14	29	0	
CI1163	18.85	15	2	12	83.33
CI1215	15.25	1	19	1	72.27
CI220	5.0	5	0	0	92.71
CI166	1.0	1	0	0	82.8
CM312	7.0	7	0	0	63.22
CM311	1.0	1	0	0	65.88

Tabela A.4: Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 6 e EXTENSAO JANELA = 1 (todas as disciplinas consideradas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	89.67	89	0	2	66.6
CE009	98.5	96	0	8	68.75
CI1210	13.42	3	13	1	67.84
CI1237	91.7	82	4	18	60.9
CI1057	49.31	1	61	3	90.58
CI1316	36.83	35	0	5	63.41
CI1007	51.0	51	0	0	95.56
CI1211	40.13	33	0	29	75.41
CI1065	32.27	25	9	4	62.05
CI1003	80.0	80	0	0	89.84
CI1055	88.0	88	0	0	70.04
CI1164	48.27	18	45	5	61.44
CI1058	56.5	32	37	6	78.16
CI1238	63.72	49	22	4	74.47
CI1068	6.0	6	0	0	77.71
CI1062	20.77	6	23	5	92.34
CI1162	33.19	28	8	0	85.19
CI1209	46.92	43	1	9	63.2
CI1221	33.5	21	15	0	74.71
CI1056	19.01	5	20	0	78.74
CI1002	31.94	18	13	12	64.11
CI1005	12.33	9	4	0	
CI1001	11.34	4	9	4	74.07
CI1212	15.39	10	5	8	75.0
CI1215	2.02	1	1	1	72.27
CI220	5.0	5	0	0	92.71
CI166	1.0	1	0	0	82.8
CM310	1.0	1	0	0	
CM311	1.0	1	0	0	65.88

Tabela A.5: Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 3 e EXTENSAO JANELA = 1 (desconsidera disciplinas não-ofertadas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	93.67	92	0	5	66.6
CE009	101.06	97	0	13	68.75
CI1210	16.17	4	14	4	67.84
CI1056	23.21	5	26	0	78.74
CI1237	97.99	85	6	23	60.9
CI1057	52.43	1	64	11	90.58
CI1002	72.65	18	67	14	64.11
CI1316	44.8	42	1	6	63.41
CI1007	68.0	68	0	0	95.56
CI1211	52.87	45	0	32	75.41
CI1221	47.0	22	30	0	74.71
CI1065	34.34	25	11	6	62.05
CI1003	80.0	80	0	0	89.84
CI1055	88.0	88	0	0	70.04
CM310	79.0	79	0	0	
CI1164	48.27	18	45	5	61.44
CI1058	61.26	34	41	7	78.16
CI1238	76.89	49	43	4	74.47
CI1162	55.45	41	23	0	85.19
CI1068	6.0	6	0	0	77.71
CI1062	20.92	6	23	7	92.34
CI1209	61.37	55	2	14	63.2
CI1005	23.5	11	15	0	
CI166	3.0	3	0	0	82.8
CI1215	8.68	1	10	1	72.27
CI1001	22.34	6	20	9	74.07
CI1212	25.78	10	17	17	75.0
CI220	5.0	5	0	0	92.71
CM311	1.0	1	0	0	65.88

Tabela A.6: Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 4 e EXTENSAO JANELA = 1 (desconsidera disciplinas não-ofertadas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	95.34	93	0	7	66.6
CE009	108.0	103	0	16	68.75
CI1210	18.87	4	15	10	67.84
CI1056	23.91	5	27	0	78.74
CI1001	26.36	6	25	11	74.07
CI1237	109.5	87	15	32	60.9
CI1057	54.95	1	66	21	90.58
CI1002	75.91	18	70	17	64.11
CI1212	66.88	10	75	24	75.0
CI1316	58.88	54	2	10	63.41
CI1007	73.0	73	0	0	95.56
CI1211	59.85	51	0	36	75.41
CI1221	54.17	22	38	2	74.71
CI1005	38.0	13	30	0	
CI1065	41.2	25	18	12	62.05
CI1003	80.0	80	0	0	89.84
CI1055	88.9	88	0	3	70.04
CM310	79.0	79	0	0	
CI1068	84.22	84	0	1	77.71
CI1164	50.2	18	45	10	61.44
CI1058	61.7	34	41	9	78.16
CI1238	82.46	51	47	8	74.47
CI1162	68.62	41	44	0	85.19
CI1209	72.79	62	2	26	63.2
CI1062	20.92	6	23	7	92.34
CI166	3.0	3	0	0	82.8
CI1215	20.1	2	25	1	72.27
CI220	5.0	5	0	0	92.71
CM311	1.0	1	0	0	65.88

