

# A evolução do perfil de estudantes de Engenharia com bom desempenho no ENADE

Eric Gondran<sup>1</sup>, Giancarlo Lucca<sup>12</sup>, Eduardo N. Borges<sup>1</sup>, Helida Santos<sup>1</sup>

<sup>1</sup> PPGComp – Universidade Federal do Rio Grande (FURG) – Rio Grande, RS – Brazil

{ericgondran, eduardoborges, helida}@furg.br

<sup>2</sup>PGEEC – Universidade Católica de Pelotas (UCPEL) – Pelotas, RS – Brazil

giancarlo.lucca@ucpel.edu.br

**Abstract.** *The National Student Performance Examination (ENADE) is annually applied to assess the quality of the Brazilian educational system. This study uses five different feature selection techniques in the ENADE databases from 2005 to 2017, aiming to analyze how personal and socioeconomic characteristics contribute to explaining the performance of graduates in Engineering. It can be concluded, among other results, that 14 attributes were important in the considered period, highlighting the age, working status, grants, financing, and parents' schooling. The main contribution of this work is to demonstrate how the profile of good students has varied over the years.*

**Resumo.** *Para avaliar a qualidade do sistema educacional brasileiro, o Exame Nacional de Desempenho dos Estudantes (ENADE) é realizado anualmente. Este estudo aplica cinco diferentes técnicas de seleção de características nas bases de dados do ENADE de 2005 a 2017, com objetivo de analisar como características pessoais e socioeconômicas contribuem na explicação do desempenho dos concluintes de graduação em Engenharia. Pode-se concluir, dentre outros resultados, que 14 atributos foram considerados importantes no período considerado, destacando-se: idade, situação de trabalho, bolsas, financiamentos e escolarização dos pais. A principal contribuição deste trabalho é demonstrar como o perfil dos bons estudantes têm variado ao longo do anos.*

## 1. Introdução

Nas últimas décadas, a taxa de matrícula mundial no ensino superior aumentou em 10% a cada dez anos [Smolentseva 2018]. No Brasil, o Exame Nacional de Desempenho dos Estudantes (ENADE), aplicado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), avalia os concluintes dos cursos de graduação. Na área das Engenharias é essencial que este tipo de avaliação seja aplicada, pois a carência de profissionais mostrou-se evidente [Fernandes 2013].

A literatura científica apresenta uma série de trabalhos relacionados à análise de dados do ENADE. Em [Gondran. et al. 2022], aplicaram-se diferentes técnicas de seleção de características em dados de 2018 com o objetivo diferenciar perfis de estudantes de cursos na modalidade presencial e EaD. Os resultados mostraram que, em ambas as modalidades, a escolarização do pai influenciou diretamente no desempenho, assim como a renda. Técnicas de Mineração de Dados também foram usadas para análise dos dados do

ENADE em outros trabalhos [Faria 2017, Corso and Resende 2018, Araújo et al. 2019, Barbosa et al. 2019], onde foram utilizados, respectivamente, os seguintes algoritmos e anos do exame: (regressão múltipla, 2012), (RIPPER, 2014), (CART, 2017) e (Regressão Logística, 2017). Dentre os resultados destes trabalhos, destacaram-se como atributos importantes a renda familiar, a idade e a escolarização dos pais. Em [Vieira et al. 2022], utilizando o algoritmo de agrupamento k-means com foco nos concluintes dos cursos da área de computação, foi proposto identificar padrões de desempenho dos estudantes nos exames realizados entre 2008 e 2017. Como resultado, observou-se questões relevantes como: estudantes oriundos de escola particular apresentaram melhor desempenho; e quanto maior a escolarização do pai e da mãe, melhor foi o resultado. A categoria das Instituições de Ensino Superior (IES) também influenciou o desempenho, com destaque para as IES públicas. Por fim, os resultados mostraram uma evolução dos grupos ao longo das edições. Estudos ao longo dos anos de ENADE apresentaram como atributos importantes a categoria da IES, a cor da pele e a escolarização dos pais [Rosa et al. 2021, Fernando Choji et al. 2021, de Almeida et al. 2021, Bertolin et al. 2022].

A partir do levantamento bibliográfico realizado, não foi possível encontrar trabalhos que utilizassem técnicas de seleção de atributos nos microdados de várias edições do ENADE, de forma a entender a evolução do perfil de estudantes de Engenharia. Considerando a expansão do ensino superior, a demanda por engenheiros qualificados, e a revisão bibliográfica realizada, este trabalho tem o objetivo de encontrar características pessoais e socioeconômicas significativas na explicação do desempenho no ENADE dos concluintes de graduação em Engenharia. Para atingir o objetivo proposto, o estudo empregou uma metodologia bastante consolidada e explorada na literatura, denominada Descoberta de Conhecimento em Base de Dados (DCBD) [Tan et al. 2016]. Um dos passos da DCBD é a Mineração de Dados (MD) [Tan et al. 2016], que tem como objetivo analisar registros, extraindo conteúdo útil por meio de algoritmos que criam um modelo com base nos dados. A aplicação das técnicas de MD no campo da educação tem apresentado resultados promissores, dando origem a uma nova área de investigação científica denominada Mineração de Dados Educacionais (MDE) [Baker et al. 2011].

O restante deste trabalho está organizado da seguinte forma. Na Seção 2 é apresentada a fundamentação teórica sobre seleção de atributos. Na Seção 3, a metodologia é especificada. Os resultados são apresentados e discutidos na Seção 4. Por fim, a Seção 5 apresenta as considerações finais e uma breve discussão acerca do trabalho futuro.

