UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS, TECNOLOGIA E SAÚDE

Joice Preuss Cardoso

APLICANDO CIÊNCIA DE DADOS PARA ANÁLISE DO PERFIL DOS ALUNOS EM CURSOS DE TECNOLOGIA DA UFSC: 2008 - 2018

Araranguá

2019

Joice Preuss Cardoso

APLICANDO CIÊNCIA DE DADOS PARA ANÁLISE DO PERFIL DOS ALUNOS EM CURSOS DE TECNOLOGIA DA UFSC: 2008 - 2018

Trabalho de Conclusão de Curso submetido ao Bacharelado em Engenharia de Computação para a obtenção do Grau de Bacharel em Engenharia de Computação.
Orientadora: Profa. Dra. Analúcia Schiaffino Morales

Araranguá

2019

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Cardoso, Joice Preuss
Aplicando ciência de dados para análise do perfil dos alunos em cursos de tecnologia da UFSC: 2008 - 2018 / Joice Preuss Cardoso ; orientadora, Analúcia Schiaffino Morales, 2019.
62 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Campus Araranguá, Graduação em Engenharia de Computação, Araranguá, 2019.

Inclui referências.

1. Engenharia de Computação. 2. Mulheres na tecnologia. 3. Ciência de dados. 4. Mineração de dados. 5. Análises estatísticas . I. Morales, Analúcia Schiaffino . II. Universidade Federal de Santa Catarina. Graduação em Engenharia de Computação. III. Título.

Joicé Preuss Cardoso

APLICANDO CIÊNCIA DE DADOS PARA ANÁLISE DO PERFIL DOS ALUNOS EM CURSOS DE TECNOLOGIA DA UFSC: 2008 - 2018

Este Trabalho de Conclusão de Curso foi julgado aprovado para a obtenção do título de "Bacharel em Engenharia de Computação" e aprovado em sua forma final pelo Bacharelado em Engenharia de Computação.

Araranguá, 05 de julho 2019.

Prof. Dr. Fabrício de Oliveira Ourique
Coordenador

Banca Examinadora:

Profa. Dra. Analúcia Schiaffino Morales
Orientadora

Profa. Ma. Tatiana Nilson Dos Santos
Universidade Federal de Santa Catarina

Ellen Larissa de Carvalho Aquino
Universidade Federal de São Carlos

Dedico este trabalho a meus pais e meu irmão.

AGRADECIMENTOS

Primeiramente gostaria de agradecer a Deus e a nossa senhora, que por meio das orações das minhas mãe e avó, conferiram alguns milagres durante toda a minha vida acadêmica. A minha mãe que sempre foi um exemplo de mãe e profissional, e indiretamente, fez eu me apaixonar pela computação. A meu pai que sempre me incentivou a leitura e ao estudo, despertando em mim o hábito pela curiosidade. Esses dois criaram um lar desconstruído e encorajador, possibilitando de todas as formas possíveis que eu chegasse até aqui. A meu irmão Josef, por ser o meu amigo mais antigo, meu copiloto, a pessoa mais esperta, engraçada e amigável que eu conheço. Obrigada por tudo.

A professora Dra. Analúcia Schiaffino Morales que me orientou neste trabalho e compartilha não só o conhecimento, mas também o dia e o mês de nascimento comigo.

A minha profunda gratidão a professora Dra. Luciana Bolan Frigo, que me orientou em toda a minha vida acadêmica e me ensinou o significado de ser uma pesquisadora, através do projeto Meninas Digitais UFSC. Agradeço a amizade, os ensinamentos e os conselhos, que vão de "a referência vai no final ou no início" até "que vinho combina mais com peixe". A toda equipe do LabTeC, onde contruí muitas experiências, conhecimentos e amizades, vocês são demais. Ao professor Dr. Éverton Fabian Jasinski pela amizade, por conseguir fazer de mim uma esportista e por estar sempre disponível para ajudar todo mundo a todo momento. E a toda família/time, softball "Araras UFSC", estar com vocês é uma das melhores experiências da vida. A professora Dra. Daiana Cristine Bundchen pela amizade e por trazer música a minha vida acadêmica. E a todo o coral "Encanta UFSC", continuem sempre cantando e encantando. A Universidade Federal de Santa Catarina por sua excelência na educação e por me acolher e transformar em uma pessoa melhor.

A minha vó Evonia por ser meu porto seguro e a meu avô Bertoldo por me ensinar o valor do trabalho braçal e sempre me dizer "calma, não pensa tanto, descansa a cabeça". Ao Marcelo, outro culpado por eu estar na computação, e a Camila por me darem o maior presente que já ganhei, ser madrinha do Theo. A minha madrinha professora Angela por ser um exemplo de dedicação, minha fã número 1 e minha parceira de trabalho nas férias. E a todos os meus primos e tios, que não são poucos, e aos meus avós e familiares que já se foram.

Aos meus amigos Antony Turra, Bruno Silva, Caroline Fontana, Ellen Larissa, Laís Dalle Mulle, Maria Teresa, Pamela Brunh, e a todos os que não citei, porque não ia caber todos aqui, mas no meu coração cabe. Obrigada por serem os grandes amores da minha vida, por tornarem meus anos na faculdade mais tranquilos e pela amizade que vai durar para sempre.

Ao meu colega, roommate, tradutor, cozinheiro, pai da Irene e amigo, Gabriel Ganzer, você foi literalmente minha bengala durante toda a vida acadêmica, desde o momento em que passamos no vestibular, devo todas as correções de português e trabalhos, traduções e chás a essa pessoa. Obrigada por enxugar muitas das minhas lágrimas e permitir que eu enxugasse as tuas, que eram poucas. Você é meu parceiro de crime e conseguiu o título de pessoa que mais me conhece.

Amo a todos.

RESUMO

O baixo ingresso de mulheres nos cursos superiores das áreas de tecnologia vem sendo observado desde a década de 90. Um dos problemas ocasionados pela falta de mulheres nos setores tecnológicos é proporcionar uma visão masculina no setor produtivo, o que pode aumentar ainda mais esta disparidade, uma vez que os produtos resultantes tendem a ser mais atraentes para os homens. Essa desigualdade pode alterar a evolução da própria tecnologia. A partir disso, esse trabalho busca compreender como se contextualiza o alto índice de evasão nos cursos de tecnologia e engenharias da Universidade Federal de Santa Catarina. Para tratar e analisar os dados de alunos, disponibilizados na base de dados dos cursos, este trabalho usa a ciência de dados e a mineração de dados. Desta forma, o objetivo deste trabalho é confirmar ou refutar algumas hipóteses sobre o ingresso, permanência e evasão de mulheres nas áreas de tecnologia. Os principais resultados desta pesquisa são a confirmação do baixo ingresso de mulheres nas áreas tecnológicas e o desempenho superior dessas alunas.

Palavras-chave: Mineração. Dados. Gênero.

ABSTRACT

A low women entry in undergraduate courses related to technological fields has been observed since the 90's decade. One of the problems caused by the lack of women inside technological departments is the provision of a mainly masculine point of view from the productive sector since the resultant goods tend to be more attractive to male consumers. The gender disparity may alter the technological evolution itself. Therefore, this thesis is concerned in understanding how the high rate of evasion is contextualized inside courses related to technology and engineering from the Universidade Federal de Santa Catarina. This thesis uses data science and data mining as a tool to address and analyze student data made available in the course database. Hence, this thesis main purpose is to confirm or refute some hypotheses about the entry, permanence and evasion of women in the areas of technology. The main results of this research regards the confirmation of the low enrollment of women into technological areas and a superior performance of these female students.

Keywords: Data. Mining. Gender.

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE ABREVIATURAS E SIGLAS

SUMÁRIO

1 INTRODUÇÃO

Este capítulo apresenta a contextualização, a justificativa, os objetivos, a metodologia utilizada e a organização deste trabalho.

1.1 CONTEXTUALIZAÇÃO

No século XIX, as mulheres propuseram uma nova questão a partir do cenário de desigualdade de gênero em que se encontravam. Essa batalha ultrapassa séculos, porém algumas vitórias já foram conquistadas: o direito ao voto, o emprego, a aposentadoria, etc. Quando são analisadas as questões como salário e posição dentro de uma organização, encontram-se ainda certas desigualdades. Em 2015, por exemplo, funcionárias mulheres de empresas de grande porte recebiam 21,5% menos do que os funcionários homens, além disso, somente 20,1% dos diretores e 30,1% dos gerentes de produção e operações eram do sexo feminino (PRONI; PRONI, 2018). A diferença salarial ao longo da carreira, assim como as oportunidades de crescimento profissional, têm sido apontadas como fatores que afastam as mulheres das carreiras ligadas à tecnologia.

Um fenômeno estudado dentro das pesquisas de gênero é o desinteresse das mulheres pelas áreas de computação. Dados relevantes, como o desempenho escolar feminino ou a situação de mulheres em universidades, por exemplo, podem apontar sinais de uma inclinação forçada para outras áreas, sendo assim, importantes ferramentas de aprendizado. Estes estudos têm promovido iniciativas de órgãos internacionais, tais como, a Organização para Cooperação e Desenvolvimento Econômico (OCDE) e a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), buscando a igualdade de gênero em todas as áreas (OLINTO, 2011).

No cenário dos cursos ligados a tecnologia, de acordo com o Censo 2012, em 2001, a porcentagem feminina era de aproximadamente 24%, e reduziu em 9% até o ano de 2012. Altas taxas de evasão têm sido verificadas para ambos os sexos nesses cursos (OLIVEIRA; MORO; PRATES, 2014). Já, segundo dados do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), de 2000 a 2013, apenas 17% dos concluintes de cursos de graduação em computação eram do sexo feminino (MAIA, 2016).

Isso se fortalece devido o que Olinto (2011) descreve como dois tipos de segregação encontradas por mulheres: a horizontal e a vertical. A primeira, se constitui da influência que a família e escola tem sobre as escolhas das meninas em toda sua vida. Nela as mulheres buscam as carreiras consideradas femininas, de acordo com o que lhes foi passado durante toda sua vida. Já a segregação vertical, atua na submissão, ou seja, ela mantém as mulheres em posições inferiores e barra as tentativas de crescimento profissional, através de um minucioso e imperceptível sistema. Esta é ainda mais perigosa, pois além de descredibilizar a profissional, ela ainda procura por meios de achar que essa sensação de incompetência é apenas um reflexo da incapacidade das mulheres em ciência e tecnologia.

Olinto (2011), ainda, destaca que a equidade de gênero está longe de ocorrer, devido ao fato dos jovens e crianças já estarem afetados pela segregação horizontal. Não existem fatores científicos que apontem para uma incapacidade feminina ou ainda para uma supremacia intelectual masculina em ciências e tecnologia, sendo então inválida a afirmação da falta de habilidade pelas mulheres nestas áreas de conhecimento (OLINTO, 2011).

Analisando os dados apresentados pelo PNAD (Pesquisa Nacional por Amostra de

Domicílios), no relatório PNAD (2009), verifica-se que no setor de tecnologia, em média, as mulheres possuem o grau de instrução "Ensino Superior Completo"e os homens "Ensino Superior Incompleto ou Equivalente", sendo a menor formação masculina “sem instrução” ou com “Ensino Fundamental Incompleto” e feminina "Ensino Fundamental Completo ou Equivalente"(CASTRO, 2013). Análises sem o devido cuidado podem apontar para uma igualdade salarial, porque quando se trata da média salarial geral, homens e mulheres aparentam receber a mesma remuneração (CASTRO, 2013). As mulheres nos cargos de analista de sistema e programadora recebem em média R$ 2.972,54 e R$ 1.485,30, com teto de R$ 10.000,00 e R$ 8.000,00, respectivamente (PNAD, 2009). No entanto, executando as mesmas funções, os homens tem um rendimento médio de R$ 3.333,29 e de R$ 1.825,62, respectivamente, chegando até R$ 19.000,00 e R$ 8.500,00 (PNAD, 2009). Com esses dados é possível observar uma das desigualdades encontradas dentro do mercado de trabalho. Homens e mulheres com uma mesma formação, tem discrepâncias significativas em seus rendimentos, e a situação só se agrava quando o teto salarial é analisado (CASTRO, 2013).

1.2 JUSTIFICATIVA

As equipes profissionais da engenharia e computação são, em geral, formadas por maioria masculina. Isso não reflete a diversidade da população mundial, tornando essas tecnologias produzidas com um viés de conhecimento, valores restritos a um grupo. Autores como Maia (2016) indicam que fatores como estereótipos e preconceitos acompanham as profissionais por toda a carreira. O afastamento de mulheres nas áreas de tecnologia podem significar um desperdício de mão de obra em áreas com escassez de profissionais qualificados (SOARES, 2001). Para que elas sejam capacitadas, há a necessidade da inserção das mulheres em ambientes dominados pelo sexo masculino, porém o baixo ingresso feminino nas áreas de tecnologia vem sendo verificado ao longo dos anos (CASTRO, 2013). Assim, analisar padrões e investigar a atuação das mulheres nos setores tecnológicos ajudam a apresentar resultados significativos, para que iniciativas e medidas possam ser tomadas, em busca de soluções para a falta de diversidade de gênero no setor.

1.3 HIPÓTESES

Todas as hipóteses estão baseadas nos cursos de ciências da computação, sistemas de informação, tecnologias da informação e comunicação e as engenharias elétrica, de computação e de controle e automação, da UFSC (Universidade Federal de Santa Catarina). Sendo assim, para este trabalho foram levantadas as seguintes hipóteses:

1.3.1 Hipótese 1

O ingresso de mulheres nos cursos vem diminuindo ao longo dos anos.

1.3.2 Hipótese 2

Existem padrões na evasão e retenção feminina nos cursos de graduação analisados.

1.3.3 Hipótese 3

As mulheres tem melhor rendimento do que os homens nesses cursos.

1.3.4 Hipótese 4

A evasão acontece no primeiro ano do curso, para ambos os sexos.

1.4 OBJETIVOS

1.4.1 Objetivo geral

O presente trabalho tem como objetivo principal analisar os dados fornecidos pela UFSC, sobre os alunos de alguns cursos de graduação de áreas tecnológicas.

1.4.2 Objetivos específicos

- Levantar o histórico da situação das mulheres nas áreas de tecnologia e engenharia;
- Realizar pesquisa sobre as principais aplicações da ciência e da mineração de dados;
- Identificar os pontos fundamentais da mineração de dados;
- Verificar o ingresso e retenção de mulheres em cursos superiores de engenharia e tecnologia;
- Encontrar padrões nos dados dos alunos da UFSC através da ciência de dados;
- Avaliar os padrões levantados;
- Minimizar os impactos da falta de mulheres na tecnologia, através da proposição de novas ações.

1.5 METODOLOGIA

O trabalho realizado caracteriza-se como uma pesquisa exploratória, em que foram realizadas análises de dados históricos de alunos de graduação, buscando informações relevantes sobre o perfil do público feminino na UFSC. Para atingir os objetivos apresentados na seção 1.4 foram realizadas as seguintes etapas:

1. Foram pesquisadas as áreas de ciência e mineração de dados;
2. Foram pesquisadas referências sobre a situação das mulheres nas carreiras de tecnologia e engenharia;
3. Foi realizado um pré-processamento dos dados para a análise;
4. Foram estudadas e definidas quais as técnicas que seriam aplicadas;

5. Foram implementadas e executadas as análises;
6. Os padrões obtidos foram analisados;
7. Os resultados da pesquisa foram avaliados;
8. E por fim, os resultados foram registrados.

1.6 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em cinco capítulos. Após a Introdução, segue o segundo capítulo com a fundamentação teórica, apresentando um levantamento da situação das mulheres nas áreas tecnológicas, e é feita uma explanação sobre as técnicas empregadas no processo de análise dos dados. O terceiro capítulo contém o desenvolvimento do trabalho, em que são apresentados os dados a serem analisados, as ferramentas adotadas e como foi realizada a análise dos dados. O quarto capítulo contém a discussão dos resultados obtidos e no quinto e último capítulo esta a conclusão do trabalho. Por fim, as devidas referências bibliográficas.

2 FUNDAMENTAÇÃO TEÓRICA

A seguir são apresentados os conhecimentos necessários para a compreensão deste trabalho. São abordadas as questões das mulheres na tecnologia, ciência e a mineração de dados. Os trabalhos relacionados serão discutidos no final do capítulo.

2.1 MULHERES NA TECNOLOGIA

Nesta seção está o levantamento histórico e a situação da participação das mulheres na ciência, tecnologia e engenharia.

2.1.1 História da mulher na ciência e tecnologia

Nos séculos XVII e XVIII, as mulheres desempenhavam papéis na ciência, pois seu lugar na área ainda não estava definido (SCHIEBINGER, 2001). Como a pesquisa era feita em laboratórios em um ambiente familiar, sem a necessidade de uma graduação e com uma organização relapsa, era possível que as mulheres desenvolvessem suas próprias análises científicas (AQUINO, 2015). Quando a ciência inevitavelmente abriu-se para a sociedade em geral, foi reservado o espaço do cuidado doméstico as mulheres, aniquilando as chances de pesquisadoras contribuírem com a ciência (SCHWARTZ et al., 2006). De acordo com Aquino (2015), esse período foi um divisor de águas. Isso por que antes, seu lugar no desenvolvimento científico só era possível através de maridos, pais ou irmãos, que atuassem na área, reservando a mulher a execução de tarefas auxiliares (SCHWARTZ et al., 2006). Este cenário apresenta alguma mudança a partir do século XX, ainda assim, insuficientemente.

Com a evolução da tecnologia, uma ciência aplicada, e com o surgimento dos primeiros computadores, cabiam as mulheres a atividade de programação das máquinas, surgindo assim, uma aparente área feminina. A maior parte dessas mulheres tinha formação nas áreas de matemática e ciência (LUBAR, 1998). Elas ingressaram neste campo científico inexplorado e despretensioso, contudo, a partir do momento que ele se tornou cientificamente e economicamente lucrativo, emergiu concomitantemente o interesse masculino pela área, atingindo finalmente a predominância do gênero no setor (CASTRO, 2013). Nesse sentido, a falta de modelos femininos conhecidos nas áreas de computação e tecnologia é destacada por Schiebinger (2001). As principais contribuições femininas permaneceram invisíveis, tendo as poucas mulheres de destaque estado associadas ao âmbito dos softwares, enquanto a história da computação é, em grande parte, contada pela perspectiva do desenvolvimento de hardware (RAPKIEWICZ, 1998). Light (1999) afirma que o desaparecimento de mulheres de destaque na história da computação contribui para a visão de que elas não mostram interesse ou não são capazes de estar nessa área.

Como exemplo, temos a história de seis jovens programadoras que trabalharam no ENIAC (*Electronic Numerical Integrator and Computer*, o primeiro computador digital eletrônico, em português "Computador Integrador Numérico Eletrônico"), porém seus nomes sumiram da história do mesmo. Nessa direção Castro (2013), reforça que na verdade há uma ilusão na ideia da exclusividade masculina das grandes descobertas científicas. Light (1999) em 1999 tentou desmistificar o papel das mulheres na computação, apresen-

tando ao mundo a importância das mulheres na história da computação. Outro ponto é o estigma de que as áreas de exatas e computação são difíceis (OLIVEIRA; MORO; PRATES, 2014).

2.1.2 Análise do contexto da mulher na área da tecnologia

A partir desse cenário apresentado, estudos voltados à compreender a redução da entrada de mulheres nos setores de ciência e tecnologia, vêm sendo realizados desde a década de 80 (CASTRO, 2013). Em 1988, Azevêdo et al. (1989) identificou, com base em dados do CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico) e FINEP (Financiadora de Estudos e Projetos), que as mulheres eram 30% dos pesquisadores do Brasil. Dos consultores científicos do FINEP na área de ciências exatas 8% eram mulheres, estando o setor de engenharia composto apenas por homens. Desse modo, Soares (2001) alertava em 2001 sobre a falta de pesquisas na área de gênero no Brasil, analisando o cenário mundial alarmante e verificando a quantidade de mulheres em ciência e tecnologia, ela menciona que o país não seria uma exceção, refletindo portando, os problemas verificados em diferentes partes do mundo.

De 1990 a 2005, os cursos de ciência e engenharia de computação têm apresentado uma diminuição do número de mulheres, passando de 30% para 5% a 10% (FILHO, 2005). Bilton (2014) revela essa tendência nas grandes empresas do setor. Em 2014, 83% dos engenheiros na Google eram homens e dos 36 executivos e gerentes apenas três são mulheres. Além disso, 85% e 80% dos profissionais em tecnologia do Facebook e da Apple, respectivamente, são homens. Natansohn (2013) destaca que mulheres engenheiras de computação, empreendedoras na área, desenvolvedoras e administradoras de sistemas ainda são minoria.