Tabela A.7: Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 5 e EXTENSAO JANELA = 1 (desconsidera disciplinas não-ofertadas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	97.01	93	0	12	66.6
CE009	111.25	105	0	20	68.75
CI1210	36.88	4	15	66	67.84
CI1056	24.34	5	27	2	78.74
CI1001	28.1	6	26	15	74.07
CI1237	116.08	87	19	42	60.9
CI1057	55.51	1	66	27	90.58
CI1002	80.93	18	70	31	64.11
CI1212	69.06	10	76	30	75.0
CI1055	89.5	88	0	5	70.04
CI1316	68.9	60	2	21	63.41
CI1007	74.0	74	0	0	95.56
CI1211	62.59	53	0	39	75.41
CI1221	58.34	22	43	2	74.71
CI1005	42.33	14	34	0	
CI1215	33.97	2	43	4	72.27
CI1065	43.54	25	21	13	62.05
CI1003	80.1	80	0	1	89.84
CM310	79.0	79	0	0	
CI1068	84.45	84	0	2	77.71
CI1164	54.83	18	45	22	61.44
CI1058	62.79	34	41	14	78.16
CI1238	84.25	51	47	15	74.47
CI1162	73.17	43	48	0	85.19
CI1209	77.61	63	4	33	63.2
CI1062	21.76	6	23	18	92.34
CI166	3.0	3	0	0	82.8
CI220	5.0	5	0	0	92.71
CM311	1.0	1	0	0	65.88

Tabela A.8: Tabela com os resultados do algoritmo para MAXIMO MATRICULAS = 6 e EXTENSAO JANELA = 1 (desconsidera disciplinas não-ofertadas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	95.34	93	0	7	66.6
CE009	104.31	99	0	17	68.75
CI1210	18.87	4	15	10	67.84
CI1056	23.91	5	27	0	78.74
CI1001	26.36	6	25	11	74.07
CI1237	108.23	86	14	33	60.9
CI1057	54.95	1	66	21	90.58
CI1002	76.27	18	70	18	64.11
CI1212	66.88	10	75	24	75.0
CI1316	49.27	45	1	10	63.41
CI1007	71.0	71	0	0	95.56
CI1211	52.85	44	0	36	75.41
CI1221	56.67	22	41	2	74.71
CI1005	38.0	13	30	0	
CI1065	45.78	25	25	12	62.05
CI1003	80.0	80	0	0	89.84
CI1055	88.9	88	0	3	70.04
CM310	79.0	79	0	0	
CI1068	84.22	84	0	1	77.71
CI1164	51.36	18	45	13	61.44
CI1058	61.92	34	41	10	78.16
CI1238	82.72	51	47	9	74.47
CI1162	68.62	41	44	0	85.19
CI1215	32.23	2	41	2	72.27
CI1062	21.0	6	23	8	92.34
CI1209	56.18	46	1	26	63.2
CI166	3.0	3	0	0	82.8
CI220	5.0	5	0	0	92.71
CM311	1.0	1	0	0	65.88

Tabela A.9: Tabela com os resultados do algoritmo para EXTENSAO JANELA = 0 e MAXIMO MATRICULAS = 5 (desconsidera disciplinas não-ofertadas)

disciplina	demanda prevista	prs completos	prs semicompletos	matriculados	taxa aprovacao
CM304	95.0	93	0	6	66.6
CE009	109.0	104	0	16	68.75
CI1210	18.87	4	15	10	67.84
CI1056	23.91	5	27	0	78.74
CI1001	26.36	6	25	11	74.07
CI1237	105.87	87	16	21	60.9
CI1057	54.95	1	66	21	90.58
CI1002	74.83	18	70	14	64.11
CI1212	66.38	10	75	22	75.0
CI1316	69.85	61	8	10	63.41
CI1007	76.61	76	1	0	95.56
CI1211	70.85	62	0	36	75.41
CI1221	50.84	22	34	2	74.71
CI1005	38.0	13	30	0	
CI1065	34.45	25	10	8	62.05
CI1003	80.0	80	0	0	89.84
CI1055	88.9	88	0	3	70.04
CM310	79.0	79	0	0	
CI1068	84.0	84	0	0	77.71
CI1164	48.66	18	45	6	61.44
CI1058	61.26	34	41	7	78.16
CI1238	81.7	51	47	5	74.47
CI1162	68.62	41	44	0	85.19
CI1209	94.27	74	17	26	63.2
CI1062	20.92	6	23	7	92.34
CI166	3.0	3	0	0	82.8
CI1215	9.73	2	11	1	72.27
CI220	5.0	5	0	0	92.71
CM311	1.0	1	0	0	65.88

Tabela A.10: Tabela com os resultados do algoritmo para EXTENSAO JANELA = 2 e MAXIMO MATRICULAS = 5 (desconsidera disciplinas não-ofertadas)

APÊNDICE B – TABELA DA OCUPAÇÃO REAL

Essa tabela foi gerada a partir da planilha de ocupação de turmas do SIGA. A planilha original foi filtrada para manter apenas turmas do período letivo de 2023/1, excluir turmas de disciplinas optativas e depois as turmas foram agrupadas por disciplina. Assim, a tabela contém os registros de ocupação das disciplinas obrigatórias ofertadas nesse período. Os números da coluna matriculados são usados para calcular os erros absolutos que são plotados nos gráficos do capítulo 4.

codigo	matriculados	vagas total
CE009	75	75
CI1001	38	60
CI1002	82	90
CI1003	82	88
CI1005	25	40
CI1007	52	55
CI1055	105	120
CI1056	30	40
CI1057	53	60
CI1058	55	55
CI1062	14	40
CI1065	51	51
CI1068	81	87
CI1162	60	80
CI1164	18	40
CI1209	43	45
CI1210	18	40
CI1211	31	40
CI1212	64	80
CI1215	42	44
CI1221	52	60
CI1237	95	100
CI1238	77	90
CI1316	41	41
CM304	88	95
CM310	32	31
CM311	60	100
CMA111	10	10

Tabela B.1: Tabela com as matrículas reais de 2023/1