

Análise estatística descritiva e Regressão da Inserção das Mulheres nos cursos de TI nos anos de 2009 a 2018.

Ariana Rodrigues Cursino

08 DE MARÇO, 2021



Apresentação

PRINCIPAIS TÓPICOS

Introdução

Desenvolvimento

Resultados e Discussões

Conclusão

Referências Bibliográficas

Dúvidas - Contato

Introdução

DEFINIÇÃO DO PROBLEMA **DESIGUALDADE DE GÊNERO NA ÁREA DA TECNOLOGIA DA INFORMAÇÃO**

- Os cursos da área de Tecnologia, são majoritariamente ocupados por homens. [1]
- Entre a década de 1960 e 1970 houve um crescimento dos cursos na área da computação e a crescente participação das mulheres. [2][3]
- A partir de 1980, iniciou-se a construção de estereótipos e identidades que associavam a computação ao universo masculino e a cultura geek. [2][3]



- 82,8% das mulheres entrevistadas relata ter vivido, ou ainda vivenciar, preconceito de gênero dentro do seu ambiente de trabalho.
- Em relação à área acadêmica, 61,8% assegura ter vivido ou vivência este preconceito.
- O estudo realizado pela empresa de consultoria é bastante revelador, trazendo outros importantes números: 91% das entrevistadas afirmam que ainda existe preconceito dentro das empresas e que essas ainda estão dando os primeiros passos direcionados à implementação de políticas de diversidade e inclusão;
- 72% afirmam que o ambiente familiar não costuma estimular meninas a gostarem de brincadeiras ou carreiras ligadas à tecnologia;
- 42% das participantes afirmam que o maior desafio é ter de provar a todo tempo que são profissionais competentes;
- **Empresa de consultoria Yoctool[4]**