O mesmo cenário, também é reforçado analisando dados do CENSO 2012 e de um questionário online. Oliveira, Moro e Prates (2014) fazem uma análise inicial do perfil feminino na computação. Em 2001, a porcentagem feminina era de aproximadamente 24%, caindo para 15% em 2012, mesmo com um aumento de cursos e estudantes, 601 para 2231 e 127 mil para 300 mil, respectivamente. Ainda é possível verificar, tanto para homens quanto para mulheres, uma alta taxa de evasão, mais de 85% não concluem o curso. Os questionários realizados focaram em mulheres estudantes e profissionais de áreas da computação.

Através de dados do INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), Maia (2016) verificou que, de 2000 a 2013, apenas 17% dos concluentes de cursos do campo da computação eram do sexo feminino, tendo o curso de robótica formado apenas por homens.

2.1.3 Influências exercidas sobre a participação da mulher

Em 2001, duas pesquisadoras suecas analisaram três motivações para a baixa participação feminina na ciência, sendo elas: que as mulheres são menos motivadas e orientadas, gerando uma diminuição na candidatura a cargos e bolsas, ou elas são menos produtivas, descredibilizando o mérito científico de seu trabalho, ou ainda, a discriminação de gênero (WENNERAS; WOLD, 2001). Ademais, Soares (2001) mostra que hipóteses como falta de controle emocional ou agressividade competitiva por parte das mulheres, são

apenas reflexos de estruturas sociais ou familiares inaptas, que as conduzem a padrões que não apresentam a verdadeira capacidade feminina. De acordo com Leslie, McClure e Oaxaca (1998), como garotas são desestimuladas as áreas consideradas masculinas muito jovens, isso acarreta posteriormente na dificuldade em acompanhar os estudos nessas áreas durante a universidade. As meninas tem a tendência de subestimar suas habilidades, enquanto os meninos as superestimam (SOARES, 2001). A frustração na carreira leva as profissionais a serem menos destemidas e confiantes, desacreditando o sucesso na carreira e levando-as a crer na inaptidão feminina. Uma pesquisa financiada pelo Conselho de Pesquisa Médica da Suécia, mostrou que os homens tem que produzir 2,2 vezes menos que mulheres para conseguir um financiamento, apresentando assim uma discriminação velada, no final do século XX (WENNERAS; WOLD, 2001).

Alguns pesquisadores persistem na ideia que as áreas da tecnologia são neutras em questões de gênero e desestimulam qualquer iniciativa em promover a participação feminina, voltando os problemas nas próprias mulheres, na sua cultura, educação e valores, eliminando qualquer influência advinda das imposições da área (NATANSOHN, 2013). O contexto de influência sobre a decisão de escolha na área da ciência e da tecnologia, apresenta segundo Soares (2001) quatro fatores dificultantes comuns a vida das profissionais de todo o mundo, em especial as das áreas de tecnologia e engenharia, sendo eles: a conciliação entre a sua profissão e a do parceiro, a jornada dupla como dona de casa e profissional, a falta de colegas mulheres em cargos de chefia, que impossibilita uma identificação dos problemas, e ainda, o não reconhecimento por parte da comunidade científica, apenas três mulheres foram laureadas com o nobel da física, cinco com o de química e uma com o de economia, por exemplo. Boicotes de posições como docência, orientação, coordenação e chefia, podem desmotivar as profissionais e ainda diminuir a possibilidade de trabalho com pesquisa dentro das universidades (OLINTO, 2011). De acordo com Olinto (2011), os obstáculos criados dentro do ambiente acadêmico e científico, são os maiores culpados responsáveis pela falta de progresso profissional, ou mesmo permanência, das mulheres na ciência e tecnologia.

Oliveira, Moro e Prates (2014) constatou que a grande maioria, tanto estudantes como profissionais da área de computação, sempre preferiram as áreas de matemática e física, e essa foi a grande influência para o ingresso na área. Quanto ao preconceito sofrido por ser mulher, houve um indicativo que a vida no campo profissional é mais complicada que a do período como discente. Ainda foi possível verificar que, por mínimo que fosse, a grande maioria das mulheres já sofreu algum tipo de preconceito durante a carreira.

Uma condição constatada por Castro (2013) é a difícil relação da vida profissional com a vida pessoal e a maternidade. A tentativa de conciliação entre essas várias jornadas leva a casos de mulheres que traçam longos planos com base em fatores como o prazo de validade da hora de ser mãe e as licenças causadas pela maternidade. Enquanto isso, homens não precisam traçar tais planos, já que na maioria dos casos, não tem necessidade de se afastar do trabalho na hora de se tornarem pais. Algumas entrevistas constantes no trabalho de Castro (2013) apontam rotinas de trabalho excessivas e duradouras, que culminaram na impossibilidade de engravidar, enquanto outras apresentam a dura realidade de ser mãe em meio a carreira. A renda e as formas como os papeis dos gêneros foram construídas, levando em conta o papel de pais e a divisão do cuidado com os filhos, podem ser fatores determinantes nas escolhas de mulheres quanto a sua carreira. Talvez por isso as mulheres são levemente mais propensas a buscar por contratos de emprego padrão, que tem uma maior chance de garantir direitos trabalhistas e sociais, enquanto os homens estão mais relacionados a contratos flexíveis, sem garantias (CASTRO, 2013).

Essas histórias podem não ser exclusividade desse setor, porém somadas as barreiras da área tecnológica, são capazes de talvez influenciar a evolução da carreira de uma mulher.

Por outro lado, Castro (2013) constatou em 2013, um padrão nas motivações que levaram homens e mulheres a optarem pelas áreas de tecnologia. Essas entrevistas foram realizadas com profissionais de São Paulo e apontaram que enquanto os homens afirmam que sempre estiveram voltados a esta escolha por suas próprias aptidões, em contrapartida as mulheres evidenciam que a influência dos pais ou mesmo pelo fato de geralmente optar por brincadeiras e áreas ditas como masculinas na infância e adolescência foi o principal fator na escolha da carreira. Isso reforça a propensão feminina em duvidar ou descredibilizar suas próprias preferências, não confiando em sua capacidade para a área.

Um fenômeno que habitualmente ocorre com mulheres, principalmente nas áreas computacionais, é mudança de setor. As mulheres tendem a buscar áreas dentro do padrão socialmente considerado feminino, que exigem geralmente a habilidade de comunicação, como designers, gerentes de projeto, entre outros (GLOVER; GUERRIER, 2010). Geralmente, os setores vistos como os mais difíceis e que exigem capacidades elevadas são os ligados à programação. Sendo essa uma área solitária, complicada e considerada a essência da computação, todos os que se afastam dela são vistos como desertores da função central (CASTRO, 2013). Cargos ligados a gestão tem habilidades relacionadas a "traços femininos"e são muitas vezes, portanto, desvalorizados (GLOVER; GUERRIER, 2010). Porém, homens que apresentem tais competências são admirados e valorizados, em contrapartida, as mulheres são apenas vistas agindo conforme suas capacidades naturais (KELAN, 2009). Para conseguir permanecer dentro das áreas de desenvolvimento e evitar assédio, muitas mulheres neutralizam seu lado feminino e assumem uma postura considerada mais séria, se vestindo ou se comportando de maneira mais masculina, com o objetivo de serem identificadas como iguais pelos seus pares homens (CASTRO, 2013).

Segundo Aquino (2015), a inserção de mulheres na tecnologia abre oportunidades para que elas sejam capacitadas, fazendo assim, afinal, parte desta área. O aumento da participação feminina nas áreas de ciência e tecnologia pode trazer impactos sociais e econômicos positivos, além de trazer outras abordagens e perspectivas para as pesquisas científicas (SOARES, 2001). Ainda de acordo com Soares (2001), a segregação feminina a determinadas carreiras simboliza um desperdício de mão de obra qualificada. Vasilescu, Serebrenik e Filkov (2015) afirma que grupos mais diversos, seja gênero, raça entre outros, têm propensão a serem mais efetivos com variedades de ideias, habilidades em resolução de problema, etc.

Os possíveis fatores que causam a desigualdade de gênero nesses setores são inúmeros, e devem ser cautelosamente analisados (CASTRO, 2013). Olinto (2011) assegura a necessidade de pesquisas e iniciativas que promovam um debate sobre a falta de equidade de gênero nas áreas de ciência, que são fundamentais para a redução da segregação vertical. Já a segregação horizontal exige medidas para que essa visão seja difundida entre famílias e escolas (OLINTO, 2011). Brinquedos, por exemplo, podem ser fonte de muitas habilidades no desenvolvimento social e mecânico das crianças, mas também podem esconder o início da diferenciação profissional dos gêneros (SANTOS et al., 2016).

2.1.4 Iniciativas brasileiras de fortalecimento do campo

Devido a baixa presença de mulheres em cursos de tecnologia e engenharia, projetos de extensão e pesquisa vêm sendo realizados no Brasil com objetivo de promover a igual-

dade de gênero. O programa Meninas Digitais, criado em 2011, é institucionalizado pela Sociedade Brasileira de Computação e busca incentivar meninas a ingressarem na área da computação. Diferentes ações são desenvolvidas de acordo com o projeto e o objetivo de cada localidade. Em geral, são ofertados minicursos, oficinas, atividades com dinâmicas de grupo, palestras com estudantes e profissionais que já atuam na área compartilhando suas experiências, realização de eventos, etc (RIBEIRO, 2019). Em 2019, aproximadamente 71 projetos foram cadastrados e estão ativos no Brasil.

O Projeto Meninas Digitais - UFSC, está em desenvolvimento desde 2013, no Campus Araranguá e já atuou em diferentes escolas em cerca de 5 municípios no sul de Santa Catarina, e já realizou atividades em outras capitais, tais como Florianópolis, Porto Alegre e São Paulo. O projeto realiza diversas atividades para as alunas de ensino fundamental e médio das escolas, com minicursos de computação desplugada, desenvolvimento de aplicativos móveis, jogos digitais, robótica e sistemas de automação, montagem de circuitos elétricos e eletrônicos, implementações e construções da domótica e de cidades inteligentes, entre outras atividades. Para o público universitário são realizadas reuniões internas semanais para discussão das atividades nas escolas e reflexões sobre temas da atualidade e situações ocorridas tanto no âmbito universitário como fora dele, com o objetivo de fortalecimento do grupo e acolhimento. Além disso, também são realizadas palestras com mulheres inspiradoras para falar sobre carreira na tecnologia e suas experiências pessoais e profissionais, estas atividades tem como foco atuar na retenção das alunas que chegam na academia (FRIGO et al., 2013).

Outro projeto parceiro do programa Meninas Digitais é o Metabotix. Ele é um projeto do Instituto Federal de Goiás (IFG), Campus Luziânia, que busca motivar meninas do ensino médio através da desmistificação do papel da mulher na área de TI, utilizando hardware e software livres, além de materiais recicláveis, mostrando o mundo da programação e robótica de uma maneira mais lúdica (SANTOS et al., 2016).

Cesário et al. (2017) cita ainda outros projetos brasileiros que publicaram no evento "X Women in Information Technology"(WIT). São eles "Trazendo Meninas para a Computação"(UCS), "Gurias na Computação"(Unipampa), "Mulheres e Jovens na Computação"(IFRS – Bento Gonçalves), "Meninas, Computação e Música"(UTFPR - Cornélio Procópio), "Emíli@s - Armação em Bits"(UTFPR – Curitiba), #include<meninas.uff> (UFF), Meninas Digitais - Regional Mato Grosso (IFMT e UFMT), Meninas.comp (UnB), SciTech Girls' Project (UFAM) e Cunhantã Digital (UFAM).

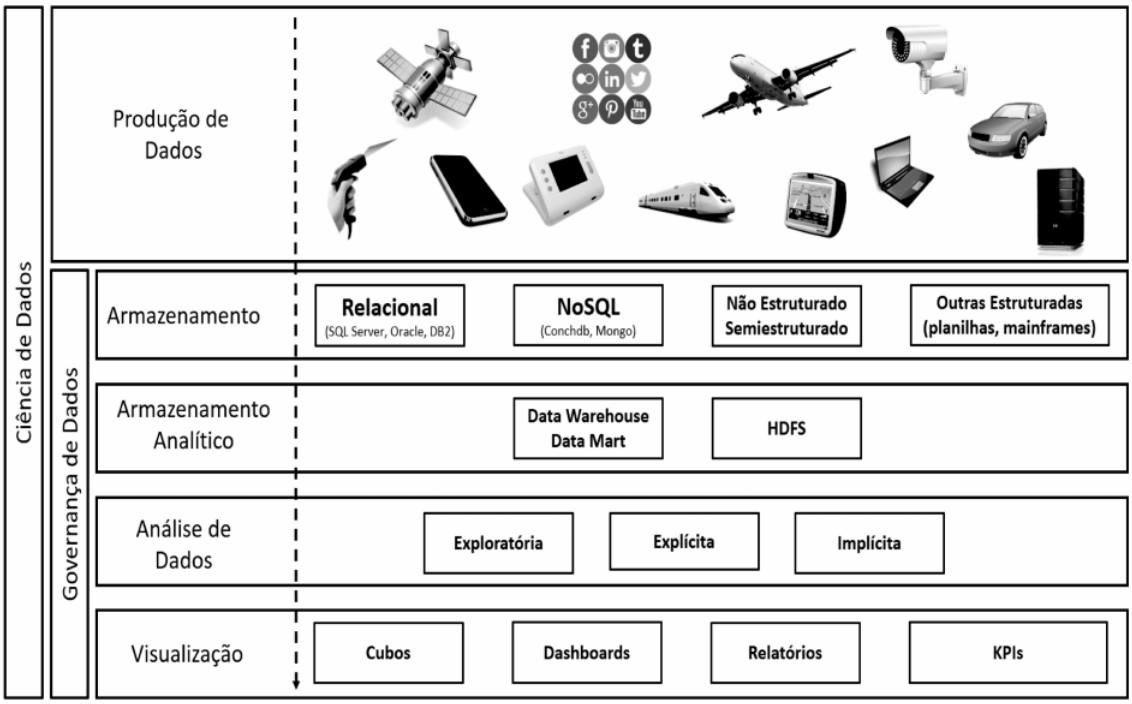
2.2 CIÊNCIA DE DADOS

Um dado é o resultado de alguma coleta ou extração, como a temperatura do motor de um carro ou um resposta em um questionário. Quando atribui-se significado para esses dados, obtém-se as informações, e a partir do momento em que estão disponíveis para um determinado fim, será gerado o conhecimento. A ciência de dados tem o dado, a informação e o conhecimento como suas principais matérias primas. Muitas vezes confundida com uma simples análise estatística, a ciência de dados compreende desde a coleta até o descarte dos dados.

A figura 1 apresenta um panorama de todas as fases da ciência de dados. Primeiro, tem-se a produção dos dados, por meio de sensoriamento, pesquisas ou coletas, esses dados geralmente são mantidos em planilhas ou bancos de dados. Posteriormente, eles são convertidos para formatos que sejam compatíveis com as ferramentas de análise, como o

HDF5, por exemplo. Depois de preparados os dados, direciona-se para a fase de extração de informações e conhecimento, onde podem ser usadas técnicas estatísticas ou de aprendizado de máquina para este fim. Após essa fase a informação é apresentada através de gráficos ou mesmo relatórios, onde o usuário possa enxergar a informação de maneira mais clara. Por fim, dependendo das regras que regem esses dados, o descarte é feito em meses ou anos. Os dados ainda podem ter questões como qualidade, segurança ou privacidade. Assim a ciência de dados acompanha todo o processo de vida do dado na busca de extrair algum conhecimento (AMARAL, 2016).

Figura 1 – Panorama da ciência de dados.



Fonte: (AMARAL, 2016)

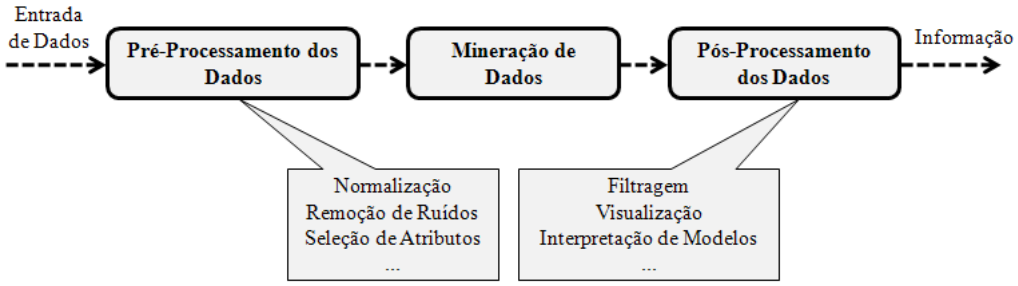
Nos dias atuais, cientistas sociais tem em mãos um interessante conjunto de informações sobre o comportamento humano, possibilitando análises em grande escala. Porém essas pesquisas envolvem um grande número de dados, além da união de diversas áreas com uma variedade de habilidades e ferramentas, como a mineração de dados, o processamento de linguagem natural, a estatística, entre outras (TAN et al., 2018).

Tendo suas primeiras citações em torno de 1980 (TAN et al., 2018), a mineração de dados surgiu como uma consequência da evolução tecnológica (HAN; PEI; KAMBER, 2012). Atualmente, existem dados sendo coletados das mais diversas formas e lugares, gerando um complexo conjunto de informações. Programas capazes de extrair resultados relevantes dessa massa de dados foram sendo necessários ao longo do crescimento da informatização. Segundo Tan et al. (2018), a mineração de dados tenta encontrar padrões que antes poderiam passar despercebidos em um grande conjunto de dados, em busca de informações interessantes e úteis para pesquisas e análises. Ela tem sido usada na melhoria de sistemas, fornecendo qualidade a busca por resultados, com base na relevância deles para suas entradas. A utilização de algoritmos de mineração de dados fornece ao analisador mais informações sobre o conjunto de dados ao qual são aplicados, assim os padrões levantados podem ofertar aos pesquisadores novas perspectivas e cenários.

Ainda, por volta de 1980, começaram a surgir uma série de oficinas sobre extração de conhecimento em bases de dados. Nelas eram feitos debates sobre as vantagens da aplicação de técnicas computacionais para extração de conhecimento em grandes bases de dados. O crescimento dessas oficinas dentro de conferências importantes e o interesse apresentado pelo mercado e a indústria, impulsionaram o crescimento da área. Fazendo uso de métodos e algoritmos já conhecidos, pesquisadores aproveitam ideias de áreas como estatística, modelagem, inteligência artificial, reconhecimento de padrões, aprendizado de máquina, aplicando essas abordagens para resolver os desafios da mineração de grandes dados (TAN et al., 2018). Atualmente, segundo os autores Tan et al. (2018) e Freitas (2002), a mineração de dados é uma das etapas dentro do processo que é a extração de conhecimento, ou também chamado de KDD (em inglês, *Knowledge Discovery in Databases*), porém vem emergindo ao longo dos anos como um campo acadêmico dentro da ciência da computação.

Segundo Han, Pei e Kamber (2012) a mineração de dados é confundida com a extração de conhecimento como um todo, por ser um termo menor (em inglês, *Data Mining*). Analisando a mineração como uma parte da extração de conhecimento, ela é realizada após o pré-processamento dos dados.

Figura 2 – Extração de conhecimento.



Fonte: adaptada de (TAN et al., 2018)

Como mostrado na figura 2, após a entrada do dado é feito o pré-processamento, considerado o estágio mais trabalhoso de todo o processo. Nele os dados são preparados para as análises que serão feitas nas próximas etapas, através da união de dados de diferentes fontes, remoção de informações dispensáveis, selecionando apenas o essencial para a mineração. O pós-processamento é a fase que garante que apenas dados válidos e relevantes sejam passados para o sistema de apoio à decisão (TAN et al., 2018). Em resumo, a fase de preparação dos dados, busca transformar os dados a fim de facilitar o trabalho da próxima etapa, enquanto a etapa de refinamento procura validar e aprimorar as informações obtidas através da mineração (FREITAS, 2002). Han, Pei e Kamber (2012) apresenta todo o processo de extração de conhecimento em sete etapas:

1. Limpeza dos dados: removendo inconsistências;
2. Integração dos dados: unindo dados semelhantes;
3. Seleção de dados: escolha de dados relevantes para a análise;
4. Transformação dos dados: conversão dos dados para uma forma mais adequada para a mineração;

5. Mineração de dados: aplicação de métodos inteligentes para a determinação de padrões;

6. Avaliação dos padrões: reconhecimento de padrões genuinamente úteis, tendo como base medidas de interesse;

7. Apresentação de conhecimento: demonstração dos resultados através de técnicas de visualização e representação de conhecimento;

As quatro primeiras etapas são formas diferentes de pré-processamento e as duas últimas são a preparação e apresentação dos padrões para a geração de novo conhecimento (HAN; PEI; KAMBER, 2012).