## **2. Seleção de Atributos**

A Seleção Atributos (SA) é o processo de identificar e remover tanto atributos irrelevantes, quanto informações redundantes [Miao and Niu 2016]. Esta técnica reduz a dimensionalidade dos dados e melhora o desempenho dos classificadores, pois elimina atributos que não agregam valor à classificação ou que deterioram os resultados. Contribui para um melhor entendimento e análise dos resultados obtidos, além de permitir que os algoritmos de aprendizagem operem de forma mais rápida e eficaz. Assim, o principal objetivo da SA é identificar o conjunto de atributos que melhor representa, para um determinado contexto, a informação útil contida nos dados [Tasca 2015]. No contexto deste trabalho, apresentam-se a seguir alguns métodos/algoritmos relevantes.

## 2.1. Ganho de Informação (GI)

O ganho de informação [Hall and Smith 1998] é definido como a quantidade de informação obtida sobre uma variável aleatória ou sinal da observação de outra variável aleatória. Ela mede a significância do atributo em relação à classe alvo, ou seja, mede a redução da incerteza (entropia) em função da divisão. Como desvantagem, tende a preferir divisões que resultam em grande número de partições, cada uma pequena, mas pura. O ganho de informação (GI) é calculado por:  $GI(A) = E(C) - \sum_{a \in A} \frac{n_a}{n} E(a)$ , onde:  $E$  é a entropia,  $C$  é a classe,  $A$  é o atributo a ser avaliado,  $n_a$  é o número de instâncias da categoria do atributo,  $n$  é o número total de instâncias e  $a$  é o valor do atributo.

## 2.2. Taxa de Ganho (TG)

A taxa de ganho [Karegowda et al. 2010] foi desenvolvida para solucionar o problema do ganho de informação. Ela é a razão do ganho de informação pela entropia do atributo. Ajusta o Ganho de Informação pela entropia do particionamento, fazendo com que o particionamento de alta entropia (grande número de pequenas partições) seja penalizado. A taxa de ganho é dada por:  $TG(A) = \frac{GI(A)}{E(A)}$ .

## 2.3. Incerteza Simétrica (IS)

A incerteza simétrica [Yu and Liu 2003] é uma medida de correlação não linear, desenvolvida com o mesmo propósito da taxa de ganho, isto é, uma tentativa de normalizar o ganho de informação de um atributo em relação à classe. É calculada de acordo com a seguinte equação:  $IS(A) = 2 \cdot \frac{GI(A)}{E(A)+E(C)}$ .

## 2.4. Correlação de Pearson (Corr)

A Correlação de Pearson [Hall 1998], também conhecida como coeficiente linear, mede o grau da correlação entre duas variáveis de escala métrica. De forma simples, é um grau de relação entre duas variáveis quantitativas e exprime o grau de correlação através de valores situados entre -1 (correlação negativa ou inversa) e 1 (relação linear positiva). Um coeficiente de correlação próximo a zero indica que não há relação entre as duas variáveis. A correlação é dada por:  $P(A) = \frac{Cov(X,Y)}{\sqrt{Var(X) \cdot Var(Y)}}$ , onde  $Cov$  é a covariância entre os dois atributos e  $Var$  é a variância de cada atributo. Para calcular a correlação de variáveis qualitativas, os dados são dicotomizados.

## 2.5. Alívio F (AF)

Ao longo dos anos, uma extensão do Alívio, chamada Alívio F, foi desenvolvida [Kononenko 1994, Kira and Rendell 1992, Urbanowicz et al. 2018]. Ela melhora o algoritmo original ao estimar probabilidades de forma mais confiável e o estende para lidar com conjuntos de dados incompletos e de várias classes, enquanto a complexidade permanece a mesma. Seu cálculo é expresso pela equação:  $W(A) = W(A) - \frac{diff(A, R_i, H)}{m} + \frac{diff(A, R_i, M)}{m}$ , onde  $W(A)$  é um vetor com a pontuação de cada atributo,  $R_i$  é a instância alvo,  $H$  é a instância mais próxima da mesma classe,  $M$  é a instância mais próxima da outra classe,  $m$  é o número de instâncias aleatórias selecionadas para fazer parte do cálculo, e  $diff$  é uma função para calcular a diferença entre os atributos.

### 3. Metodologia

As bases de dados utilizadas neste trabalho são os microdados do ENADE disponibilizadas pelo Inep [Inep 2022]. O exame possui um ciclo avaliativo de três anos, avaliando um conjunto de cursos diferentes a cada ano. Ressalta-se que o foco desta pesquisa são alunos de Engenharia, dada a falta de engenheiros experientes no mercado, aliado ao problema de pouca qualidade dos engenheiros formados. No entanto, evidencia-se a carência de dados confiáveis em relação a estes profissionais [Salerno et al. 2014]. Assim, foram coletados os anos do ciclo avaliativo *Ano II*, sendo eles: 2005, 2008, 2011, 2014 e 2017. No total 37 cursos de engenharias compõem esta seleção.

A cada ano foram selecionados os alunos concluintes dos cursos de Engenharia, na modalidade presencial. Somente os concluintes presenciais foram escolhidos porque, de acordo com [Gondran. et al. 2022], há diferenças significativas nos perfis dos alunos na modalidade EaD e presencial. Após análise de todos os atributos de cada base de dados, o estudo focou nos aspectos pessoais e socioeconômicos. Um total de 22 atributos atendeu a este requisito, sendo eles: Idade, Sexo, Estado Civil, Cor, Escolarização pai e mãe, Onde e com quem mora, Pessoas que moram com você, Renda familiar, Situação financeira, Situação de trabalho, Bolsa/financiamento, Bolsa permanência, Bolsa acadêmica, Ação afirmativa ou inclusão social, Tipo escola ensino médio, Modalidade ensino médio, Família com curso superior, Estudo semanal, Escolha curso, Escolha instituição e Nota geral. Além dos atributos de semântica equivalente ao nome, ressalta-se a situação financeira que diz respeito à origem e ao destino da renda do indivíduo, enquanto a situação de trabalho refere-se a estar trabalhando e ao quantitativo de horas semanais trabalhadas. Os atributos bolsa/financiamento, bolsa permanência e bolsa acadêmica caracterizam os tipos de bolsa ou financiamento. Escola ensino médio categoriza os estudantes em: pública ou privada, no Brasil ou no exterior. Estudo semanal refere-se a quantidade de horas de estudo extraclasse. Os atributos escolha curso e escolha instituição dizem respeito aos motivos pela escolha do curso e da instituição, respectivamente. Família com curso superior indica se alguém na família concluiu uma graduação. Por fim, a nota geral refere-se ao desempenho no ENADE. O número de atributos em cada base de dados variou ligeiramente porque alguns atributos não estavam presentes em todos os anos considerados.