Dentro da mineração de dados existem tarefas que podem ser realizadas para extrair informações interessantes dos conjuntos de dados. Segundo Tan et al. (2018) essas tarefas são geralmente divididas em duas grandes categorias: preditivas e descritivas. De acordo com Tan et al. (2018) as duas tarefas tem por objetivo minimizar o erro entre a predição e os verdadeiros valores de uma variável objetivo. Ambas possuem um conjunto de técnicas que podem ser aplicadas aos conjuntos de dados.

2.2.1 Tarefas preditivas

As tarefas preditivas preveem valores de atributos, conhecidos como variável alvo (*target*) ou variável independente, com base em outros atributos, chamados de explicativos (*explanatory*) ou variáveis independentes (TAN et al., 2018). Em resumo, as tarefas preditivas tem por objetivo prever eventos futuros, com base em dados já existentes. As técnicas preditivas de classificação e regressão serão apresentados nas próximas seções.

2.2.1.1 Classificação

Os algoritmos de classificação são utilizados em variáveis independentes discretas (TAN et al., 2018). É o processo de descrever e distinguir classes ou conceitos através de modelos baseados na análise de dados, onde os rótulos das classes são conhecidos, assim encontrando o rótulo desconhecido de outras classes de objetos (HAN; PEI; KAMBER, 2012). Exemplo: Usuários comprando algum objeto online, já que a variável alvo é um valor binário, ou seja, a compra ou não do objeto (TAN et al., 2018).

2.2.1.2 Regressão

Mais frequentemente utilizada na predição numérica, a regressão engloba a identificação da distribuição de tendências baseada nos dados disponíveis (HAN; PEI; KAMBER, 2012). É empregada para variáveis independentes contínuas. Exemplo: Prever o aumento ou a diminuição no preço de um objeto a venda online, já que o preço do produto é um atributo de valor contínuo (TAN et al., 2018).

2.2.2 Tarefas descritivas

Diferente das tarefas preditivas, as descritivas não necessitam da supervisão dos dados de entrada e saída. Elas unem dados a partir de semelhanças compartilhadas. Seu objetivo é formar padrões, sendo na maioria das vezes exploratória, necessitando frequentemente de pós-processamento para validação e fundamentação dos resultados (TAN et al., 2018).

2.2.2.1 Associação

Algoritmos de associação são utilizados para perceber padrões que descrevam características fortemente associadas nos dados, extraindo-os de maneira eficiente (TAN et al., 2018). Exemplos: Identificação de páginas online que são acessadas paralelamente ou conhecer o relacionamento entre os diferentes elementos do clima terrestre (TAN et al., 2018).

2.2.2.2 Agrupamento (*Cluster*)

Ao contrário da classificação e da regressão, que analisam conjuntos de dados através do rótulo das classes, *clustering* ou agrupamento analisa objetos de dados sem consultar os rótulos. Os objetos são agrupados baseados nos princípios de maximizar a similaridade intraclasse e minimizar a similaridade interclasse, ou seja, agrupamentos de objetos são formados de modo que dentro de um grupo os objetos tenham alta similaridade entre si e apresentem diferenças visíveis de objetos de outros agrupamentos(HAN; PEI; KAMBER, 2012). Cada grupo pode ser visto como uma classe de objetos, a partir da qual as regras podem ser extraídas. O agrupamento pode facilitar a taxonomia, isto é, organizar os grupos com base em características comuns e dar nomes a esses grupos, além de ser usado para gerar rótulos de classes (HAN; PEI; KAMBER, 2012). Exemplos: agrupar conjuntos de clientes relacionados e comprimir dados (TAN et al., 2018).

2.2.2.3 Detecção de Anomalias

A detecção de anomalias identifica objetos que contenham características significativamente diferentes do resto dos dados (TAN et al., 2018). A análise de dados atípicos que será referenciada neste trabalho como detecção de anomalias, também pode ser encontrada na literatura como análise de *outlier* ou *anomaly mining* (HAN; PEI; KAMBER, 2012). Conforme Tan et al. (2018) o objetivo de um algoritmo de detecção de anomalias é descobrir reais anomalias, evitando falsos positivos, em outras palavras, uma boa detecção de anomalias deve ter uma alta taxa de detecção e uma baixa de alarmes falsos. Muitos métodos de mineração de dados tratam anomalias como ruídos ou exceções, contudo, eventos raros podem ser mais interessantes que os regulares em algumas aplicações. Objetos atípicos podem ser detectados usando testes estatísticos que assumem uma distribuição ou um modelo probabilístico, ou usando médias de distâncias, onde os objetos que estão longe de qualquer grupo são considerados anômalos (HAN; PEI; KAMBER, 2012). Como por exemplo, detecção de fraude e padrões incomuns de doenças (TAN et al., 2018).

2.3 TRABALHOS RELACIONADOS

Esta seção apresenta os trabalhos relacionados ao tema, além de projetos que auxiliaram na produção e análise do conjunto de dados.

Couto e Dantas (2014) procuram através de técnicas de mineração, encontrar padrões em conjuntos de dados de alunos dos cursos de Ciência da Computação, Licenciatura em Computação e Engenharia da Computação da Universidade de Brasília, além de alunas do Ensino Fundamental e Médio do Distrito Federal. Com esses padrões eles buscam analisar as diferenças de gênero nos cursos de graduação, além de identificar o pensamento das alunas do ensino fundamental e médio sobre a área. A ferramenta escolhida pelos pesquisadores foi WEKA (*Waikato Environment for Knowledge Analysis*), que já tem algoritmos de aprendizagem de máquina e várias ferramentas de pré-processamento de dados, possibilitando um rápido desenvolvimento.

No estudo com as meninas em fase escolar, Couto e Dantas (2014) identificaram predições sobre a intenção de carreira das estudantes. As garotas do ensino fundamental que acreditam que a profissão é bem remunerada mostraram interesse em ingressar em um curso superior em computação. No ensino médio, em geral, as estudantes que tem algum conhecimento prévio de computação, programação e banco de dados, por exemplo, somado a crença no prestigio e boa remuneração da profissão, são as que escolheriam a computação como carreira profissional.

Com os alunos do ensino superior foram extraídas informações como motivos de desligamento do curso, médias nas disciplinas, etc. Ainda segundo Couto e Dantas (2014), as mulheres tendem a abandonar o curso enquanto os homens não cumprem as condições para a formação, sendo essas as formas mais comuns de desligamento. As disciplinas com médias mais baixas são Física I, Cálculo I e Computação Básica. A análise individual aponta que a média das alunas de Ciência da Computação na disciplina de Física I é menor que a dos homens, enquanto nos outros cursos essa média é significativamente maior que a dos homens. No caso de Computação Básica, as mulheres da Licenciatura ficaram abaixo da média. Os períodos com maior desligamento do curso, no caso dos homens, são os de segundo ao quarto semestre. Para as mulheres a análise definiu o terceiro semestre sendo um período decisivo.

Ainda, a pesquisa mostrou que existem mais estudantes homens vindos de escola pública do que mulheres, provindas, em sua maioria, de escolas particulares. Outro ponto ressaltado é a maior variedade racial entre as acadêmicas em comparação aos do outro gênero. Estudos mais aprofundados dos resultados e verificação de outros padrões são apontados como trabalhos futuros (COUTO; DANTAS, 2014).

Em lugares como a Asia e a Africa, as pessoas demostram preferencias por filhos de um determinado gênero, por questões como pressão econômica ou social. Nesta linha, existem pesquisas médicas voltadas a escolha do gênero de maneira natural. Assim em Sabir et al. (2017) são aplicas diferentes técnicas de classificação de mineração de dados em fatores macroscópicos que afetam o gênero do feto no útero da mãe, identificando o gênero como masculino ou feminino. Três técnicas de classificação foram aplicadas: árvore de decisão[1], *Naive Bayes*[2] e redes neurais[3], através da ferramenta *RapidMiner*. Neste trabalho, o classificador *Naive Bayes* obteve resultados mais precisos em comparação

[1]Método de classificação baseado em árvores.

[2]Classificador probabilístico baseado no teorema de Bayes, utiliza suposições de independência condicional entre os valores dos atributos.

[3]Modelos computacionais ou matemáticos que são construídos com base em redes neurais biológicas.

com os outros, portanto eleito o mais eficiente.

Devido ao estilo de vida mais sedentário da sociedade, doenças cardíacas tornaram-se mais comuns nos dias de hoje. Como cada indivíduo tem diferentes valores de pressão arterial, colesterol e taxa de pulso, em busca da prevenção de doenças cardíacas, Thomas e Princy (2016) utilizaram técnicas de classificação, como *Naive Bayes*, k-ésimo vizinho mais próximo (KNN)[4], árvores de decisão, redes neurais e mineração de dados para prever o nível de risco de cada pessoa desenvolver doenças cardíacas com base na idade, sexo, pressão, colesterol, pulsação. Com a maior quantidade de atributos a precisão do nível de risco se torna alta. Com os algoritmos KNN e *iterative dichotomiser 3* a taxa de risco de doença cardíaca pode ser detectada.

[4]Algoritmo de classificação baseado na distância entre os dados.

3 PREPARAÇÃO, ANÁLISE E MINERAÇÃO DOS DADOS

Este capítulo apresenta os dados e as técnicas utilizadas para a realização do trabalho. Assim como, demonstra as análises que foram realizadas no conjunto de dados.

3.1 DADOS UTILIZADOS

Os dados utilizados no trabalho foram fornecidos em abril de 2019 pelo SETIC (Superintendência de Governança Eletrônica e Tecnologia da Informação e Comunicação) e autorizados pela PROGRAD (Pró-Reitoria de Graduação). São informações semestrais de estudantes de 7 cursos de graduação. São esses:

- O ano e semestre ao qual as informações do aluno são referentes;
- O curso a qual o aluno está alocado:

  EEL - Engenharia elétrica (Campus Florianópolis);

  CCO - Ciências da computação (Campus Florianópolis);

  SIN - Sistemas de informação (Campus Florianópolis);

  ECA-FLN - Engenharia de controle e automação (Campus Florianópolis);

  ECA-BLU - Engenharia de controle e automação (Campus Blumenau);

  TIC-DIU - Tecnologias da informação e comunicação (Diurno - Campus Araranguá);

  TIC-NOT - Tecnologias da informação e comunicação (Noturno - Campus Araranguá);

  ENC - Engenharia de computação (Campus Araranguá).

- O centro ao qual o curso pertence, sendo eles: CTC, CTS e Blumenau. O primeiro localizado em Florianópolis, o segundo em Araranguá e o terceiro na cidade de mesmo nome. Além disso, também é disponibilizado as siglas destes centros:

  CTC - Centro Tecnológico - Campus Universitário Reitor João David Ferreira Lima

  CTS - Centro de Ciências, Tecnologias e Saúde - Campus Araranguá

  BLN - Campus Blumenau

- O modo de ingresso na universidade. Exemplo: vestibular.
- A categoria de ingresso do estudante. Exemplo: classificação geral.
- O ano e o semestre de ingresso do estudante;
- A situação do estudante ao final do semestre em questão. Exemplo: regular.
- O índice de aproveitamento semestral (IA[1]);

[1] O IA é uma média simples que é calculada através da soma das notas finais dos alunos nas disciplinas dividida pela quantidade de disciplinas feitas pelo aluno.

- O sexo do estudante, sendo a classificação binária (feminino e masculino);
- A quantidade de disciplinas matriculadas;
- A quantidade de disciplinas aprovadas;
- A quantidade de frequências insuficientes (FI[2]) nas disciplinas.

Os dados analisados são correspondentes a um intervalo de 10 anos, do segundo semestre de 2008 ao segundo semestre de 2018. Alguns cursos foram criados durante esse intervalo. Engenharia de computação foi criado no início de 2011, Engenharia de Controle e Automação (Campus Blumenau) foi criado no início de 2014. O curso de Tecnologias da Informação e Comunicação foi criado no segundo semestre de 2009, porém ele coexistiu por 3 semestres (2009-2, 2010-1 e 2010-2) com 2 códigos de cursos diferentes. Um se refere ao curso diurno, com alunos que não transferiram a matrícula para a modalidade noturna do curso, enquanto o outro se refere a modalidade noturna. Como os dados fornecidos pela universidade se tratam binariamente do sexo do estudante, quando é feita uma referência a gênero, neste trabalho, significa sexo masculino ou feminino.

Nos cursos CCO, SIN, EEL e ECA-BLU, abrem 100 vagas anuais, o curso de ECA-FLN, 72 vagas, e no curso de ENC são 60 vagas anuais. O curso de TIC até o ano de 2017 recebia 100 alunos anualmente, atualmente, recebe 60 estudantes por ano. Os cursos de engenharia possuem 10 fases, o curso de SIN, 9 fases, o curso de CCO, 8 fases, e TIC, 6 fases.

3.2 FERRAMENTAS UTILIZADAS

Nesta seção são apresentadas a linguagem e as ferramentas utilizadas.

3.2.1 Linguagem de programação Python

Python é uma linguagem interpretada, que permite uma programação rápida além da integração com inúmeras bibliotecas e ferramentas. Ela possibilita um fácil aprendizado para novos ou experientes programadores. A linguagem possui estruturas de dados de alto nível e uma abordagem simples, mas eficaz, para programação orientada a objetos (PYTHON, 2019).

Por ser uma linguagem com código aberto, ela possibilita a utilização e distribuição de aplicações, além de permitir uma vasta portabilidade, podendo ser usada em diversas plataformas de programação (OLIPHANT, 2007). Além disso, a vasta comunidade de programadores Python possibilita a existência de muitos tutoriais e materiais de apoio disponíveis gratuitamente.

3.2.1.1 Pandas

A biblioteca Pandas Python transforma a linguagem em uma ferramenta análise de dados. Possui variados métodos de análise estatística e permite a visualização dos dados

[2]Um aluno obtém FI quando tem menos de 75% de presença na disciplina, sendo então reprovado por frequência e recebendo IA igual a 0.

como uma tabela, além da importação e exportação de planilhas em diferentes formatos, facilitando a compreensão e agilizando o processo de análise (MCKINNEY et al., 2010).

3.2.1.2 Scikit-learn

O Scikit-learn é uma biblioteca Python que conta com uma variedade de algoritmos de aprendizado de máquina (PEDREGOSA et al., 2011). Com ela é possível fazer o pré-processamento dos dados, a redução de dimensão, além da utilização de algorítimos de agrupamento como o *K-means*[3] ou classificação como KNN.

3.3 PRÉ-PROCESSAMENTO

Os dados repassados pela UFSC estavam em boas condições para as análises, o que não é comum em conjuntos de dados encontrados em repositórios públicos, por exemplo. Em geral há a necessidade de substituir valores ausentes ou mesmo trocar o tipo de categorização. A UFSC repassou dados de alunos ingressos desde o ano 2000, porém o registro das informações desses alunos só é observado a partir de 2008/2. Sendo assim, o único pré-processamento realizado em todos os dados foi, a exceção da análise estatística de formandos, a remoção desses dados nas demais análises.

A biblioteca pandas da linguagem Python foi utilizada para o carregamento do arquivo contendo os dados, no formato .csv. A figura 3 apresenta um exemplo de como os códigos e tabelas são visualizadas.

Figura 3 – Demonstração de uso da biblioteca pandas.

```
In [1]: import pandas as pd
```

Transformação da planilha em um DataFrame

```
In [2]: str_table = "./PlanilhaRevisao20082.csv"
        df = pd.read_csv(str_table)
```

Seleção dos cinco primeiros dados de algumas colunas

```
In [3]: df[["semestre","centro","siglacentro","anoingresso","semingresso","situacao","sexo"]].head()
```

Out[3]:

	semestre	centro	siglacentro	anoingresso	semingresso	situacao	sexo
0	20082	Tecnologico	CTC	2008	2	Regular	M
1	20091	Tecnologico	CTC	2008	2	Regular	M
2	20092	Tecnologico	CTC	2008	2	Regular	M
3	20101	Tecnologico	CTC	2008	2	Regular	M
4	20102	Tecnologico	CTC	2008	2	Regular	M

Fonte: elaborada pela autora.

[3] Algorítimo de agrupamento que divide os dados dentre k grupos onde cada dado faz parte do grupo mais próximo da média.

3.4 ANÁLISE DOS DADOS

Nesta seção realiza-se as análises dos dados fornecidos pela universidade.

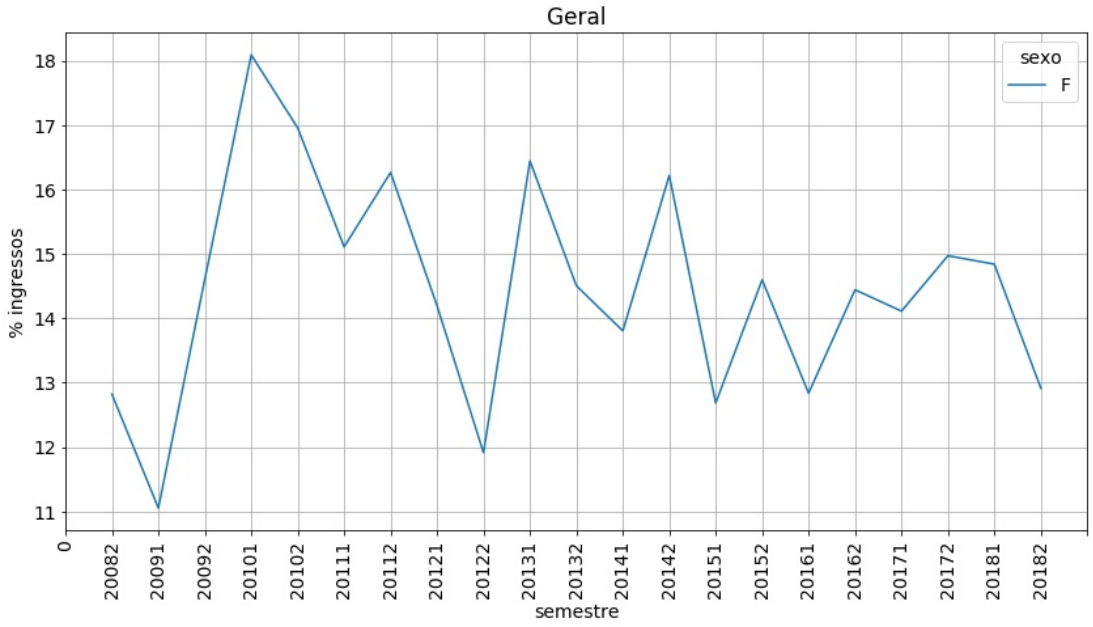
3.4.1 Análises estatísticas

Um dos primeiros passos da extração do conhecimento é conhecer os dados a serem analisados. Como ponto de partida deste trabalho, foram realizadas análises estatísticas, que tem como objetivo extrair informações de maneira analítica.

3.4.1.1 Número de Ingressos

A primeira análise verifica a entrada de alunos por curso e em geral, podendo ou não confirmar a menor entrada feminina e conferir o padrão de entrada dos estudantes. Para tal, os dados dos alunos foram organizados em ordem crescente através da função pandas *sort_values*. Posteriormente foram removidos os dados repetidos dos alunos pela matrícula através da função pandas *drop_duplicates*. Com isso apenas restam as informações dos alunos referentes ao primeiro semestre cursado. Então foi realizado o agrupamento por semestre através da função pandas *groupby*, contando quantos alunos de cada gênero ingressaram por semestre e posteriormente fazendo a média de ingresso (função pandas *mean*). Por fim, é feito um percentual entre a quantidade de homens e mulheres.

Gráfico 1 – Percentual das mulheres ingressantes.