Após a seleção, foi aplicado um filtro para retirar os participantes: i) ausentes, com prova objetiva ou discursiva em branco, e ii) com o resultado desconsiderado pelo Inep. A terceira etapa da limpeza consistiu em verificar dados nulos ou incompletos, esses casos incluem: notas da prova em branco, parte do questionário em branco ou aqueles que não quiseram responder o questionário. Operações realizadas na transformação incluíram: i) criação ou derivação de novos atributos, por exemplo, se a família tinha concluído um curso superior; ii) padronização na nomenclatura dos atributos em todos os períodos; iii) categorização das notas obtidas pelos candidatos em desempenho baixo, médio e alto, mantendo a distribuição original a cada ano. A Tabela 1 sintetiza quantitativamente os dados considerados no estudo.

Finalizado o pré-processamento, os cinco métodos distintos de SA descritos na Seção 2 foram aplicados, em cada um dos cinco anos selecionados (2005, 2008, 2011, 2014 e 2017). Consideraram-se os *top-10* atributos mais frequentes em pelo menos quatro *rankings* gerados pelos algoritmos de SA para cada ano do ENADE. Por fim, os resultados obtidos a cada ano e a variação do perfil ao longo do tempo foram analisados.

**Tabela 1. Conjunto de dados do ENADE utilizado no estudo.**

Ano	Estudantes	Selecionados		Após pré-processamento e transformação	
		Concluintes Engenharia	Atributos	Concluintes Engenharia	Atributos
2005	323.338	20.763	210	8.996	18
2008	461.776	28.145	198	12.281	19
2011	376.180	47.808	115	30.400	20
2014	481.720	91.635	154	66.081	22
2017	537.436	149.300	150	112.285	22

## 4. Resultados

Os resultados dos cinco algoritmos de SA aplicados no ENADE 2005 estão disponíveis na Tabela 2, onde as linhas são os 17 atributos considerados (neste ano) e nas colunas, os diferentes métodos de SA. Os valores de cada célula referem-se à posição no *ranking* e ao resultado da operação obtidos pelo atributo. Os dez melhores valores, para cada método, são sublinhados. Além disso, estão destacados em **negrito** os atributos selecionados nas dez primeiras posições em pelo menos quatro dos algoritmos utilizados. Ao final da tabela, a coluna *Rank* contém a soma das quatro melhores posições obtidas pelos diferentes métodos de SA.

Analisando a Tabela 2, podemos observar que os atributos que mais aparecem entre as duas primeiras posições do ranking são: Idade e Situação de Trabalho. Ainda podemos observar os atributos Estudo Semanal, Família com Curso superior e Bolsa Acadêmica também mantêm-se nas primeiras colocações. Por último, destacamos os atributos Onde e com quem mora, Escolarização da Mãe, Renda Familiar e Escolarização do Pai, fechando a seleção dos *top-10* considerados os mais importantes. Perceba que Estado Civil não foi considerado porque, apesar de bem classificado por três algoritmos, ficou em 11º quando considerado o Ganho de Informação (GI) e 16º para o Alívio F (AL).

De forma análoga, os resultados dos cinco algoritmos de SA aplicados no ENADE 2008 estão disponíveis na Tabela 3. Observando os resultados, percebe-se que os atributos Idade, Situação de Trabalho, Bolsa Acadêmica, Família com Curso Superior e Escolarização Pai são os que mais se destacam. Também aparecem como *top-10*, em quatro dos algoritmos, a Escolarização Mãe e o Estado Civil.

**Tabela 2. Resultados obtidos com diferentes algoritmos de seleção de atributos, considerando o ENADE 2005 para concluintes de Engenharia.**

#	Atributos	Corr	GI	TG	IS	AL	Rank
1	<b>Idade</b>	<u>4</u> 0,02862	<u>1</u> 0,02431	<u>1</u> 0,01252	<u>1</u> 0,01597	17 0,00363	7
2	<b>Situação de Trabalho</b>	<u>2</u> 0,03612	<u>2</u> 0,01558	<u>5</u> 0,00784	<u>2</u> 0,01008	<u>5</u> 0,04007	11
3	<b>Estudo Semanal</b>	<u>3</u> 0,02992	<u>3</u> 0,01448	<u>7</u> 0,00674	<u>6</u> 0,00890	<u>1</u> 0,06886	13
4	<b>Família com Curso Superior</b>	<u>1</u> 0,03821	<u>9</u> 0,01040	<u>2</u> 0,01040	<u>3</u> 0,00989	15 0,01329	15
5	<b>Bolsa Acadêmica</b>	<u>8</u> 0,02307	<u>4</u> 0,01434	<u>6</u> 0,00720	<u>5</u> 0,00926	<u>3</u> 0,04598	18
6	<b>Onde e com quem Mora</b>	15 0,01169	<u>5</u> 0,01228	<u>4</u> 0,00800	<u>4</u> 0,00931	<u>10</u> 0,02682	23
7	<b>Escolarização Mãe</b>	11 0,01597	<u>6</u> 0,01154	<u>8</u> 0,00592	<u>7</u> 0,00756	<u>7</u> 0,03787	28
8	Estado Civil	<u>7</u> 0,02356	11 0,00704	<u>3</u> 0,00862	<u>8</u> 0,00733	16 0,00947	29
9	<b>Renda Familiar</b>	<u>10</u> 0,01744	<u>8</u> 0,01065	<u>10</u> 0,00558	<u>10</u> 0,00707	<u>2</u> 0,05023	30
10	<b>Escolarização Pai</b>	<u>9</u> 0,02201	<u>7</u> 0,01104	<u>9</u> 0,00565	<u>9</u> 0,00723	<u>8</u> 0,03773	33
11	Bolsa/Financiamento	<u>6</u> 0,02476	<u>10</u> 0,00761	11 0,00493	11 0,00575	11 0,02564	38
12	Escola Ensino Médio	<u>5</u> 0,02669	12 0,00650	13 0,00385	12 0,00466	<u>9</u> 0,02776	38
13	Situação Financeira	13 0,01345	13 0,00598	15 0,00281	14 0,00371	<u>6</u> 0,03864	46
14	Modalidade Ensino Médio	12 0,01516	14 0,00466	12 0,00410	13 0,00416	12 0,02089	49
15	Pessoas Moram com Você	16 0,01014	15 0,00352	16 0,00186	16 0,00235	<u>4</u> 0,04547	51
16	Cor	14 0,01305	16 0,00342	14 0,00317	15 0,00313	13 0,01996	56
17	Sexo	17 0,00945	17 0,00112	17 0,00147	17 0,00120	14 0,01800	65

**Tabela 3. Resultados obtidos com diferentes algoritmos de seleção de atributos, considerando o ENADE 2008 para concluintes de Engenharia.**

#	Atributos	Corr	GI	TG	IS	AL	Rank					
1	<b>Idade</b>	3	0,03840	1	0,03271	3	0,01479	1	0,02033	18	0,00533	8
2	<b>Situação de Trabalho</b>	2	0,04780	2	0,02610	4	0,01399	2	0,01818	4	0,05835	10
3	<b>Bolsa Acadêmica</b>	1	0,05400	4	0,02106	8	0,01056	5	0,01404	2	0,06452	12
4	<b>Família com Curso Superior</b>	5	0,03780	8	0,01609	1	0,01609	3	0,01605	15	0,01459	17
5	<b>Escolarização Pai</b>	10	0,02470	3	0,02206	6	0,01121	4	0,01483	7	0,04880	20
6	Estudo Semanal	12	0,02230	7	0,01660	11	0,00777	9	0,01057	1	0,07785	28
7	<b>Escolarização Mãe</b>	8	0,02830	5	0,01914	9	0,00985	7	0,01298	8	0,04733	28
8	Onde e com quem Mora	11	0,02250	6	0,01699	7	0,01066	6	0,01307	12	0,03051	30
9	<b>Ação Afirmativa ou Inclusão Social</b>	4	0,03810	15	0,00771	2	0,01596	10	0,01036	17	0,00862	31
10	<b>Estado Civil</b>	9	0,02600	10	0,01183	5	0,01382	8	0,01271	16	0,01261	32
11	Situação Financeira	17	0,01750	9	0,01345	14	0,00623	12	0,00850	5	0,05744	40
12	Modalidade Ensino Médio	6	0,03160	14	0,00838	10	0,00779	14	0,00806	13	0,02681	43
13	Escola Ensino Médio	14	0,02110	11	0,01139	12	0,00713	11	0,00875	10	0,03422	44
14	Renda Familiar	15	0,02100	13	0,00880	15	0,00467	15	0,00609	3	0,06437	46
15	Bolsa/Financiamento	16	0,01850	12	0,01099	13	0,00657	13	0,00821	9	0,03770	47
16	Sexo	7	0,02830	18	0,00144	17	0,00191	18	0,00164	11	0,03101	53
17	Pessoas Moram com Você	18	0,01640	17	0,00281	18	0,00153	17	0,00198	6	0,05711	58
18	Cor	13	0,02130	16	0,00375	16	0,00343	16	0,00357	14	0,02434	59

Na Tabela 4 são apresentados os resultados do ENADE 2011. O atributo Bolsa/Financiamento ocupa a primeira colocação em todos os algoritmos, mostrando a sua importância neste ano de exame. Idade, Situação de Trabalho, Bolsa Permanência, Estudo Semanal, Família com Curso Superior e Escola Ensino Médio são os atributos que se mantêm no top-10 em quatro das técnicas de seleção de atributos. Nota-se que alguns atributos antes considerados importantes agora são penalizados pela Taxa de Ganho e pelo Alívio F.

A Tabela 5 apresenta os resultados do ENADE 2014. Destacam-se os atributos Bolsa/Financiamento e Bolsa Acadêmica por terem retornado para primeiro ou segundo no ranking da maioria dos algoritmos. Escolha Instituição, Escola Ensino Médio, Idade, Escolarização Pai, Situação de Trabalho e Escolarização Mãe são os atributos que se mantêm no top-10 na maioria dos algoritmos, com exceção do Alívio F.