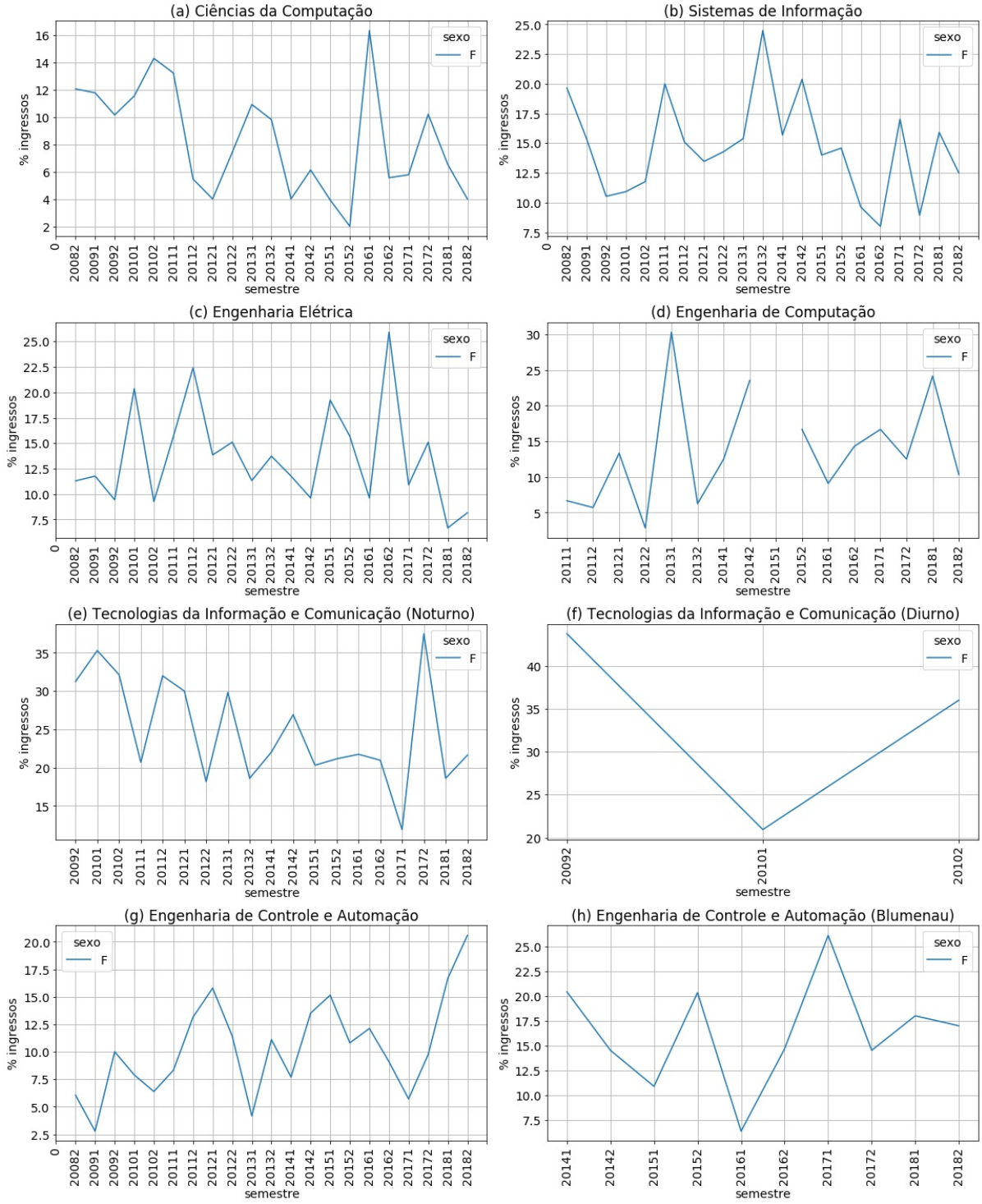


Fonte: elaborado pela autora.

O Gráfico 1 apresenta o percentual de ingressantes mulheres nos cursos analisados (Como o total do percentual é a soma de homens e mulheres, os dados dos alunos do sexo masculino foram retirados). O primeiro semestre de 2009 obteve o menor percentual de estudantes ingressantes, 11,05%, e o primeiro semestre de 2010 foi o que obteve o maior

percentual de alunas ingressas, 18,09%. O pequeno aumento de mulheres em 2010, mais 3,56% no percentual total de alunas, pode ser consequência do início do curso de TIC no segundo semestre de 2009. A média percentual de entrada feminina é de 14,44%, com um desvio padrão de 1,73.

Gráfico 2 – Porcentagem de estudantes femininas ingressantes, de todos os cursos analisados.



Fonte: elaborado pela autora.

Quando observa-se a entrada de alunos por curso, Gráfico 2, nota-se que os cursos de ENC e ECA-FLN são os que apresentam a maior disparidade de gênero entre os cursos analisados. Em média, os ingressantes do sexo feminino são 13,66% no curso de ENC, com um desvio padrão de 8,147, e 10,39% em ECA-FLN, com um desvio padrão de 4,42. A ENC contou em 2012/2 com apenas uma menina e em 2015/1 com nenhuma. Em 2009/1 e em 2015/2 os cursos de ECA e CCO, ambos alocados em Florianópolis, também só obtiveram a entrada de uma estudante. Ressalta-se que o curso de ECA em Florianópolis atingiu o seu maior percentual de ingressantes do sexo feminino em 2018/2 com cerca de 21%, enquanto o mesmo curso na localidade de Blumenau obteve em 2014/1 e 2015/2 percentuais acima de 20%, atingindo seu pico em 2017/1 com 26,15% de ingressantes.

Dentre os cursos analisados o curso de CC tem o menor percentual de ingresso de mulheres, com uma média de 8,33% de ingressantes femininas e um desvio padrão de 4,01, enquanto o curso com maior número de representantes do sexo feminino é o de TIC, com em média 24,78% de mulheres em sua modalidade noturna, com desvio padrão de 6,89, e 33,56% ingressantes na modalidade diurna, com desvio de 11,6 (O curso no período diurno só incluiu novos alunos entre 2009/2 e 2010/2). Os cursos EEL e SIN seguem a média geral, com desvios de 4,95 e 4,13, respectivamente.

3.4.1.2 Rendimento acadêmico

Para avaliar o rendimento acadêmico dos estudantes foram utilizados os índices acadêmicos semestrais (IA). Vale ressaltar que, para ser aprovado, um aluno deve obter índice superior a 6 na disciplina cursada. Sendo assim, se um aluno é aprovado em todas as disciplinas, no semestre seu índice deve ser igual ou superior a 6.

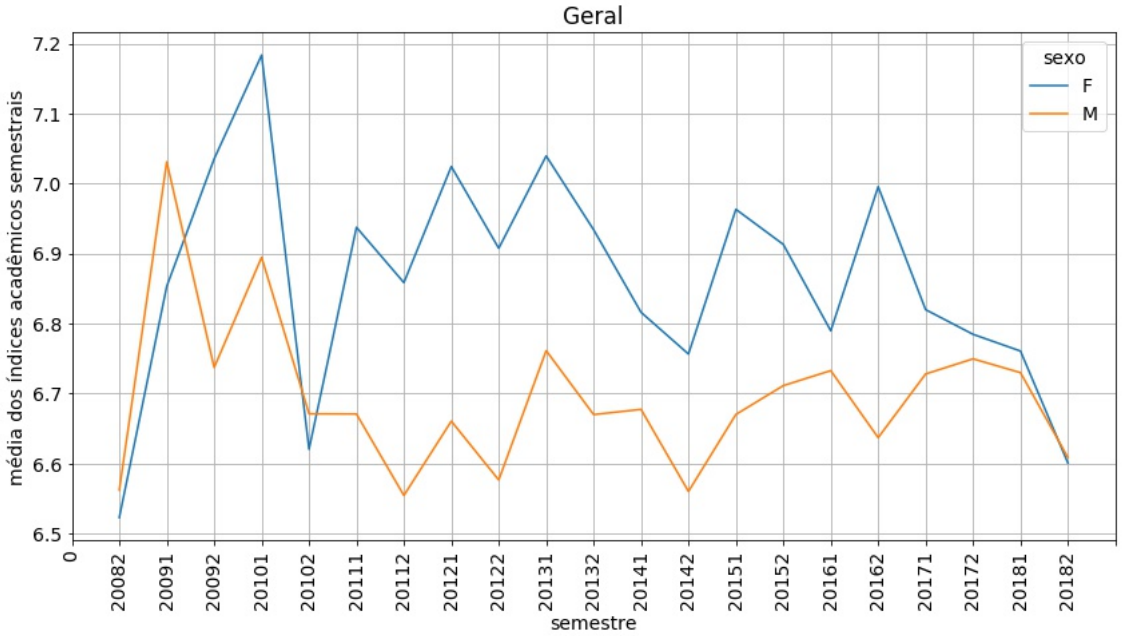
No processo de limpeza dos dados, foram retirados os dados onde o aluno, naquele semestre, obteve IA igual a zero e/ou não obteve aprovação em nenhuma disciplina, pois isto não reflete uma dedicação do aluno, mostrando que possivelmente o mesmo desistiu do semestre em questão. Também foram removidos dessa análise, os semestres terminados em 3, que se referem a disciplinas feitas durante o período de férias ou atrasadas, por exemplo, portanto não refletindo o desempenho de todos os alunos, mas sim, apenas da parcela que fez a determinada disciplina. Isso torna esses semestres específicos demais para serem acrescentados em um cálculo geral de desempenho.

Fez-se uma análise do rendimento acadêmico dos alunos por sexo. Através da função *groupby*, da biblioteca pandas, os dados foram agrupados por semestre e por gênero do aluno. Além disso foi possível obter a média dos índices acadêmicos semestrais gerais e por curso através da função *mean* do pandas.

O Gráfico 3 mostra a média dos índices semestrais considerando todos os cursos avaliados de um modo geral.

As mulheres apresentam índices maiores quando olhamos para a média de todos os cursos. Nos semestres 2008/2, 2009/1, 2010/2 e 2018/2 os estudantes do sexo masculino obtiveram índices superiores, contudo, com exceção de 2009/1, em todos os outros semestres esses valores superiores estão atrelados a quedas nos índices de ambos os sexos. A maior média de IA feminino foi de 7,18 em 2010/1, enquanto o masculino foi de 7,03 em 2009/2. Os menores índices acadêmicos gerais foram de 6,52 e 6,55, feminino e masculino, respectivamente. Além disso, os estudantes homens mantiveram uma média de IA de 6,69 com desvio padrão de 0,11, enquanto as estudantes do sexo feminino obtiveram uma média de 6,86 com desvio de 0,16. Há uma tendência geral nos IA. Percebe-se que

Gráfico 3 – Média de índices acadêmicos dos estudantes por semestre e sexo, de todos os cursos analisados.



Fonte: elaborado pela autora.

apesar do número de mulheres ser menor o Gráfico 3 corrobora a hipótese 3, que diz que o desempenho acadêmico das mulheres que escolhem a carreira de tecnologia é melhor.

Os Gráficos 4 apresentam as médias dos índices de acordo com os cursos tecnológicos.

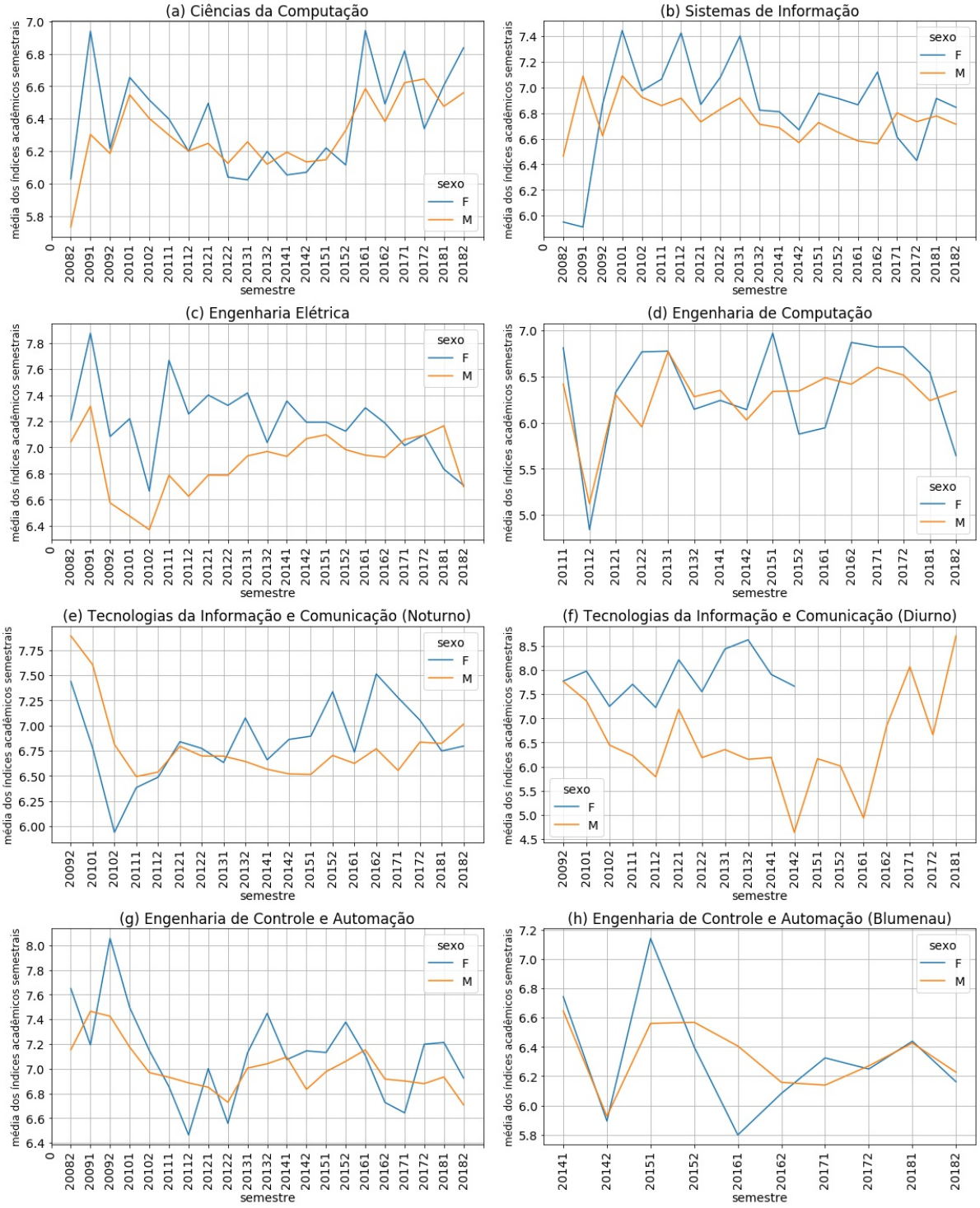
No curso de CCO, Gráfico 4 (a), dos 21 semestres analisados, em 7 deles os estudantes homens obtiveram índices superiores. O maior IA médio entre os homens foi de 6,64 e o mínimo 5,73, uma média de 6,30 com desvio padrão de 0,22. Já entre as mulheres a maior média de índice acadêmico foi de 6,94 e o mínimo 6,02, com uma média de IA de 6,39 e desvio padrão de 0,31. Os homens em 2008/2 obtiveram em média notas abaixo do necessário para aprovação no semestre.

Em SIN, Gráfico 4 (b), os índices acadêmicos masculinos foram maiores em 4 semestres. Entre os estudantes homens o maior IA foi de 7,09 e o menor 6,46, com uma média de 6,76 e desvio padrão de 0,17. Entre as estudantes o maior índice acadêmico foi de 7,44 e o mínimo 5,91, com uma média de 6,85 e desvio padrão de 0,40. As mulheres em 2008/2 e 2009/1 obtiveram em média notas abaixo do necessário para aprovação no semestre.

Os estudantes homens do curso de EEL, Gráfico 4 (c), só obtiveram média de índice acadêmico superior ao feminino em dois semestres, 2017/1 e 2018/1. Além disso, sua maior média de IA foi 7,31 e a menor 6,37, com uma média de 6,89 e desvio padrão de 0,24. Já as mulheres atingiram um índice acadêmico de 7,87 no máximo e de 6,66 no mínimo, com uma média de 7,20 e desvio padrão de 0,28.

Em ENC, Gráfico 4 (d), as mulheres obtiveram em média o IA de 6,34 com desvio padrão de 0,57, enquanto os homens, conseguiram uma média de índice acadêmico de 6,28 com desvio de 0,36. A maior nota e a menor nota alcançada entre os estudantes masculinos foi 6,77 e 5,12, respectivamente. As estudantes obtiveram a média máxima 6,97 e mínima de 4,84 na média dos índices acadêmicos. O menor IA feminino e masculino

Gráfico 4 – Média de desempenho (IA) dos estudantes por semestre e sexo, por curso



Fonte: elaborado pela autora.

foi atingido no semestre de 2011/2.

No curso de ECA-FLN, Gráfico 4 (g), os índices acadêmicos das alunas, em geral, foi superior ao dos alunos. As médias dos IA's atingidas pelos sexos são de 7,12 para as mulheres, com desvio padrão de 0,37, e 7,00 para os homens, com desvio padrão de 0,19. Os índices acadêmicos máximos e mínimos femininos ficaram entre 8,06 e 6,46, enquanto

os masculinos ficaram entre 7,46 e 6,71.

Sendo o único dos cursos analisados onde os homens atingem mais semestres com índices acadêmicos maiores que as mulheres o curso de ECA-BLU, Gráfico 4 (h), também é o mais novo. Os homens atingiram o IA máximo de 6,64, enquanto as mulheres atingiram um IA de 7,14. Já os valores mínimos são 5,80 e 5,92, para mulheres e homens, respectivamente.

O TIC, Gráfico 4 (e) e (f), em suas duas modalidades as mulheres obtiveram índices acadêmicos maiores que os homens. Na modalidade noturna, os estudantes homens obtiveram 8 semestres de 19 com notas superiores. As mulheres obtiveram um IA superior e inferior de 7,51 e 5,94, respectivamente, com índice acadêmico médio de 6,85 e desvio padrão de 0,38. Já os homens obtiveram o maior índice acadêmico de 7,89 e o menor de 6,49, com média de 6,79 e desvio de 0,37. No curso diurno as mulheres sempre obtiveram índices superiores, porém no segundo semestre de 2014 apenas restaram homens nesse período. Os índices acadêmicos máximos e mínimos atingidos pelas mulheres são de 8,63 e 7,22, com média de 7,85 e desvio de 0,45. Já os homens de 2009/2 até 2018/1, obtiveram IA máximo de 8,71 e mínimo de 4,64, totalizando uma média de 6,54 e um desvio de 1,01.

Dos cursos verificados, os maiores índices acadêmicos, de ambos os sexos, vieram da modalidade diurna do curso de TIC, e os menores, 4,84 para as mulheres, do curso de engenharia de computação e 4,64, para os homens de TIC diurno. Outros dados a serem levados em consideração é o grande desvio padrão das notas do curso de TIC diurno e o fato do curso com maior média de índice acadêmico entre as mulheres ser o de EEL, curso com uma maior proporção feminina entre os observados. Entre os homens o campeão na média de IA é o curso de ECA-FLN, que conta com uma das porcentagens mais baixas de ingresso de meninas.

3.4.1.3 Formandos

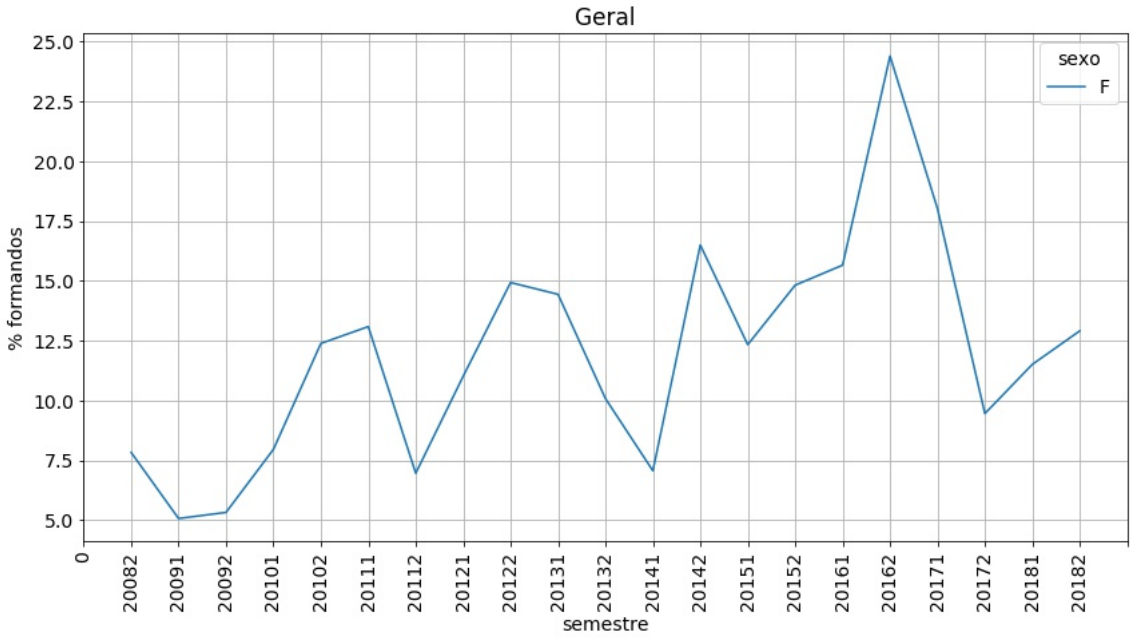
Esta análise possibilita verificar o percentual de formandos. A análise de formandos foi a única onde foi necessário a utilização de todos os dados fornecidos pela UFSC (2000-2008), para o número de formandos divulgado pela UFSC ser o mesmo que o encontrado na análise.

Foram mantidos os dados de alunos com a situação "Formado", ou seja, as linhas onde um aluno obteve a situação que define que esse estudante se formou. Através da função *groupby* da biblioteca pandas foi feita a contagem de alunos por sexo e semestre. O percentual semestral de ingresso de homens e mulheres foi realizado através da função *apply*. Lembrando que os dados dos homens foram removidos por serem um espelho do percentual de mulheres.

O Gráfico 5 apresenta o percentual de estudantes formados do sexo feminino por semestre. O curso de ECA-BLU, por ser um curso novo, ainda não contou com nenhum formado e os cursos do CTS ainda não haviam realizado a formatura quando os dados foram extraídos. O semestre com menor percentual de formandas é o de 2009/1, 5,06%, e o maior é o de 2016/2, com 24,39%. Enquanto os dados gerais de ingresso apresentam uma média de 14,44% , com um desvio padrão de 1,73, de mulheres ingressantes, tem-se 11,98% no somatório de formandas em todos os cursos, com um desvio padrão de 4,667. Isso mostra que o percentual de formandas é aproximadamente um reflexo da entrada.

Analisando os Gráficos 6 pode-se observar que os cursos SIN e EEL dispõem de formandas em todos os semestres analisados. O Gráfico 6 (b) mostra uma tendência de

Gráfico 5 – Porcentagem de formandas em todos os cursos analisados.



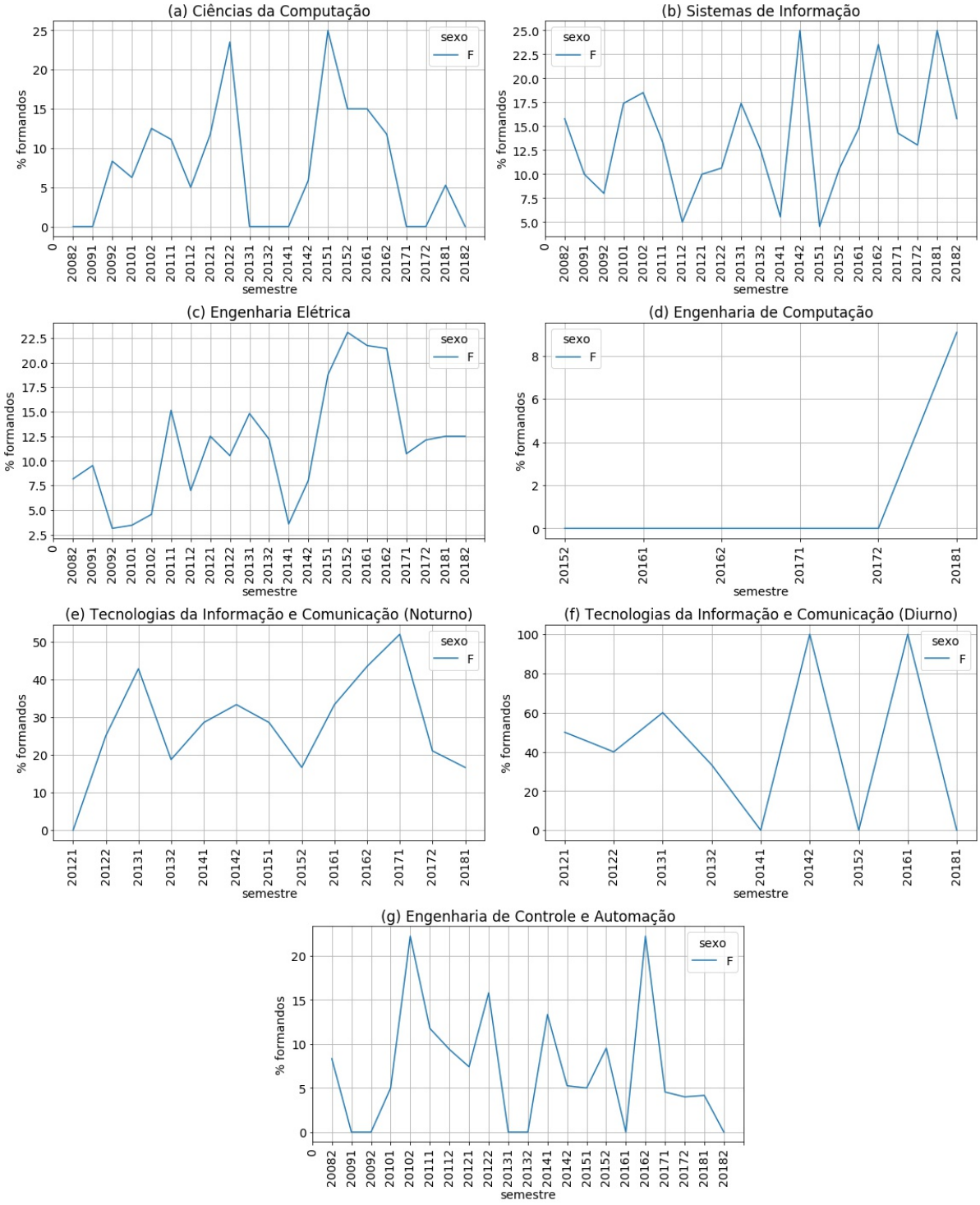
Fonte: elaborado pela autora.

aumento no percentual de mulheres que concluíram a graduação do curso de SIN desde 2015. O mesmo tem uma média de 13,84% de formandas com um desvio padrão de 6. Em 2014/2 e 2018/1 atingiu 25% de concluintes do sexo feminino e seu menor resultado foi em 2015/1, 4,54%. Já o curso de EEL obteve seu maior resultado em 2015/2, 23.077% de formandas, e o menor em 2009/2, com 3,12% de formandas. Os dois cursos apresentam uma certa tendência a igualdade nos formandos, com números de formandas crescendo.

Os cursos de ENC e ECA-FLN são os que apresentam o menor ingresso de mulheres dentre os cursos analisados. O curso de ENC só teve uma estudante formada, em 2018/1. Já o curso de ECA-FLN, contou com 6 semestres onde não houveram formandas, porém o curso já atingiu 22,22% de formandas nos semestres 2010/2 e 2016/2. A média de concluintes mulheres de ECA-FLN é de 7,04%, com um desvio padrão de 6,87. O curso de CCO tem 8 semestres sem a presença de formandas mulheres no período analisado, e um percentual máximo de 25%. A média de concluintes mulheres é de 7,45% com um desvio de 7.8.

Os Gráficos 6 (e) e (f), do curso de TIC, apresentam os maiores percentuais dentre os cursos analisados, atingindo um número máximo de 52% de formandas, TIC-NOT, e 100% de concluintes, TIC-DIU. A modalidade noturna contou apenas com o semestre de 2012/1 zerado e tem uma média de formandas de 27.71% com um desvio padrão de 13.77. A modalidade diurna conta com uma média de 42,59% de concluintes e desvio padrão de 39.50.

Gráfico 6 – Porcentagem de formandas por curso.



Fonte: elaborado pela autora.

3.4.1.4 Modo de Ingresso

Essa análise foi realizada devido ao acesso a esses dados, o que foi visto como uma oportunidade de verificar o desempenho desses alunos de acordo com o modo de ingresso

e tentar extrair outras informações além da análise por sexo.

Primeiramente os códigos fornecidos pela UFSC foram agrupados em 5 categorias através das funções *isin*[4] e *loc*[5]. As formas de ingresso são:

Vestibular Onde o aluno realiza uma prova para talvez ingressar na universidade;

Transferência interna O aluno pede transferência dentro da UFSC de um curso para outro;

Transferência externa O estudante ingressa através de outra instituição de ensino;

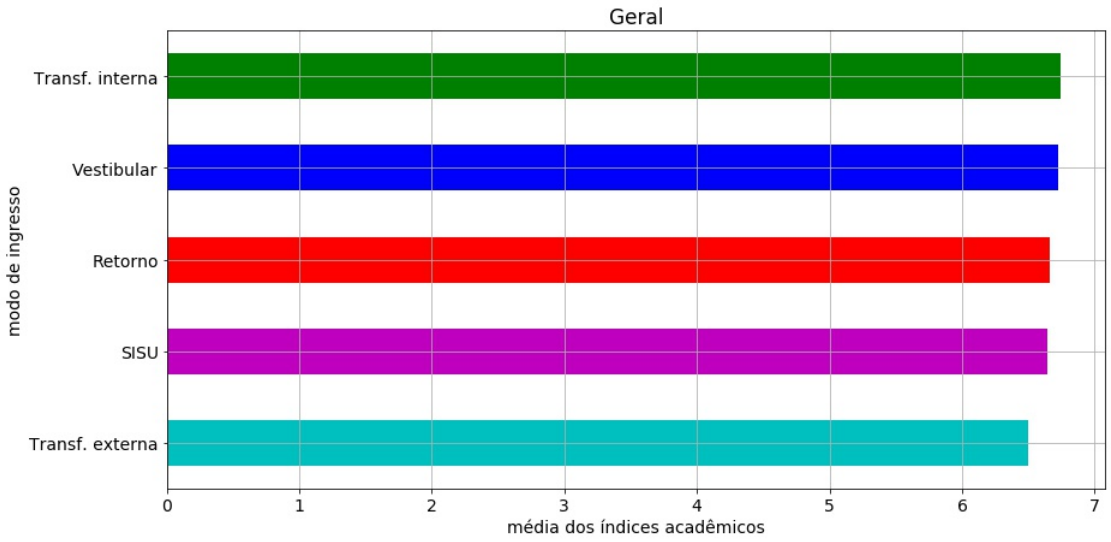
Retorno Alunos já formados na UFSC pode requisitar o mesmo para ingressar em outro curso;

SISU (Sistema de Seleção Unificada) Onde o aluno procura o ingresso submetendo a nota do ENEM (Exame Nacional do Ensino Médio).

Como na análise do desempenho, os semestres onde o aluno não mostrou um comprometimento com as disciplinas foram removidos. Os dados foram agrupados, *groupby*, de acordo com a forma de ingresso e a média dos índices acadêmicos, função *mean*, foi realizada.

O Gráfico 7 apresenta o modo de ingresso dos estudantes na universidade pela média dos IA's. Transferência interna obteve a maior média, porém os valores são relativamente semelhantes, sendo Transferência externa o mais distinto. O melhor resultado do primeiro se deve provavelmente ao prévio conhecimento do aluno as práticas da universidade, diferente da transferência externa.

Gráfico 7 – Modo de ingresso dos estudantes pela média dos índices acadêmicos de todos os cursos

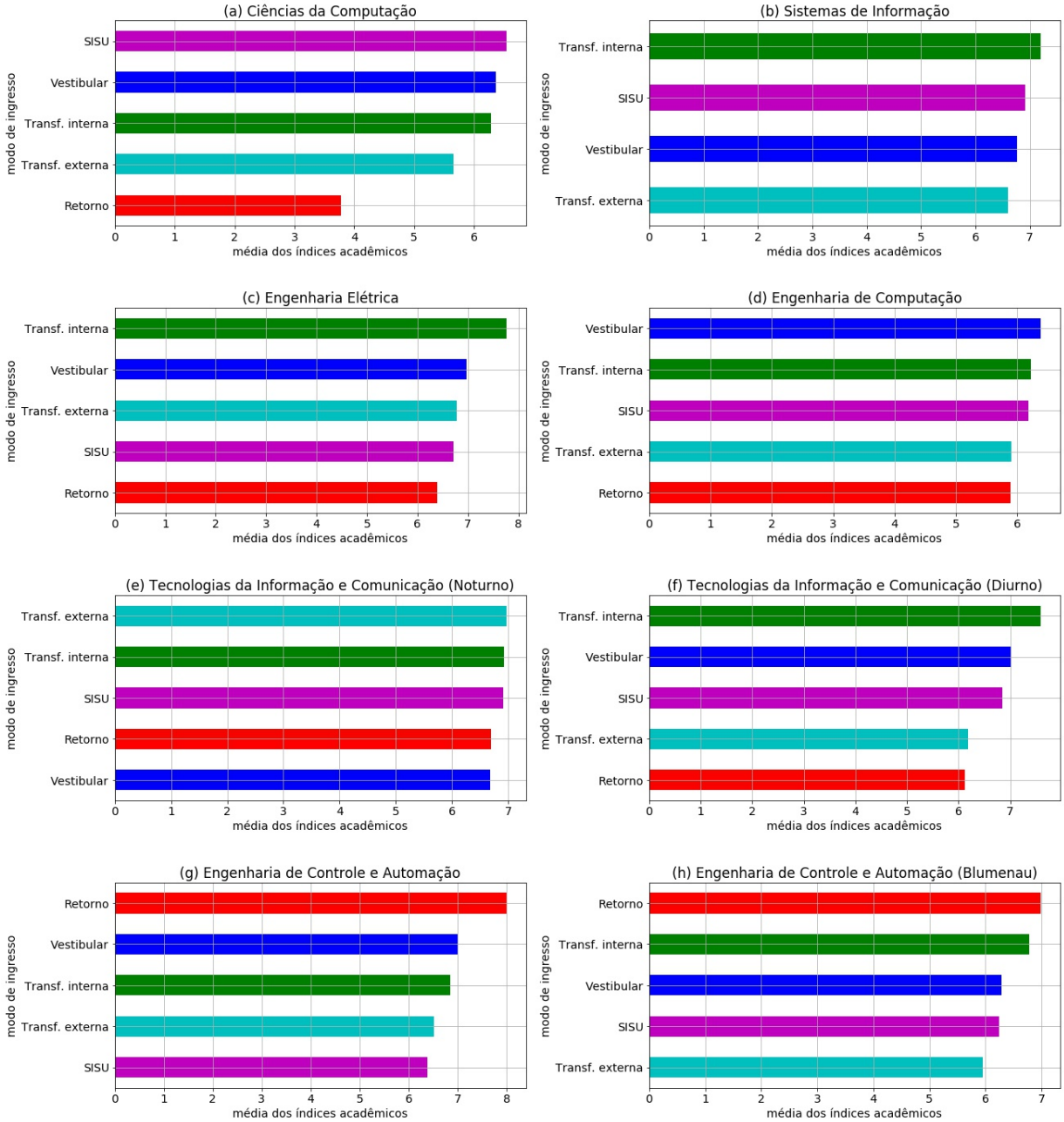


Fonte: elaborado pela autora.

No Gráfico 8 observa-se o modo de ingresso dos estudantes na universidade pela média dos IA's dos cursos analisados separadamente.

[4]Função pandas utilizada para verificar se um dado pertence uma coluna.
[5]Esta função da biblioteca pandas é usada para a localização de dados.

Gráfico 8 – Modo de ingresso dos estudantes pelos índices acadêmicos, por curso.



Fonte: elaborado pela autora

O curso de CCO, Gráfico 8 (a), é o único onde o SISU tem os melhores índices. A média de IA dos retornos e das transferências externas é abaixo do IA necessário para aprovação dos alunos. Este fato se repete no curso de ENC. Esse baixo índice de retorno pode ser motivado pelos estudantes já serem formados e já estarem imersos profissionalmente, os levando a terem um menor desempenho acadêmico devido a falta de tempo para se dedicarem aos estudos.

Os cursos de SIN, EEL e TIC-DIU, tem altos índices de transferência interna. Isso levanta a hipótese de uma troca entre cursos com bases semelhantes. O curso de TIC-DIU está localizado na mesma cidade que o curso de ENC, por ter um menor tempo de conclusão, muitos alunos da ENC se transferem para o curso de TIC. O mesmo pode

acontecer com os cursos de CCO e SIN, onde os alunos do primeiro migram para o outro, e EEL com as outras engenharias do CTC.

Os cursos de ECA-BLU e ECA-FLN tem altos IA's de alunos ingressos por retorno, porém esses alunos são minoria em quantidade, 2,5% e 0.31% do total de alunos, respectivamente. Isso pode ser um fato isolado, devido a esses alunos em especial obterem boas notas, em comparação com os outros cursos.

3.4.1.5 Categoria de Ingresso

Esta análise tem a mesma motivação que a anterior, visto a possibilidade de verificar social e economicamente o ingresso desses alunos.

Primeiramente os códigos fornecidos pela UFSC foram agrupados em 3 categorias através das funções *isin* e *loc*. Vale ressaltar que apenas os alunos ingressos nos modos vestibular e SISU possuem essa informação. As categorias de ingresso são:

Classificação geral O estudante não optou por nenhum tipo de cota;

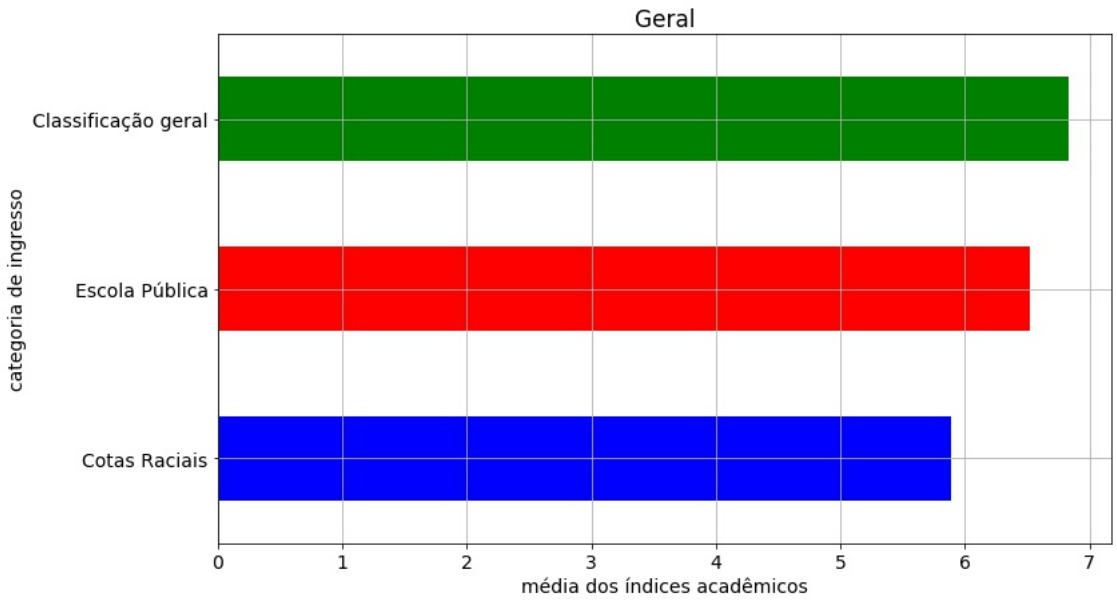
Cotas raciais Aluno se enquadra na categoria pretos, pardos, quilombolas e indígenas;

Escola pública Estudante fez o ensino médio em escola pública.

Os semestres onde o aluno não mostrou um comprometimento com as disciplinas foram removidos, como já observado em análises anteriores. Os dados foram agrupados, *groupby*, de acordo com a categoria de ingresso e a média dos índices acadêmicos.

O Gráfico 9 apresenta a média dos IA's pela categoria de ingresso do estudante. Os alunos oriundos da classificação geral apresentam índices acadêmicos superiores aos demais. Os estudantes provenientes de cotas raciais (pretos, pardos, quilombolas e indígenas) obtiveram IA inferior a 6, a nota necessária para conseguir aprovação em uma disciplina.