Por fim, os resultados do ENADE 2017 estão disponíveis na Tabela 6. Novamente, vê-se o destaque para os atributos relacionados às bolsas. Ainda foram selecionados Es-

**Tabela 4. Resultados obtidos com diferentes algoritmos de seleção de atributos, considerando o ENADE 2011 para concluintes de Engenharia.**

#	Atributos	Corr	GI	TG	IS	AL	Rank					
1	<b>Bolsa/Financiamento</b>	1	0,04934	1	0,05766	1	0,02415	1	0,03381	1	0,07220	4
2	<b>Idade</b>	5	0,03063	2	0,02141	3	0,01125	2	0,01463	19	0,00414	12
3	<b>Situação de Trabalho</b>	4	0,03107	3	0,01965	4	0,01114	3	0,01410	9	0,03495	14
4	<b>Bolsa Permanência</b>	2	0,04846	8	0,01117	2	0,01156	4	0,01122	15	0,01557	16
5	<b>Estudo Semanal</b>	9	0,02060	4	0,01276	9	0,00638	7	0,00844	4	0,05699	24
6	<b>Família com Curso Superior</b>	7	0,02955	9	0,01044	5	0,01045	5	0,01032	16	0,01249	26
7	<b>Escolarização Mãe</b>	14	0,01396	5	0,01229	11	0,00543	9	0,00747	7	0,04409	32
8	Onde e com quem Mora	18	0,00849	6	0,01190	8	0,00717	6	0,00887	12	0,02457	32
9	Bolsa Acadêmica	6	0,02963	11	0,00923	13	0,00366	13	0,00520	5	0,04614	35
10	Estado Civil	8	0,02694	13	0,00721	7	0,00846	8	0,00768	18	0,00722	36
11	Escolarização Pai	12	0,01468	7	0,01156	12	0,00507	12	0,00700	6	0,04538	37
12	<b>Escola Ensino Médio</b>	10	0,01744	10	0,00970	10	0,00617	10	0,00747	11	0,02653	40
13	Sexo	3	0,04429	18	0,00257	15	0,00322	16	0,00282	10	0,02985	44
14	Situação Financeira	13	0,01431	12	0,00742	14	0,00352	14	0,00474	8	0,04260	47
15	<b>Ação Afirmativa ou Inclusão Social</b>	16	0,01193	15	0,00631	6	0,00855	11	0,00717	17	0,01051	48
16	Renda Familiar	17	0,01102	14	0,00646	17	0,00243	15	0,00351	3	0,05750	49
17	Pessoas Moram com Você	19	0,00788	16	0,00353	19	0,00135	19	0,00194	2	0,05881	56
18	Cor	11	0,01599	17	0,00259	18	0,00219	18	0,00235	13	0,02238	59
19	Modalidade Ensino Médio	15	0,01291	19	0,00245	16	0,00269	17	0,00253	14	0,01956	62

**Tabela 5. Resultados obtidos com diferentes algoritmos de seleção de atributos, considerando o ENADE 2014 para concluintes de Engenharia.**

#	Atributos	Corr	GI	TG	IS	AL	Rank					
1	<b>Bolsa/Financiamento</b>	2	0,05304	1	0,06162	2	0,02480	1	0,07589	5		
2	<b>Bolsa Acadêmica</b>	1	0,06830	2	0,03637	1	0,02679	2	0,03076	10	0,02973	6
3	<b>Escolha Instituição</b>	7	0,02936	3	0,02458	5	0,00996	4	0,01415	3	0,05364	15
4	<b>Escola Ensino Médio</b>	5	0,03613	6	0,01958	3	0,01297	3	0,01556	12	0,02837	17
5	<b>Idade</b>	4	0,03655	4	0,02353	6	0,00971	5	0,01372	21	0,00322	19
6	<b>Escolarização Pai</b>	9	0,02534	5	0,02054	7	0,00902	6	0,01251	7	0,04747	25
7	<b>Situação de Trabalho</b>	3	0,04121	9	0,01625	8	0,00884	8	0,01142	11	0,02904	28
8	<b>Escolarização Mãe</b>	11	0,02332	7	0,01923	9	0,00846	7	0,01172	8	0,04462	31
9	<b>Estado Civil</b>	6	0,03270	12	0,00999	4	0,01128	9	0,01056	19	0,00712	31
10	<b>Renda Familiar</b>	12	0,02032	8	0,01868	13	0,00715	10	0,01032	2	0,06569	32
11	<b>Estudo Semanal</b>	15	0,01861	11	0,01018	15	0,00505	14	0,00673	6	0,04782	46
12	<b>Família com Curso Superior</b>	10	0,02418	16	0,00651	10	0,00826	12	0,00726	16	0,01850	48
13	<b>Onde e com quem Mora</b>	18	0,01253	10	0,01192	12	0,00718	11	0,00894	15	0,02240	48
14	<b>Escolha Curso</b>	14	0,01970	14	0,00698	19	0,00320	17	0,00438	5	0,05051	50
15	<b>Situação Financeira</b>	16	0,01559	13	0,00923	16	0,00403	16	0,00560	9	0,04127	54
16	<b>Cor</b>	13	0,01973	17	0,00644	14	0,00505	15	0,00564	13	0,02783	55
17	<b>Ação Afirmativa ou Inclusão Social</b>	20	0,00759	15	0,00657	11	0,00723	13	0,00686	17	0,01604	56
18	<b>Pessoas Moram com Você</b>	19	0,00886	18	0,00537	20	0,00207	19	0,00298	4	0,05119	60
19	<b>Sexo</b>	8	0,02805	21	0,00174	21	0,00199	21	0,00185	14	0,02701	64
20	<b>Modalidade Ensino Médio</b>	17	0,01280	19	0,00311	17	0,00379	18	0,00340	18	0,01590	70
21	<b>Bolsa Permanência</b>	21	0,00747	20	0,00195	18	0,00342	20	0,00247	20	0,00505	78