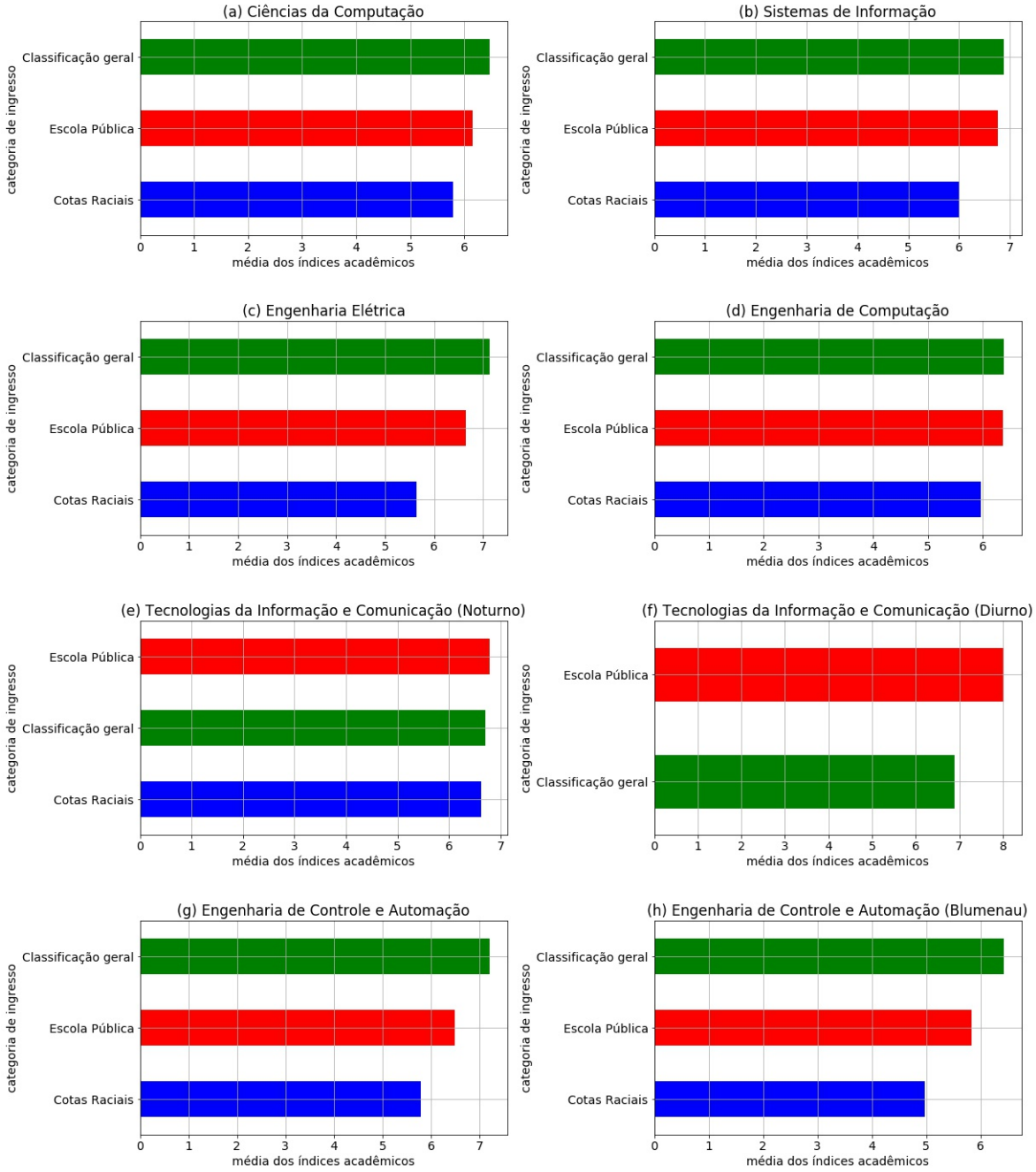
Gráfico 9 – Categoria de ingresso somando o IA de todos os cursos analisados.



Fonte: elaborado pela autora.

O Gráfico 10 apresenta a média dos IA's pela categoria de ingresso do estudante por curso. Com exceção dos cursos de TIC-NOT e SIN, nos demais cursos as cotas raciais obtiveram IA abaixo de 6, destacando o curso de ECA-BLU, com IA abaixo de 5. Porém, esse baixo índice dos alunos de ECA-BLU pode ser relativo a baixa quantidade de estudantes nessa situação, 2,17% do total. No curso de TIC, nas duas modalidades, os estudantes provenientes de escola pública obtiveram índices superiores aos demais, sendo que, na modalidade noturna, os IA's são praticamente iguais em todas as categorias. Nos cursos de EEL e ECA-FLN, os alunos provindos da classificação geral alcançaram IA superior a 7.

Gráfico 10 – Categoria de ingresso por IA de cada curso.



Fonte: elaborado pela autora.

Essa inferioridade nos IA's de alunos provenientes de cotas (tanto escola pública quanto raciais) apenas confirma a necessidade das mesmas, pois esses estudantes provavelmente nunca teriam condições de competir com alunos provenientes de escolas particulares ou com realidades sociais mais abastadas. Esses dados confirmam a importância vital deste incentivo e de outras políticas de acompanhamento na vida desses estudantes.

A partir do ano de 2013 a UFSC começou a separar o ingresso por alunos não optantes e duas grandes subcategorias de alunos oriundos de escola pública, renda até 1.5 salario mínimo e renda maior que 1,5 salários mínimos. Assim os alunos não optantes, em geral, são oriundos de escolas privadas, sendo os demais obrigatoriamente provenientes de escolas públicas. Sendo assim, os códigos foram separados de acordo com essa classificação, sendo eles:

Não optantes O estudante não optou por nenhum tipo de cota;

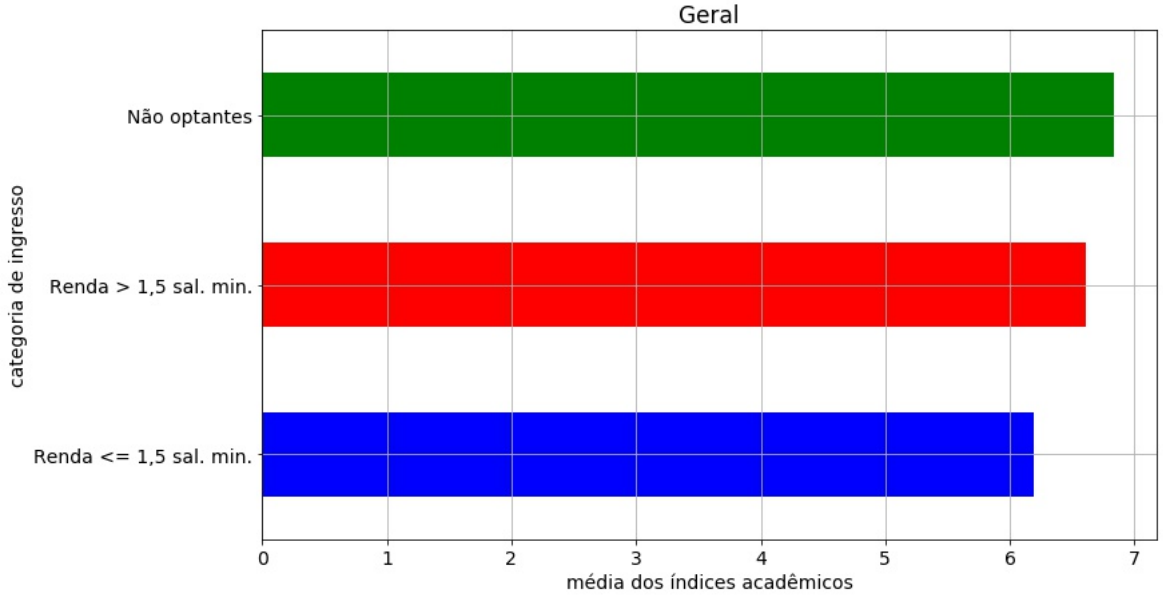
Renda <= 1,5 sal. min. Estudante que cursou o ensino médio em escola pública e a família possui renda inferior a um salário mínimo e meio por pessoa;

Renda > 1,5 sal. min. Estudante que cursou o ensino médio em escola pública e a família possui renda superior a um salário mínimo e meio por pessoa.

Essa análise possibilita verificar as diferenças entre alunos provenientes de diferentes tipos de escolaridade e condições financeiras.

O Gráfico 11 apresenta a média dos IA's de todos os cursos, de acordo com a categoria de ingresso. Na média geral, os alunos não optantes obtiveram IA superior, porém todas as categorias alcançaram índices acima de 6. Isso fortalece a ideia de que o nível de escolaridade e a situação financeira do estudante afeta seu desempenho acadêmico, tendo os alunos de baixa renda os menores índices.

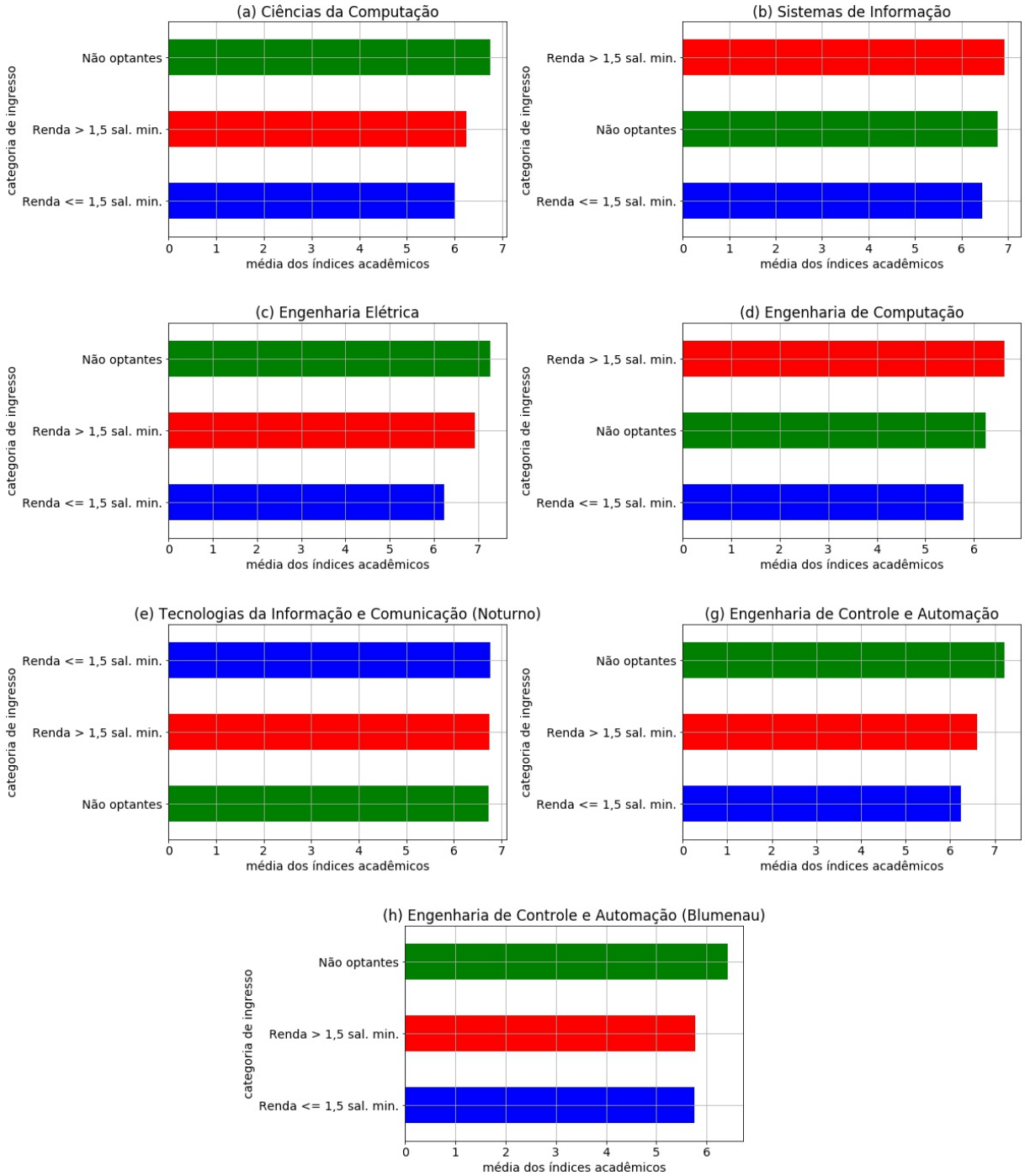
Gráfico 11 – Categoria de ingresso por renda e IA de todos os cursos analisados.



Fonte: elaborado pela autora.

No Gráfico 12 podemos observar, por curso, a média dos IA's, de acordo com a categoria de ingresso. O curso de TIC-DIU só recebeu ingressantes antes da mudança na categorização, ou seja, nesta análise o curso não apresenta dados.

Gráfico 12 – Categoria de ingresso por renda e IA, por curso.



Fonte: elaborado pela autora.

Nos cursos de ENC e ECA-BLU os estudantes oriundos de escola pública com renda até 1,5 salário mínimo obtiveram IA inferior a 6, porém na modalidade noturna do curso de TIC, os estudantes dessa mesma categoria atingiram índices superiores aos das outras categorias.

Os alunos dos cursos SIN e ENC com os melhores índices, são os pertencentes a categoria escola pública com salário mínimo superior a 1,5. Os estudantes pertencentes a categoria de não optantes dos cursos EEL e ECA-BLU, foram os únicos a conseguirem

índices superiores a 7.

A análise mostra que alunos de baixa renda e/ou participantes de cotas raciais tem um baixo desempenho acadêmico. Essa condição pode ser fruto de um despreparo na fase escolar ou mesmo inaptidão da universidade em atender esses alunos. Vieira, Dell'Agli e Caetano (2019) destacam que a política de cotas se mostra efetiva diante das desigualdades sociais e raciais. Elas servem como ferramenta de inclusão social e educativa, garantindo acesso aos bens culturais e intelectuais por todos os cidadãos, desmistificando assim, a concepção tradicional de uma elite intelectual e econômica (VIEIRA; DELL'AGLI; CAETANO, 2019).

3.4.1.6 Padrão de evasão

Para conseguir avaliar o padrão de evasão dos alunos analisados, fez-se necessário uma estimativa. Como os dados não continham as disciplinas cursadas pelos estudantes, a solução encontrada foi analisar dois pontos:

Onde o aluno estaria quando desistiu Dividindo a soma total de disciplinas matriculadas pela média das disciplinas feitas semestralmente no curso;

Onde o aluno estava quando desistiu Dividindo a soma total de disciplinas aprovadas pela média das disciplinas feitas semestralmente no curso.

A média das disciplinas feitas semestralmente no curso corresponde a quantidade de matérias, em média, que o currículo do curso determina que um estudante faça para se formar no tempo estimado pela coordenação. As médias utilizadas nessa estimativa são:

5 disciplinas SIN e ECA-BLU;

6 disciplinas CCO, EEL, TIC e ECA-FLN;

7 disciplinas ENC.

Os Gráficos 13 apresentam a distribuição de evasão dos discentes por sexo, sendo que, o tamanho do circulo mostra a quantidade de alunos ligados ao ponto em questão. No eixo x se encontra o semestre onde o aluno estaria se possuísse aprovação em todas as disciplinas matriculadas e o eixo y onde ele estava quando desistiu do curso.

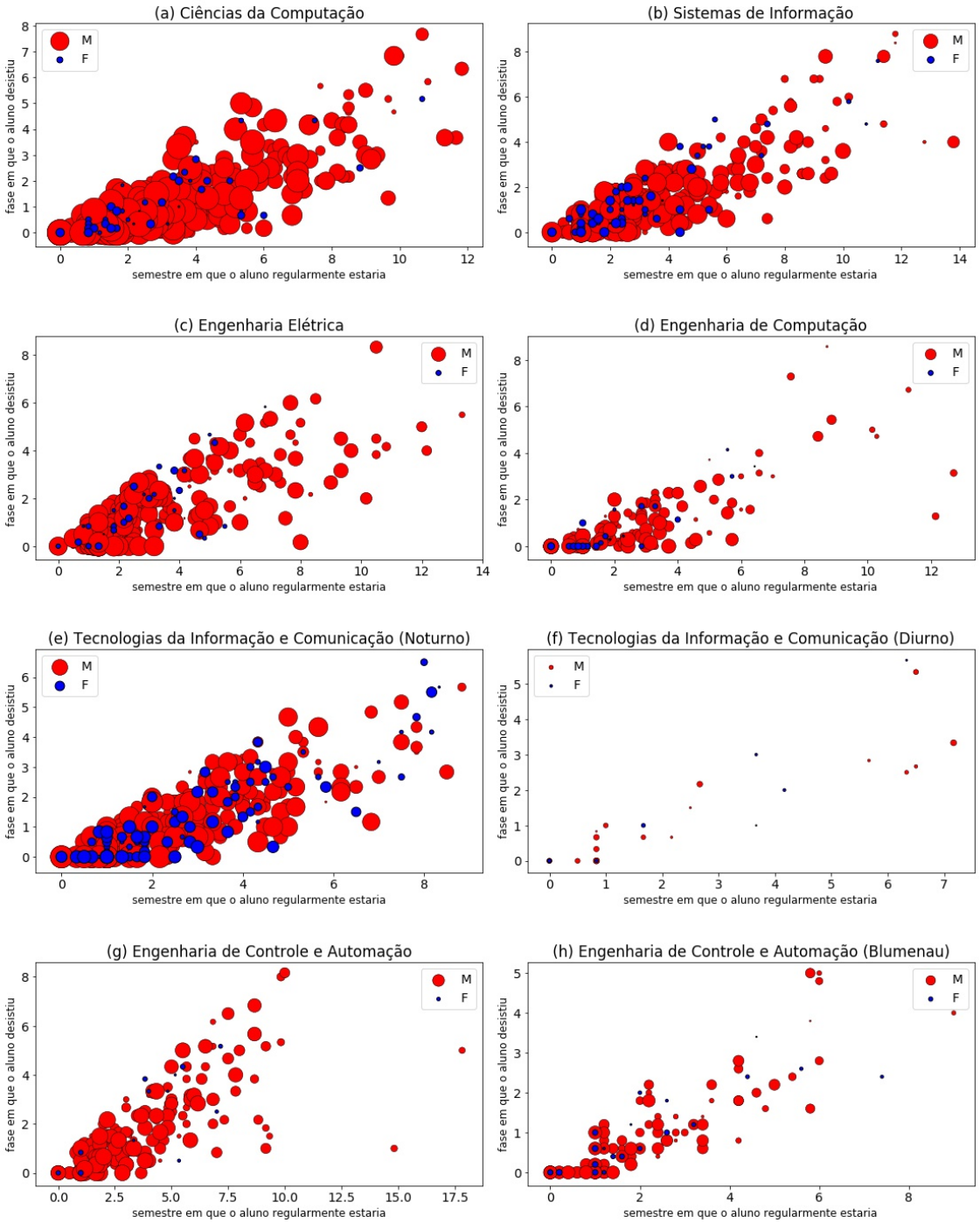
Em todos os cursos pode-se verificar uma tendência de evasão nas fases iniciais. Vale ressaltar que o ingresso no curso TIC-DIU ocorreu durante apenas 3 semestres, resultando em apenas alguns alunos ligados ao curso, por isso o Gráfico 13 (f) apresenta poucas ocorrências.

Todos os cursos apresentam um agrupamento nas fases iniciais, segundo os gráficos os alunos desistem do curso, em geral, antes da quarta fase. Os cursos de CCO, SIN, EEL e TIC abrem 50 vagas de ingresso por semestre, por isso tem uma maior concentração de dados. Porém vale ressaltar que, com base na média de ingresso dos dados fornecidos, a média de ingresso dos cursos SIN e CCO é similar, 52 e 55 alunos por semestre, respectivamente, significando que o o número de evasão do curso de CCO é maior que os demais. Os cursos de ENC e ECA-FLN tem aproximadamente o mesmo número de ingressantes, porém o curso de ENC-FLN tem um número de desistências superior.

Essa análise confirma a hipótese 4, constatando que em geral, a evasão acontece nos dois primeiros anos do curso, para ambos os sexos . As primeiras fases dos cursos,

com exceção aos cursos de SIN e TIC, contam com disciplinas semelhantes como cálculo, física, álgebra e geometria analítica, essas disciplinas podem estar relacionadas com as desistências nas fases iniciais. O curso de TIC é o que conta com a maior igualdade de entrada entre homens e mulheres, e isso pode ser visto como um reflexo nas desistências, já que o Gráfico 13 conta com uma maior distribuição de homens e mulheres.

Gráfico 13 – Evasão dos alunos com base na quantidade de disciplinas cursadas.



Fonte: elaborado pela autora.

3.4.1.7 Reprovação

Para verificar se há influência da quantidade de reprovações na situação do aluno, extraiu-se a média de reprovação pela última situação do estudante. Através da função *sort_values* os dados foram ordenados por semestre em ordem decrescente. A quantidade de disciplinas reprovadas foi obtida através da subtração da quantidade de disciplinas matriculas e aprovadas. Com as funções da biblioteca pandas *groupby*, *transform* e *drop_duplicates*, foi feita a soma do número de reprovações dos alunos de acordo com a matrícula e apenas a primeira ocorrência da matrícula foi mantida (Última situação do aluno nos dados analisados, lembrando que os dados estavam em ordem decrescente).

As funções *groupby* e *mean* extraíram a média do número de reprovações dos alunos por situação final. As situações "Regular"e "Trancado", foram removidas da análise pelo aluno ainda estar vinculado a UFSC e a situação "Falecido"foi removida por apresentar uma saída por força maior. A cinco situações finais são:

Desistente O aluno desistiu do curso;

Formado Estudante concluiu a graduação;

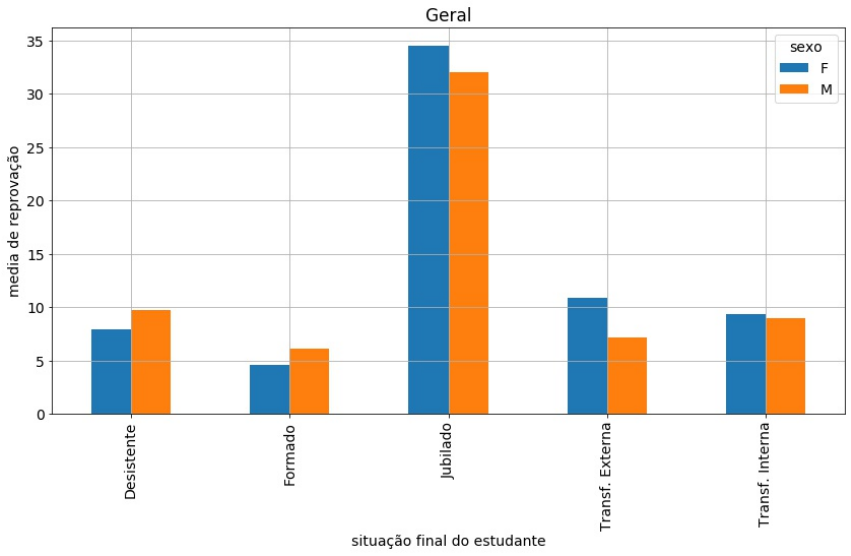
Transferência interna O aluno pediu transferência para outro curso da UFSC;

Transferência externa Estudante foi transferido para outra instituição de ensino;

Jubilado O discente excedeu o prazo máximo de conclusão do curso, ou seja, se o aluno não termina o curso dentro do prazo estabelecido, o mesmo perde a matrícula na universidade.

O Gráfico 14 apresenta os dados gerais com as médias de todos os cursos analisados. Os alunos jubilados apresentam uma média de reprovação expressivamente maior que os demais, devido ao fato desses estudantes se matricularem para não perder sua matrícula na universidade, porém excedem o tempo de formação máximo exigido.

Gráfico 14 – Média de disciplinas reprovadas dos alunos pela situação final do estudante, de todos os cursos analisados.



Fonte: elaborado pela autora.

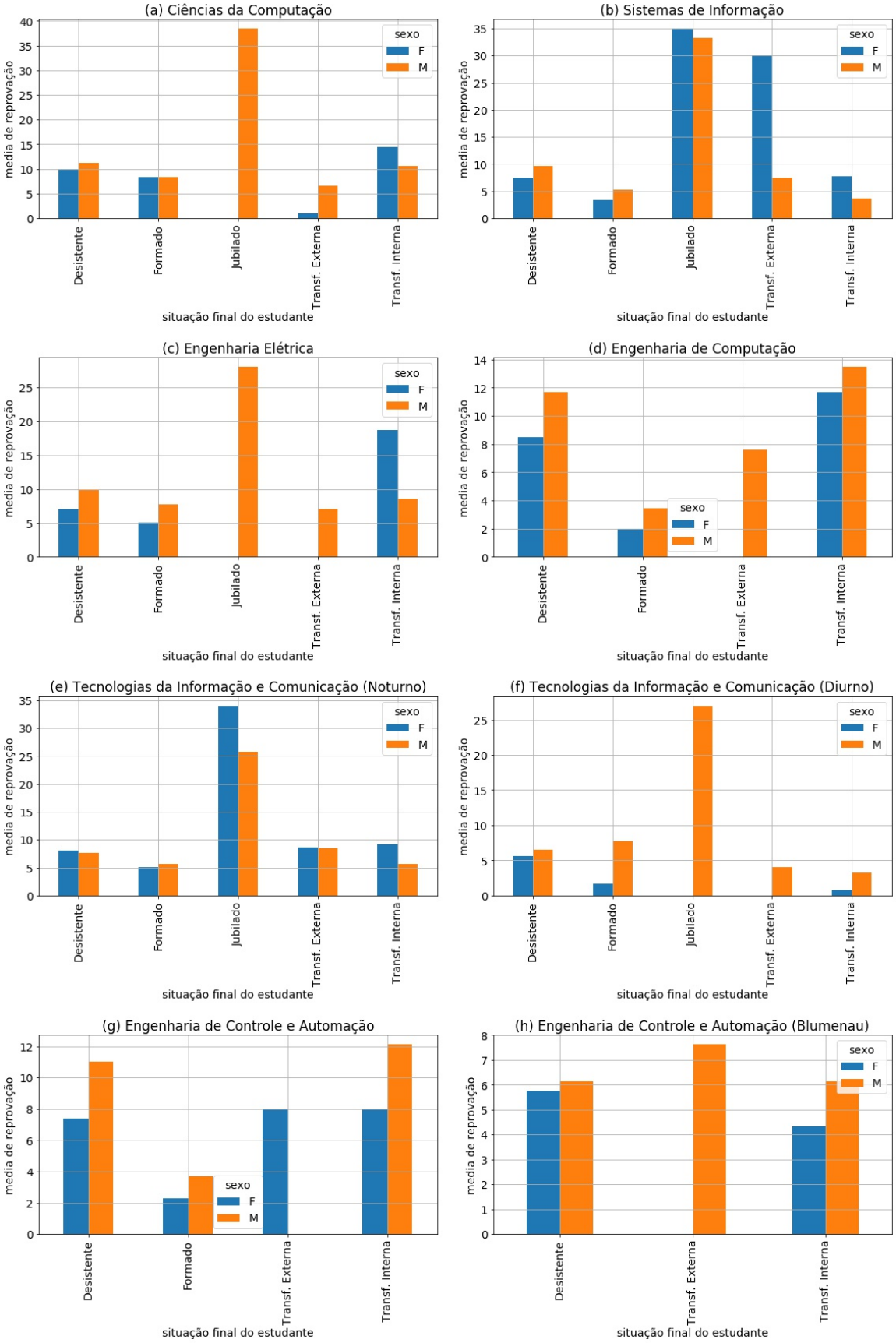
A segunda maior média de reprovação masculina é dos alunos que abandonaram o curso, enquanto a média feminina é das estudantes transferidas para outras instituições de ensino. A situações com menores médias gerais de reprovações são dos alunos formados.

Os Gráficos 15 apresentam as médias de reprovação pela situação final do aluno na universidade de cada curso. Nenhum dos alunos dos cursos de ENC, ECA-FLN e ECA-BLU, durante o período analisado, foi jubilado. Com exceção dos cursos citados, os demais apresentam a maior média de reprovação em alunos jubilados, sendo que, nos cursos de CCO e EEL, nenhuma aluna foi jubilada no período analisado.

Sem levar em consideração os números do alunos jubilados, nos cursos de CCO e EEL, as mulheres que requisitaram transferências para outros cursos da UFSC e os homens que desistiram do curso, foram as segundas maiores médias de reprovação. No curso de SIN as estudantes que fizeram transferência externa e os estudantes que desistiram do curso, obtiveram as maiores médias de reprovação. No curso de ENC ambos os sexo que requisitaram transferência interna atingiram as maiores médias de reprovação. No curso de TIC-NOT, as maiores médias de reprovação são das mulheres que requisitaram transferência interna e dos homens que pediram transferência externa. Na modalidade diurna os formados e as desistentes atingiram as maiores médias de reprovação. As alunas do curso de ECA-FLN em transferência externa e interna e os estudantes do sexo masculino que requisitaram transferência interna, alcançaram as maiores médias de reprovação. No curso de ECA-BLU as médias de reprovação superiores foram das alunas desistentes e dos alunos em transferência interna.

Vale ressaltar que as médias de reprovação de formados são similares as dos alunos que trocaram ou abandonaram o curso, mostrando, assim, que provavelmente o número de reprovações não interfere na escolha do aluno de conclusão ou não do curso.

Gráfico 15 – Média de disciplinas reprovadas dos alunos por situação no último semestre analisado por curso.



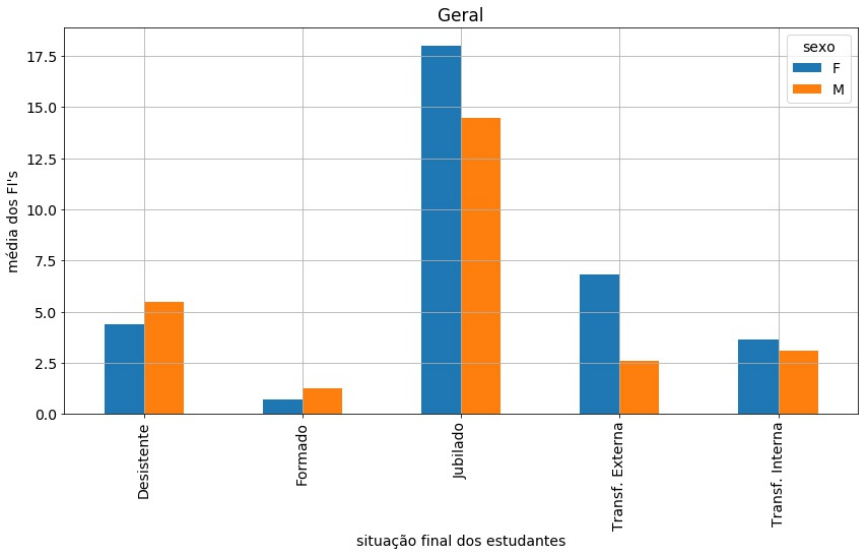
Fonte: elaborado pela autora.

3.4.1.8 Frequência Insuficiente

A seguinte análise possibilita verificar se a média de FI influência na situação final do aluno na universidade. O pré-processamento e a transformação dos dados realizados na análise do número de reprovações é idêntico ao da análise desta análise. A única diferença é que em vez de analisar a quantidade de reprovações, aqui é utilizado diretamente o número de FI.

O Gráfico 16 apresenta as médias de FI's dos alunos, unindo todos os cursos analisados. Como no número de reprovações, os alunos jubilados de ambos os sexos tem os maiores números de FI e os menores números são dos alunos formados. Nos cursos CCO, EEL e TIC-NOT, nenhuma aluna foi jubilada. A segunda maior média feminina são das transferências externas, e masculina, dos alunos desistentes.

Gráfico 16 – Média da soma de FI dos alunos por situação no último semestre, de todos os cursos analisados.



Fonte: elaborado pela autora.

Na análise por curso, Gráfico 17, tem-se os as médias de FI pelas situações finais dos alunos. Não houveram alunos jubilados nos cursos de ENC, ECA-FLN e ECA-BLU, durante o período analisado, porém nos demais cursos a média de FI dos estudantes jubilados é expressivamente superior as demais situações.

A média de FI dos formandos só é inferior as outras situações no curso de TIC-DIU, e também somente os alunos do sexo masculino, e no curso de ECA-BLU devido a ausência de formandos.

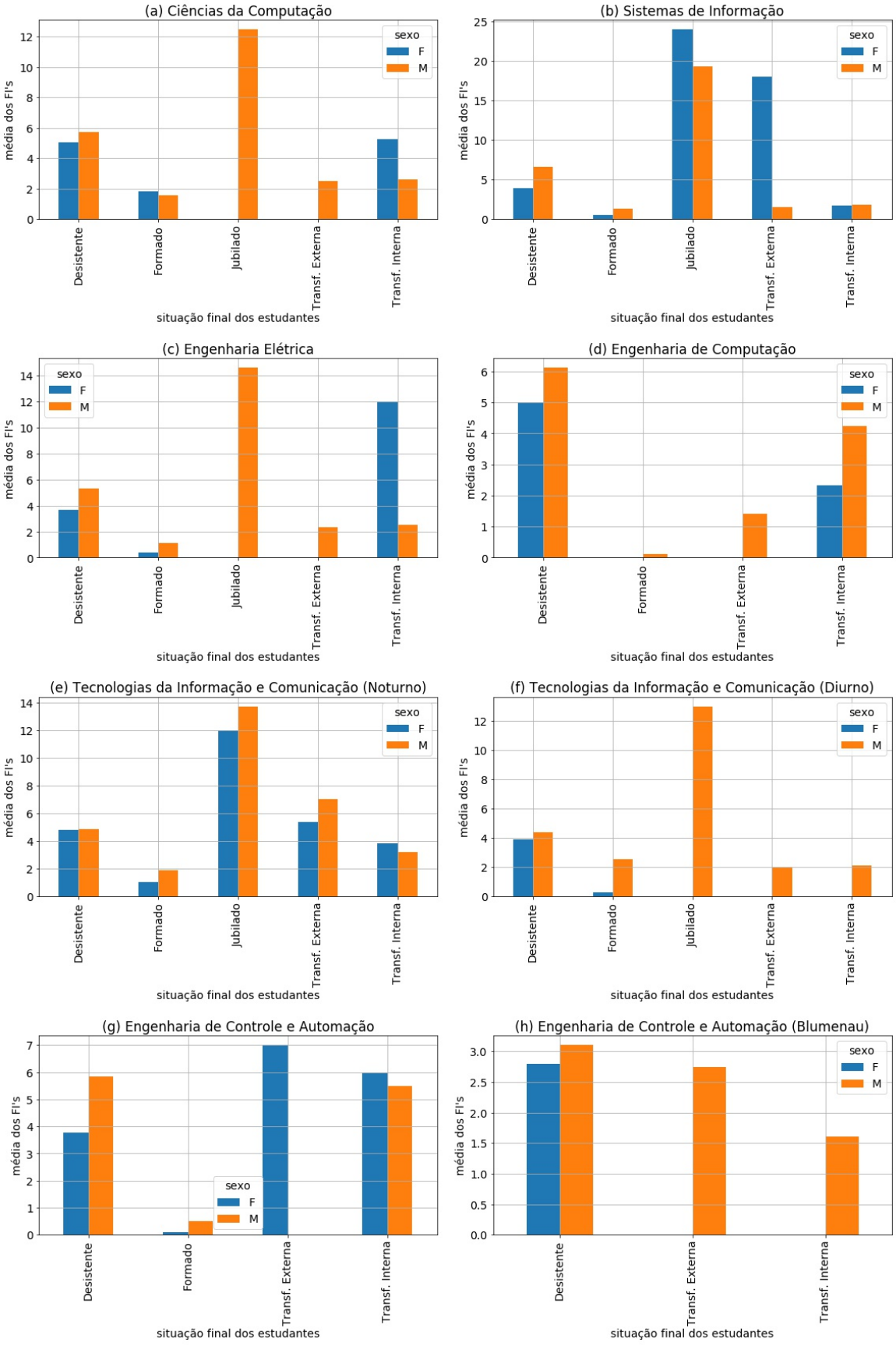
Sem levar em consideração os valores atingidos pelos alunos jubilados, é possível analisar como o número de FI influência nas demais situações finais dos alunos. As mulheres dos cursos de CCO e EEL que optaram por transferência dentro da própria universidade, atingiram maiores médias de FI, enquanto os homens desistentes alcançaram médias superiores. Os cursos de SIN e ECA-FLN, as estudantes que fizeram transferência para outra instituição de ensino obtiveram maiores médias de FI, enquanto os alunos desistentes, alcançaram as maiores médias. Nos cursos ENC, TIC-DIU e ECA-BLU, os discentes desistentes de ambos os sexos obtiveram as maiores médias de FI. No curso

de TIC-NOT todos os estudantes transferidos para outras universidades obtiveram as segundas maiores médias de FI.

Analisando as números gerais de FI os homens com maior média são do curso TIC-NOT, 6 disciplinas, e as mulheres são de SIN, 9 disciplinas em média. Os menores números femininos são do curso de TIC-DIU, 1 disciplinas em média, e masculinas do curso de ECA-BLU, 2 disciplinas em média. As menores médias podem estar ligadas ao curto período de ambos os cursos.

Com essa análise pode-se verificar que os estudantes que são reprovados por falta, tendem a trocar o curso ou são jubilados, enquanto os alunos não faltantes tendem a se formar.

Gráfico 17 – Média da soma de FI dos alunos por situação no último semestre, por curso.



Fonte: elaborado pela autora.

3.4.2 Mineração de dados

Nesta seção serão realizadas análises com a técnica de árvores de decisão. Devido a pequena quantidade de dados, esta técnica de classificação se mostrou a mais adequada para a predição de informações. Foram realizadas predições com outros tipos de algoritmos de mineração, como KNN e *K-means*, porém não foi possível chegar a nenhuma conclusão, devido a baixa acurácia das técnicas.

3.4.2.1 Formado vs Desistente

As árvores de decisão são algorítimos de classificação, ou seja, seu aprendizado é supervisionado. A árvore funciona como um fluxograma, sendo os nós os testes realizados em um determinado atributo, as arestas são as decisões tomadas e as folhas (base da árvore) são a classificação feita pela mesma (ZUEGE, 2018). Cada dado é testado na árvore, o que torna seu algorítimo de aprendizado lento porém eficaz. O índice *Gini* encontrado na árvore mede a impureza da classificação, ou seja, quando o índice é igual a 0 significa que a classificação obteve êxito (HAN; PEI; KAMBER, 2012).

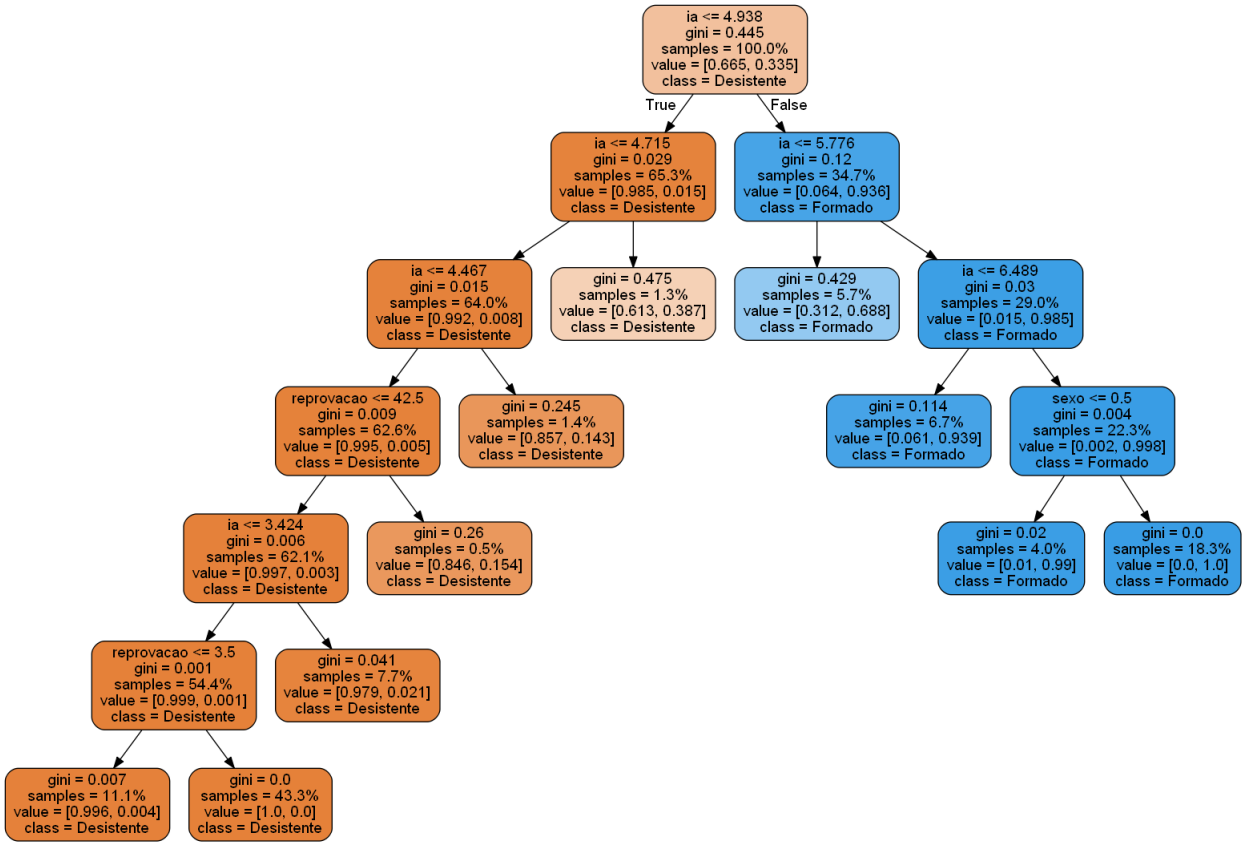
Alguns pré-processamento são realizados para a utilização da biblioteca *scikit-learn*, como a transformação de texto (*string*) para uma categoria (número inteiro), através do método *LabelEncoder* da classe *preprocessing*. Para a criação da árvore de decisão se utilizou alguns módulos da biblioteca *scikit-learn*. Do módulo *tree* utilizou-se as classes *DecisionTreeClassifier* para criação da árvore e *export_graphviz* para exportar a imagem da mesma. Para divisão do conjunto de dados entre treino e teste utilizou-se a classe *train_test_split* do *model_selection*, selecionando 80% para o conjunto de treino e 20% para o de teste.