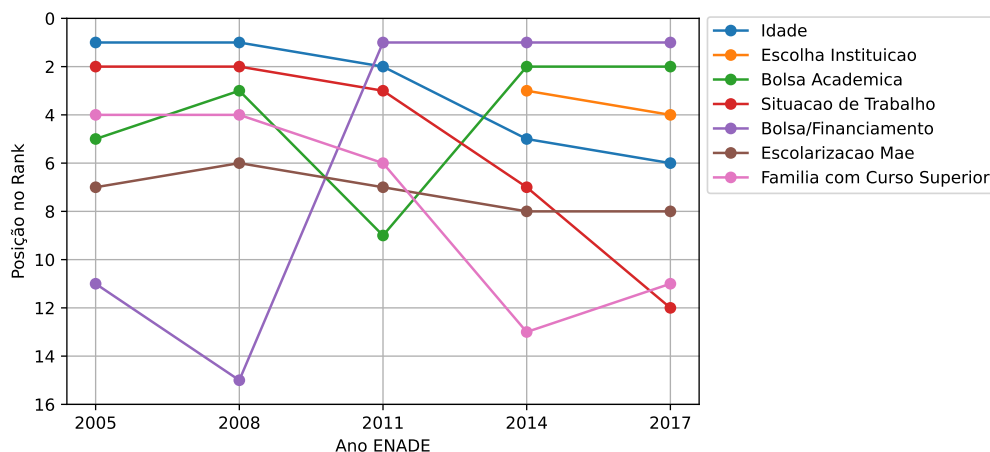
cola Ensino Médio, Escolha Instituição, Renda Familiar, Idade e Escolarização Mãe.

Para analisar a variação do perfil dos estudantes de Engenharia com bom desempenho no ENADE ao longo dos anos considerados neste estudo, foi calculada a média dos rankings para cada atributo e ano de aplicação do exame. Como a visualização de todos os atributos de uma única vez dificultaria muito o entendimento, optou-se por dividi-los em três subconjuntos, apresentando os comportamentos de sete atributos em cada gráfico, agrupados de acordo com a ordem geral de classificação relativa a coluna *Rank*.

A Figura 1 mostra os sete primeiros atributos. É possível perceber que Idade e Situação de Trabalho vêm perdendo importância ao longo dos anos, assim como o atributo Família com Curso Superior, embora esse atributo acene uma ligeira recuperação em 2017. Escolha Instituição é um atributo que surgiu somente em 2014, com um bom

**Tabela 6. Resultados obtidos com diferentes algoritmos de seleção de atributos, considerando o ENADE 2017 para concluintes de Engenharia.**

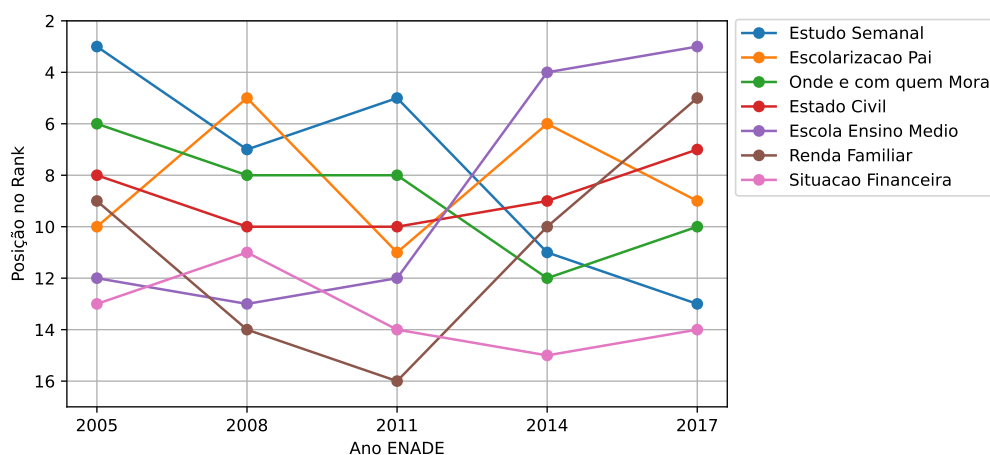
#	Atributos	Corr	GI	TG	IS	AL	Rank					
1	<b>Bolsa/Financiamento</b>	4	0,04244	1	0,06501	2	0,02604	1	0,03696	1	0,07026	5
2	<b>Bolsa Acadêmica</b>	6	0,03290	2	0,03763	1	0,02681	2	0,03103	10	0,03045	11
3	<b>Escola Ensino Médio</b>	3	0,04810	6	0,02039	3	0,01418	3	0,01658	13	0,02518	15
4	<b>Escolha Instituição</b>	9	0,02935	3	0,02627	5	0,01047	4	0,01489	5	0,04450	17
5	<b>Renda Familiar</b>	13	0,02276	5	0,02242	8	0,00857	5	0,01233	2	0,05864	20
6	<b>Idade</b>	1	0,07175	4	0,02347	10	0,00833	6	0,01223	21	0,00280	21
7	<b>Estado Civil</b>	2	0,05161	13	0,01113	4	0,01357	7	0,01209	20	0,00572	26
8	<b>Escolarização Mãe</b>	14	0,02210	7	0,01918	9	0,00848	8	0,01168	8	0,03929	32
9	<b>Escolarização Pai</b>	17	0,01792	8	0,01870	11	0,00819	9	0,01132	7	0,04041	35
10	<b>Onde e com quem Mora</b>	11	0,02892	9	0,01419	7	0,00882	10	0,01079	15	0,01960	37
11	<b>Família com Curso Superior</b>	7	0,03066	14	0,00735	6	0,00898	12	0,00799	17	0,01507	39
12	<b>Situação de Trabalho</b>	8	0,03041	10	0,01409	12	0,00762	11	0,00982	11	0,02789	40
13	<b>Estudo Semanal</b>	12	0,02849	11	0,01178	13	0,00580	13	0,00772	6	0,04365	42
14	<b>Situação Financeira</b>	15	0,02121	12	0,01114	16	0,00485	14	0,00672	9	0,03576	50
15	<b>Sexo</b>	5	0,03398	19	0,00340	18	0,00372	18	0,00351	12	0,02752	53
16	<b>Cor</b>	10	0,02905	15	0,00620	17	0,00412	16	0,00491	14	0,02358	55
17	<b>Escolha Curso</b>	18	0,01166	18	0,00547	20	0,00244	19	0,00335	3	0,04551	58
18	<b>Ação Afirmativa ou Inclusão Social</b>	20	0,00944	16	0,00591	14	0,00536	15	0,00556	16	0,01940	61
19	<b>Pessoas Moram com Você</b>	21	0,00778	17	0,00552	21	0,00214	20	0,00306	4	0,04494	62
20	<b>Bolsa Permanência</b>	16	0,01997	20	0,00285	15	0,00494	17	0,00357	19	0,00585	67
21	<b>Modalidade Ensino Médio</b>	19	0,00958	21	0,00272	19	0,00343	21	0,00300	18	0,01415	77



**Figura 1. Evolução do perfil de estudantes de Engenharia com bom desempenho no ENADE ao longo dos anos (parte 1).**

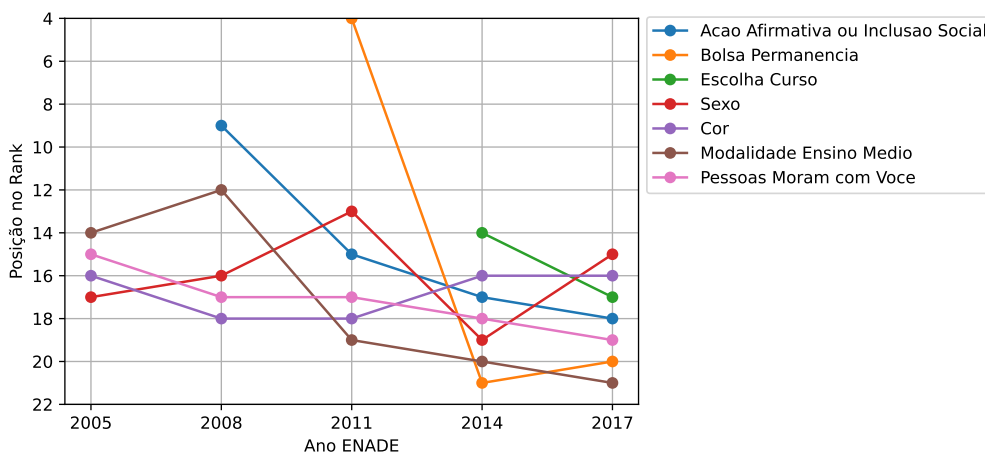
resultado (em terceiro), caindo para a quarta colocação em 2017. O atributo Bolsa/Financiamento inicia em 2005 com modesta importância, mas a partir de 2011 torna-se o mais relevante. Bolsa Acadêmica ficou em quinto em 2005, caindo para nono em 2011. No entanto, tornou-se o segundo atributo mais importante nos dois últimos anos considerados neste estudo. O último atributo analisado neste gráfico é Escolarização Mãe, que se mostrou estável ao longo dos anos, ficando sempre entre o sexto e o oitavo lugar.

Os atributos classificados da 8<sup>a</sup> a 14<sup>a</sup> posição do *ranking* geral são apresentados na Figura 2. Estudo Semanal e Onde e com quem Mora são atributos que perderam a importância ao longo dos anos. Escola Ensino Médio se manteve estável na posição 12<sup>a</sup> de 2005 a 2011. Depois, subiu no ranking até se tornar o terceiro atributo mais importante. Renda Familiar começou em sétimo, caiu até a 16<sup>a</sup> posição em 2011 e terminou em quinto em 2017, mostrando uma forte recuperação. Três atributos apresentaram certa variação, dependendo do ano, sendo eles: Escolarização Pai, Situação Financeira e Estado Civil.



**Figura 2. Evolução do perfil de estudantes de Engenharia com bom desempenho no ENADE ao longo dos anos (parte 2).**





**Figura 3. Evolução do perfil de estudantes de Engenharia com bom desempenho no ENADE ao longo dos anos (parte 3).**

Finalmente, a Figura 3 mostra os últimos atributos do *ranking* geral. Nota-se que Bolsa Permanência saiu de quarto em 2011 e despencou para a 20ª posição em 2017. Outros quatro atributos apresentaram um comportamento decrescente ao longo dos anos, sendo eles: Ação Afirmativa ou inclusão Social, Escolha Curso, Modalidade Ensino Médio, e Pessoas Moram com Você. O atributo sexo apresentou bastante variação ao longo dos anos, oscilando da 13ª até a 19ª posição. Por fim, o atributo Cor é o mais estável, iniciando em 2005 e finalizando em 2017 na 16ª posição.

## 5. Conclusão

Este trabalho propôs a análise do conjunto de dados composto pelos concluintes dos cursos de Engenharia que realizaram o ENADE no período de 2005 a 2017. Foram aplicadas diferentes técnicas de mineração de dados educacionais, mais especificamente, cinco diferentes algoritmos para seleção de atributos, os quais selecionaram dez características pessoais e socioeconômicas mais relevantes para o bom desempenho no exame.

Quatorze atributos diferentes foram considerados importantes, variando em cada ano do exame. Os relevantes em todos os anos, com o quantitativo de anos entre parênteses, foram: Idade (5), Situação de Trabalho (4), Bolsa Acadêmica (4), Escolarização Mãe (4), Família com Curso Superior (3), Escolarização do Pai (3), Bolsa/Financiamento (3), Escola Ensino Médio (3), Estudo Semanal (2), Renda Familiar (2), Escolha Instituição (2), Onde e Com Quem Mora (1), Estado Civil (1) e Bolsa Permanência (1).

Por fim, considerando a variação do perfil do estudante ao longo dos anos, observamos que os atributos Idade, Situação de Trabalho, Família com Curso Superior, Estudo Semanal, Onde e com quem Mora, Bolsa Permanência, Ação Afirmativa ou Inclusão Social e Modalidade Ensino Médio perderam relevância. Já os atributos Bolsa/Financiamento, Escola Ensino Médio e Renda Familiar ganharam importância.