O Gráfico 18 apresenta a árvore de decisão que recebe como entrada o sexo do estudante, o IA, o número de reprovações, o curso e a forma de ingresso. Na saída se espera a situação do estudante, entre formado e desistente. Esta árvore apresenta 97% de acurácia. As arestas esquerdas estão relacionadas a afirmação da questão levantada pelo nó, enquanto as mais a direita negam a afirmação.

Segundo esta predição um aluno tem 34,7% de probabilidade de se formar e 65,3% de desistir do curso. O fato de a árvore não estar apresentado resultados por curso, pode significar que existe um padrão de desistência e formatura em todos os cursos. Segundo as análises estatísticas, dos alunos ingressos, em todos os cursos, desde 2008/2, 66,9% desistiram do curso enquanto 33,11% conseguiram se formar, esses dados mostram que a árvore reflete a realidade. De acordo com a árvore, o IA tem muita influência na situação final do aluno, o que faz sentido, já que o índice acadêmico é retirado a partir das notas das disciplinas cursadas pelos alunos. Quando a média dos IA's semestrais do aluno é abaixo de 4,3 o aluno é reprovado. Com uma maior quantidade de dados, talvez fosse possível verificar outras ocorrências.

Removendo o IA da entrada, Gráfico 19, se observa novas decisões tomadas pelo algoritmo, porém a precisão cai para 72%, sendo a máxima atingida 73%, com uma árvore extremamente complexa. Observa-se que o número de reprovações é importante para a decisão da situação do aluno, sendo que alunos que possuem mais de 4 reprovações tem uma probabilidade 68,8% de desistir do curso e para alunos ingressantes por retorno, SISU ou transferência externa existe uma probabilidade de 3,6%. Para alunos de ECA-BLU existe a chance de 0,8% de desistência se obterem até 3 reprovações e ingressarem

Gráfico 18 – Árvore de decisão para análise da influência dos atributos na formação ou desistência do curso.

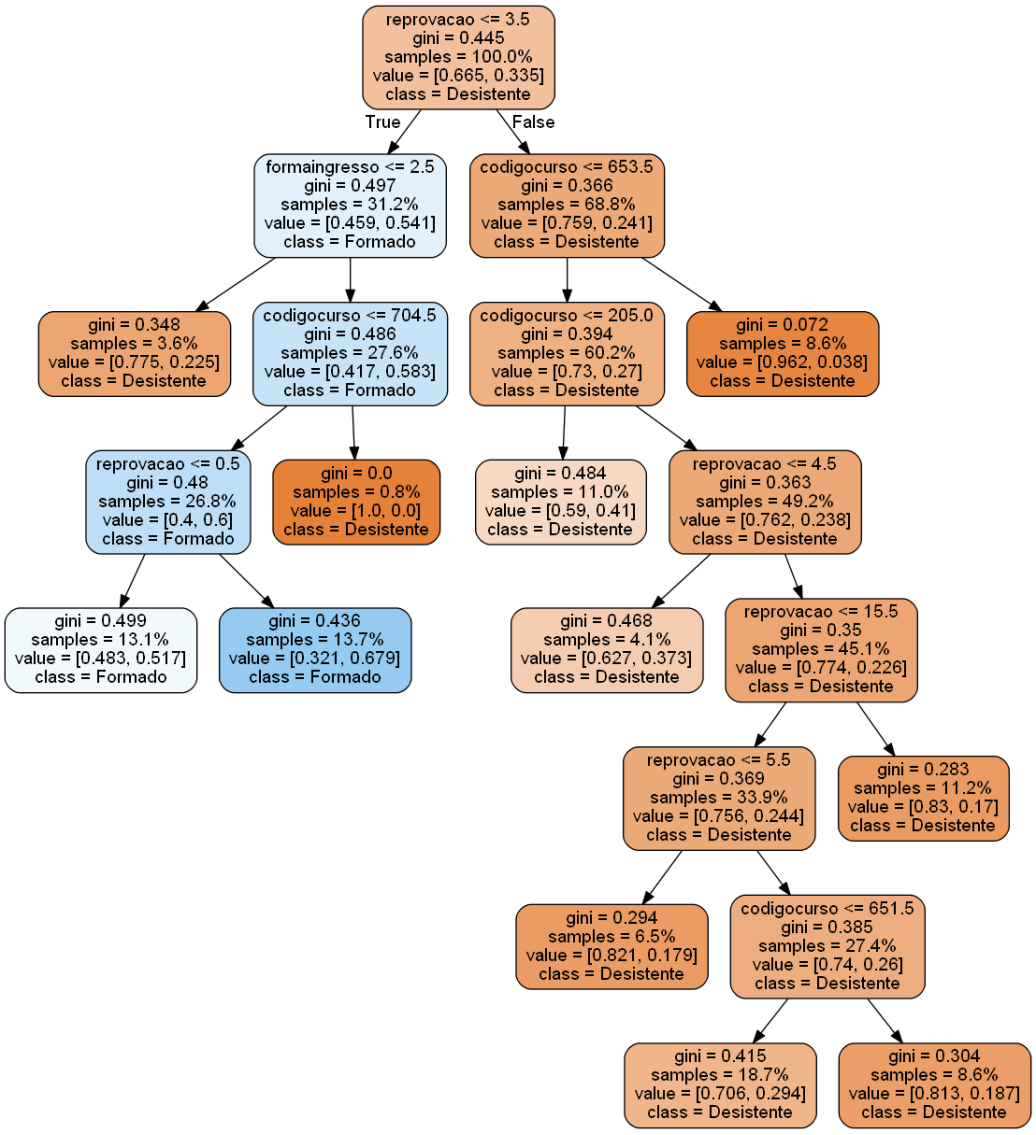


Fonte: elaborado pela autora.

por transferência interna e vestibular. A chance de formatura, segundo essa árvore é de 26,8%.

Como essa árvore não tem uma precisão alta, seus valores não são um total reflexo da realidade. O baixo valor de acurácia pode ser devido ao tamanho do conjunto de dados não ser suficiente para que a árvore consiga generalizar tais resultados

Gráfico 19 – Árvore de decisão para análise da influência dos atributos na formação ou desistência do curso.



Fonte: elaborado pela autora.

4 ANÁLISE DOS RESULTADOS

A pequena quantidade de mulheres pode ser verificada em toda as análises feitas. Os cursos apresentam um leve aumento na quantidade de alunas, porém essa mudança não é considerável. A média do ingresso de mulheres é de 14,44%, o que não mostra alteração na pesquisa de Oliveira, Moro e Prates (2014), que em 2012, verificou a presença de 15% de mulheres. Dos alunos que ingressaram no segundo semestre de 2008, 36% abandonaram o curso e 30% se formaram, apresentando melhores resultado que a pesquisa de Oliveira, Moro e Prates (2014). As análises mostraram que a evasão do curso acontece nas fases iniciais, para ambos os sexos. No período analisado as mulheres consistiram em 14% dos formandos, dados abaixo dos encontrados por Maia (2016), que verificou 17%, de 2000 a 2013. As estudantes possuem IA maior que os homens, demostrando assim que não há motivos para mulheres não serem vistas como incapacitadas para as áreas de tecnologia.

As análises por categoria mostraram que os estudantes cotistas, tanto por raça quanto por renda, apresentam IA inferior aos demais. Porém, a diferença entre não optantes e alunos de baixa renda não apresentou grande discrepância no IA, sendo a diferença ficado nas casas decimais.

O número de reprovações não apresentou grande diferença entre formados e desistentes, porém os alunos que reprovam por frequência insuficiente tem maior chance de não concluir do que os alunos que reprovam por nota. Enquanto os homens desistem mais do curso, as mulheres tendem a pedir transferência.

A análise com algoritmos de mineração de dados mostrou que essa técnica depende de muitos fatores para trazer um bom desempenho. Como a quantidade de dados utilizada neste trabalho é pequena, algoritmos como os de regressão e agrupamento não conseguiram generalizar respostas. Sendo assim foram utilizadas árvores de decisão para buscar informações desconhecidas. Foi possível verificar com uma precisão de 97% que um aluno tem 34,7% de chance de se formar e 65,3% de desistir do curso, valores estes que condizem com os valores encontrados na análise estatística.

5 CONCLUSÃO

Nesta seção são apresentadas as considerações finais e possíveis trabalhos futuros.

5.1 CONSIDERAÇÕES FINAIS

A pesquisa mostrou que a participação das alunas na universidade, mesmo pequena, apresenta um desempenho superior aos alunos homens. Sendo assim, no geral as mulheres que entram e permanecem nos cursos tem rendimento melhor do que os colegas dos sexo masculino, o que confirma a primeira hipótese do trabalho. É conhecido no meio acadêmico que muitas estudantes e pesquisadoras sofrem da síndrome do impostor, onde elas acreditam que seu desempenho é inferior ao que realmente é, ou que em algum momento vão ser descobertas como incapazes (MARTIMIANO et al., 2018). As análises comprovam que na realidade as estudantes não só tem um bom desempenho, como superam seus colegas do sexo masculino. Pesquisas utilizando a ciência de dados são fundamentais para representarem e contextualizarem problemas na relação de gênero e adesão aos cursos de tecnologia.

A segunda hipótese levantada diz que o ingresso de mulheres nos cursos de tecnologia e engenharia vem diminuindo ao longo dos anos, porém segundo as análises feitas o ingresso de estudantes do sexo feminino se mantém constante, variando entre semestres com entrada maior e outros com baixo ingresso. O curso de TIC se mostrou o curso com maior ingresso de mulheres.

A pesquisa também conseguiu verificar a hipótese 3 que diz existirem padrões na evasão e retenção feminina nos cursos de graduação analisados, já que retenção feminina é similar ao número da entrada, ou seja, o percentual de formandas segue a mesma linha do ingresso. Isso leva a concluir que se houvessem mais mulheres ingressando nesses cursos, haveriam, também, mais formandas. Além disso, existe um padrão de evasão nos cursos, onde os estudantes de ambos os sexos costumam desistir nas fases iniciais dos cursos, principalmente no primeiro ano, o que corrobora as hipóteses 3 e 4.

Além de confirmar as hipóteses levantadas as análises feitas observaram também outras informações. O desempenho inferior de alunos cotistas pode revelar a baixa qualidade das escolas públicas ou ainda o despreparo da universidade na recepção desses estudantes, já que antes era preparada para receber discentes oriundos de cursinhos e boas escolas.

A descrição das disciplinas cursadas pelos alunos poderia ter proporcionado uma maior precisão na verificação da evasão e retenção dos alunos, já que se teria a certeza da fase em que o aluno conseguiu chegar.

Devido a pequena quantidade de dados, as tentativas de análises preditivas e descritivas de mineração de dados não obtiveram bons resultados. A única tarefa que conseguiu um resultado aceitável foi a classificação através de árvores de decisão, ainda assim, a quantidade de dados não é significante para que conclusões mais profundas possam ser feitas. Porém a ciência de dados abrange todas as técnicas de na busca do conhecimento e a análise através da estatística foi a que melhor trouxe resultados para o presente trabalho.

Ademais, vê-se necessário a continuidade das atividade de apoio pedagógico da universidade, perante a dificuldade de alunos nas fases iniciais, podendo essa ser um dos motivadores para a desistência dos discentes. A sociedade em geral ainda tem muito o que caminhar na busca da igualdade de gêneros, porém atividades e projetos voltados

ao incentivo do ingresso de mulheres nas áreas de engenharia e tecnologia, como aulas para meninas do ensino médio ou divulgação desses cursos para a comunidade, podem ser realizadas pela universidade. Existem ações isoladas como o projeto meninas digitais, Frigo et al. (2013), no campus Araranguá, porém para serem efetivas, seria necessária uma ampliação das atividades. A diversidade de estudantes na universidade possibilita o crescimento do aluno como cidadão, melhorando o respeito as diversas formas de pensamentos, atitudes e culturas (VIEIRA; DELL'AGLI; CAETANO, 2019).

5.2 TRABALHOS FUTUROS

Os trabalhos futuros estão relacionados a análise de outros dados da UFSC ou ainda dados de outras universidades, que podem apontar novos conhecimentos sobre a participação, não só feminina, mas também de alunos cotistas. A verificação de padrões em outros cursos também pode fornecer um comparativo com os cursos onde há maioria masculina. Essas pesquisas são importantes para que as formas existentes de discriminação sejam verificadas e para que soluções possam ser levantadas para se criar uma maior diversidade em cursos de graduação.

REFERÊNCIAS

AMARAL, F. *Introdução à Ciência de Dados: mineração de dados e big data*. [S.l.]: Alta Books Editora, 2016.

AQUINO, E. L. d. C. Da participação ao ativismos: As tecnologias da informação e comunicação aliadas ao feminismo. Araranguá-SC, 2015.

AZEVÊDO, E. S. et al. A mulher cientista no brasil. dados atuais sobre sua presença e contribuição. *Ciência e cultura*, v. 41, n. 3, p. 275–283, 1989.

BILTON, N. *As mulheres que a tecnologia esqueceu*. 2014. <http://m.folha.uol.com.br/tec/2014/10/1539110-as-mulheresque-a-tecnologia-esqueceu.shtml?mobile>.

CASTRO, B. *Afogados em contratos: o impacto da flexibilização do trabalho na trajetória dos profissionais em TI. 2013*. Tese (Doutorado) — Tese (Doutorado em Ciências Sociais)-Instituto de Filosofia e Ciências . . . , 2013.

CESÁRIO, G. et al. Por mais mulheres na computação: análise dos trabalhos publicados no x women in information technology. In: SBC. *11º Women in Information Technology (WIT 2017)*. [S.l.], 2017. v. 11, n. 1/2017.

COUTO, G. C.; DANTAS, M. A. d. N. A. Utilizando mineração de dados para análise de gênero nos cursos de computação na unb. 2014.

FILHO, M. A. Por uma computação mais democrática (e feminina). *Jornal da Unicamp, edição*, v. 298, p. 22, 2005.

FREITAS, A. A. *Data mining and knowledge discovery with evolutionary algorithms*. [S.l.]: Springer Science & Business Media, 2002.

FRIGO, L. B. et al. Tecnologias computacionais como práticas motivacionais no ensino médio. In: *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*. [S.l.: s.n.], 2013. v. 2, n. 1.

GLOVER, J.; GUERRIER, Y. Women in hybrid roles in it employment: A return to'nimble fingers'? *Journal of technology management & innovation*, Universidad Alberto Hurtado. Facultad de Economía y Negocios, v. 5, n. 1, p. 85–94, 2010.

HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2012.

KELAN, E. *Performing gender at work*. [S.l.]: Springer, 2009. 264 p.

LESLIE, L. L.; MCCLURE, G. T.; OAXACA, R. L. Women and minorities in science and engineering: A life sequence analysis. *The journal of higher education*, Taylor & Francis, v. 69, n. 3, p. 239–276, 1998.

LIGHT, J. S. When computers were women. *Technology and culture*, JSTOR, v. 40, n. 3, p. 455–483, 1999.

LUBAR, S. Men/women/production/consumption. *His and hers: Gender, consumption, and technology*, University Press of Virginia Charlottesville, p. 7–37, 1998.

MAIA, M. M. Limites de gênero e presença feminina nos cursos superiores brasileiros do campo da computação. *cadernos pagu*, n. 46, p. 223–244, 2016.

MARTIMIANO, L. A. et al. Um estrato do perfil das profissionais de tic na cidade de maringá-pr. In: SBC. *12º Women in Information Technology (WIT 2018)*. [S.l.], 2018. v. 12, n. 1/2018.

MCKINNEY, W. et al. Data structures for statistical computing in python. In: AUSTIN, TX. *Proceedings of the 9th Python in Science Conference*. [S.l.], 2010. v. 445, p. 51–56.

NATANSOHN, L. G. Internet em código feminino: Teorias e práticas. La Crujía, 2013.

OLINTO, G. A inclusão das mulheres nas carreiras de ciência e tecnologia no brasil. *Inclusão Social*, v. 5, n. 1, 2011.

OLIPHANT, T. E. Python for scientific computing. *Computing in Science Engineering*, v. 9, n. 3, p. 10–20, May 2007. ISSN 1521-9615.

OLIVEIRA, A. C.; MORO, M. M.; PRATES, R. O. Perfil feminino em computação: Análise inicial. In: *XXXIV Congresso da Sociedade Brasileira da Computação–CSBC*. [S.l.: s.n.], 2014.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, v. 12, n. Oct, p. 2825–2830, 2011.

PNAD, I. Pesquisa nacional por amostra de domicílios. *Rio de Janeiro: IBGE*, 2009.

PRONI, T. T. da R. W.; PRONI, M. W. Discriminação de gênero em grandes empresas no brasil. *Estudos Feministas*, JSTOR, v. 26, n. 1, p. 1–21, 2018.

PYTHON. *Python.org*. 2019. Acessado em 14/06/2019. <https://www.python.org/>.

RAPKIEWICZ, C. E. Informática: domínio masculino? *cadernos pagu*, n. 10, p. 169–200, 1998.

RIBEIRO, K. *Meninas Digitais*. 2019. Acessado em 06/05/2019. <http://meninas.sbc.org.br/index.php/sobre/>.

SABIR, S. et al. A preconception gender assessment using data mining techniques based on implementation of natural laws & favoring factors. In: ACM. *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*. [S.l.], 2017. p. 5.

SANTOS, C. B. et al. Robotics and programming: Attracting girls to technology. In: *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. [S.l.: s.n.], 2016. p. 2052–2056.

SCHIEBINGER, L. O feminismo mudou a ciência. *Bauru: Edusc*, p. 32, 2001.

SCHWARTZ, J. et al. Mulheres na informática: quais foram as pioneiras. *cadernos pagu*, SciELO Brasil, v. 27, n. 1, p. 255–278, 2006.

SOARES, T. A. Mulheres em ciência e tecnologia: ascensão limitada. *Química Nova*, v. 24, n. 2, p. 281–285, 2001.

TAN, P.-N. et al. *Introduction to data mining*. 2. ed. [S.l.]: Pearson, 2018. An optional note. ISBN 9780133128901 and 0133128903.

THOMAS, J.; PRINCY, R. T. Human heart disease prediction system using data mining techniques. In: IEEE. *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on*. [S.l.], 2016. p. 1–5.

VASILESCU, B.; SEREBRENIK, A.; FILKOV, V. A data set for social diversity studies of github teams. In: IEEE PRESS. *Proceedings of the 12th Working Conference on Mining Software Repositories*. [S.l.], 2015. p. 514–517.

VIEIRA, K. E.; DELL'AGLI, B. A. V.; CAETANO, L. M. Desempenho acadêmico de alunos cotistas antes da lei de cotas: Revisão. *Revista da Universidade Vale do Rio Verde*, v. 17, n. 1, 2019.

WENNERAS, C.; WOLD, A. Nepotism and sexism in peer-review. *Women, sience and technology: A reader in feminist science studies*, p. 46–52, 2001.

ZUEGE, T. J. *Aplicação de técnicas de mineração de dados para detecção de perdas comerciais na distribuição de energia elétrica*. Dissertação (B.S. thesis), 2018.