Os resultados apresentados nesse estudo, combinados com dados coletados de outras fontes, podem ser utilizados para melhorar a tomada de decisão de gestores da Educação em nível nacional. Por exemplo, garantir o financiamento para a atual de-

manda de projetos de ensino, pesquisa, extensão, desenvolvimento e inovação submetidos para as agências de fomento aumentará de forma significativa a oferta de quotas de bolsas acadêmicas nas instituições, as quais se mostraram essenciais no perfil dos estudantes com bom desempenho no ENADE. Outras ações importantes incluem direcionar a elaboração de políticas públicas que promovam a elevação da qualidade do ensino médio. Por fim, políticas que retomem o crescimento econômico são necessárias. Percebe-se claramente, após mais de uma década de incremento do PIB e do PIB per capita, que o período de recessão compreendido entre 2014 e 2016 influenciou diretamente na inversão de importância do atributo Renda Familiar.

## 6. Agradecimentos

Esta pesquisa foi parcialmente financiada pelas agências de fomento: CNPq (305805/2021-5) e FAPERGS (23/2551-0000126-8).

## Referências

- Araújo, R. A. et al. (2019). Análise dos microdados do enade: Proposta de uma ferramenta de exploração utilizando mineração de dados. Dissertação de mestrado, Ciência da Computação, Universidade Federal de Goiás, Goiânia, Goiás.
- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19(02):03.
- Barbosa, E., Cavalcanti, R., Mateus, J., and Ferreira, L. (2019). Identificação dos perfis de estudantes com nota elevada no enade. *RGCTI*, 3:17–29.
- Bertolin, J. C. G., Fioreze, C., and Barão, F. R. (2022). Higher education and educational inequality in Brazil: Elitist heritage in a context of expanding access. *SciELO Preprints*.
- Corso, G. and Resende, M. (2018). Mineração de dados aplicada ao enade. Trabalho de conclusão de curso, Licenciatura em Ciência da Computação, Universidade de Brasília.
- de Almeida, D. A., Almeida, S. P. N. d. C. e., and Amorim, M. M. T. (2021). Profile of graduates in mathematics degree: An analysis from enade's data (2005-2017). *SciELO Preprints*.
- Faria, E. d. O. (2017). Uma análise dos fatores determinantes do desempenho dos alunos dos cursos superiores em administração do distrito federal. Dissertação de mestrado, Administração, Universidade de Brasília, Brasília, DF.
- Fernandes, J. C. (2013). O mapa do apagão de engenheiros no Brasil. In *XLI Congresso Brasileiro de Educação em Engenharia*, volume 41, pages 1–9.
- Fernando Choji, M., Diego N. Damasceno, C., Ibert Bittencourt, I., and Isotani, S. (2021). Mineração de dados do enade de 2016 a 2018: uma análise sobre o município de araçatuba/sp. *RENOTE*, 19(2):183–192.
- Gondran, E., Lucca, G., Berri, R., Santos, H., and Borges, E. (2022). Analyzing the determinant characteristics for a good performance at enade Brazilian exam stratified by teaching modality: Face-to-face versus online. In *Proceedings of the 24th International Conference on Enterprise Information Systems - Volume 1: ICEIS*, pages 234–242. INSTICC, SciTePress.

- Hall, M. A. (1998). *Correlation-based Feature Subset Selection for Machine Learning*. PhD thesis, University of Waikato, Hamilton, New Zealand.
- Hall, M. A. and Smith, L. A. (1998). Practical feature subset selection for machine learning. In *Australasian Computer Science Conference ACSC'98*, volume 20, pages 181–191, Berlin. Springer.
- Inep (2022). Microdados do enade. Disponível em <https://www.gov.br/inep/pt-br/area-informacao/dados-abertos/microdados/enade>. Acesso: junho de 2023.
- Karegowda, A. G., Manjunath, A., and Jayaram, M. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management*, 2(2):271–277.
- Kira, K. and Rendell, L. A. (1992). A practical approach to feature selection. In Sleeman, D. H. and Edwards, P., editors, *Ninth International Workshop on Machine Learning*, pages 249–256. Morgan Kaufmann.
- Kononenko, I. (1994). Estimating attributes: Analysis and extensions of relief. In *Machine Learning: ECML-94*, pages 171–182, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Miao, J. and Niu, L. (2016). A survey on feature selection. *Procedia Computer Science*, 91:919–926.
- Rosa, E. R., Ferreira, D. J., Silva, N. F. F. d., and Assis, A. (2021). Estudo exploratório através de análises longitudinais aplicado à ciência da computação a partir da base de dados do enade. *Revista Brasileira de Informática na Educação*, 29:1463–1486.
- Salerno, M. S., Lins, L. M., Araújo, B. C., Gomes, L. A. V., Toledo, D., and Nascimento, P. A. M. M. (2014). Uma proposta de sistematização do debate sobre falta de engenheiros no brasil. Technical report, Texto para Discussão.
- Smolentseva, A. (2018). *High Participation Systems of Higher Education*, chapter High Participation Society, pages 184–200. Oxford University Press.
- Tan, P.-N., Steinbach, M., and Kumar, V. (2016). *Introduction to data mining*. Pearson Education India.
- Tasca, M. (2015). *Contribuições ao Problema de Seleção de Atributos*. Tese de doutorado, Computação, Universidade Federal Fluminense, Niterói, Rio de Janeiro.
- Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., and Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, 85:189–203.
- Vieira, A., Bertolini, D., and Schwerz, A. (2022). Análise do desempenho no enade dos concluintes de computação usando técnica de agrupamento. In *Anais do XXXIII Simpósio Brasileiro de Informática na Educação*, pages 834–845, Porto Alegre, RS, Brasil. SBC.
- Yu, L. and Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the Twentieth International Conference on Machine Learning*, pages 856–863. AAAI Press.