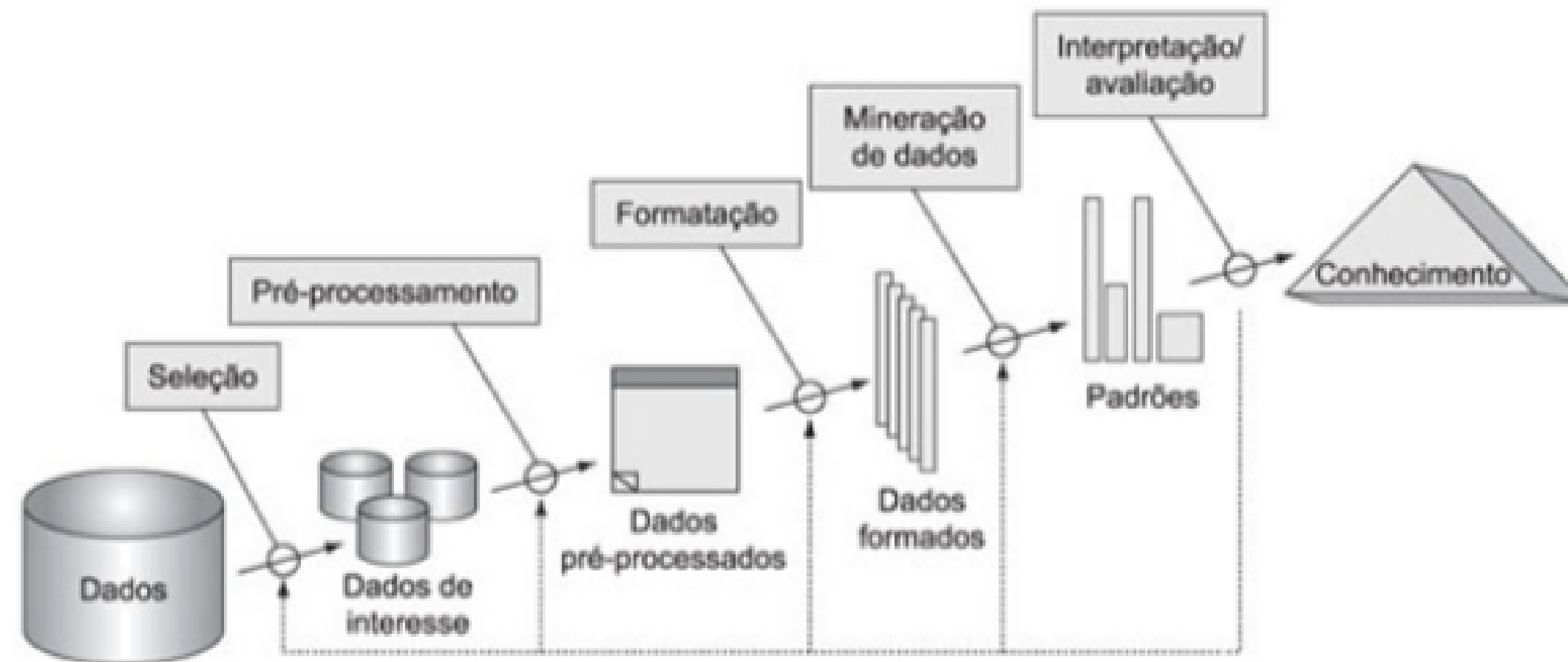
Introdução

OBJETIVO

- Análise descritiva dos dados do Censo da Educação Superior dos anos de 2009 a 2018.
- Criar métricas (como o perfil dos alunos - idade e gênero) de apoio às coordenações e também para as comunidades que incentivam as mulheres.
- Regressão, para traçar a tendência da taxa de crescimento da inserção dos alunos nos cursos de Tecnologia da Informação (TI).

Desenvolvimento

MÉTODO KNOWLEDGE DISCOVERY IN DATABASES (KDD)



Fonte: Adaptado[5]

Desenvolvimento



DADOS E SELEÇÃO

- Os dados brutos dos anos de 2009 a 2018, foram baixados do site do INEP.
- Arquivos .csv com dados dos alunos, dados dos cursos e dicionário das variáveis.

Desenvolvimento

PRÉ-PROCESSAMENTO E FORMATAÇÃO

- Linguagem de programação Python 3.
- Bibliotecas de suporte Pandas, Numpy, Matplotlib, Seaborn, Statsmodels.
- Aplicativo da web Jupyter Notebook.
- O volume de dados de 10 anos da educação superior é muito grande.
- Filtro com os seguintes termos: "Computação", "Tecnologia da Informação", "Informática" e "Análise de Desenvolvimento de Sistemas"

Desenvolvimento

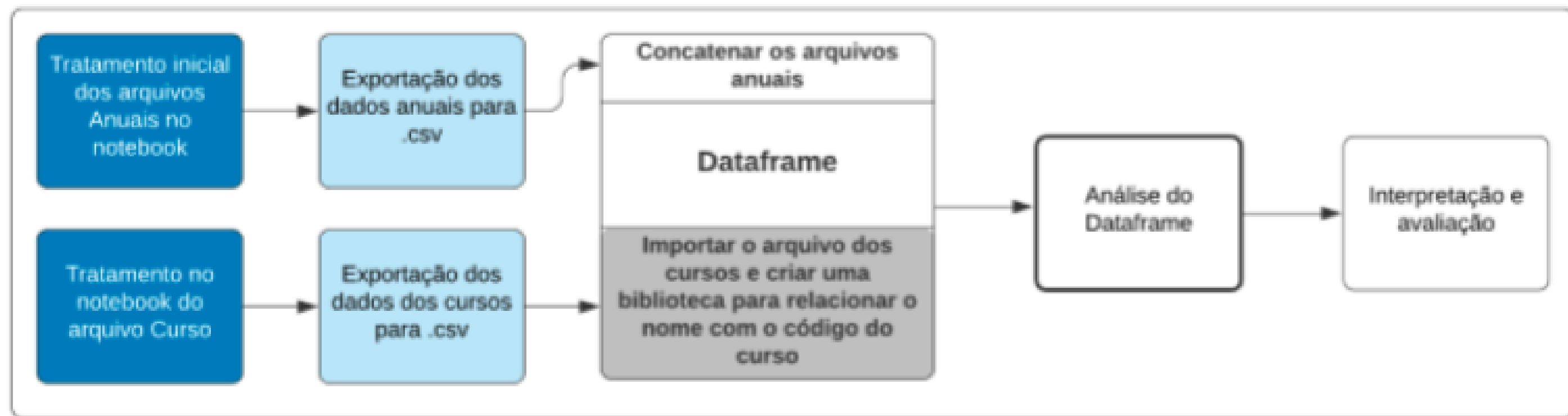
MINERAÇÃO E ANÁLISE DOS DADOS

- Contagem de alunos por curso e as medidas de tendência central.
- Cruzamento dos dados divididos por gênero e por idade, assim como a porcentagem de mulheres e homens nos cursos.
- Cruzamento dos dados por gênero e por ano - Inserção das mulheres na área ao longo dos anos analisados.
- Relação dos dados e sua causalidade, a correlação e a covariância.
- Gráfico de dispersão e, com o apoio da biblioteca Statmodels, os gráficos de regressão da taxa de crescimento dos homens e das mulheres nos cursos de TI.
- Para validar o modelo, foi utilizada a biblioteca sklearn.metrics, para o cálculo dos erros médios.

Desenvolvimento

INTERPRETAÇÃO/AVALIAÇÃO E CONHECIMENTO

- Será melhor discutido no próximo Capítulo - Resultados e Discussões.
- Fluxograma da Implementação :



Resultados e Discussões

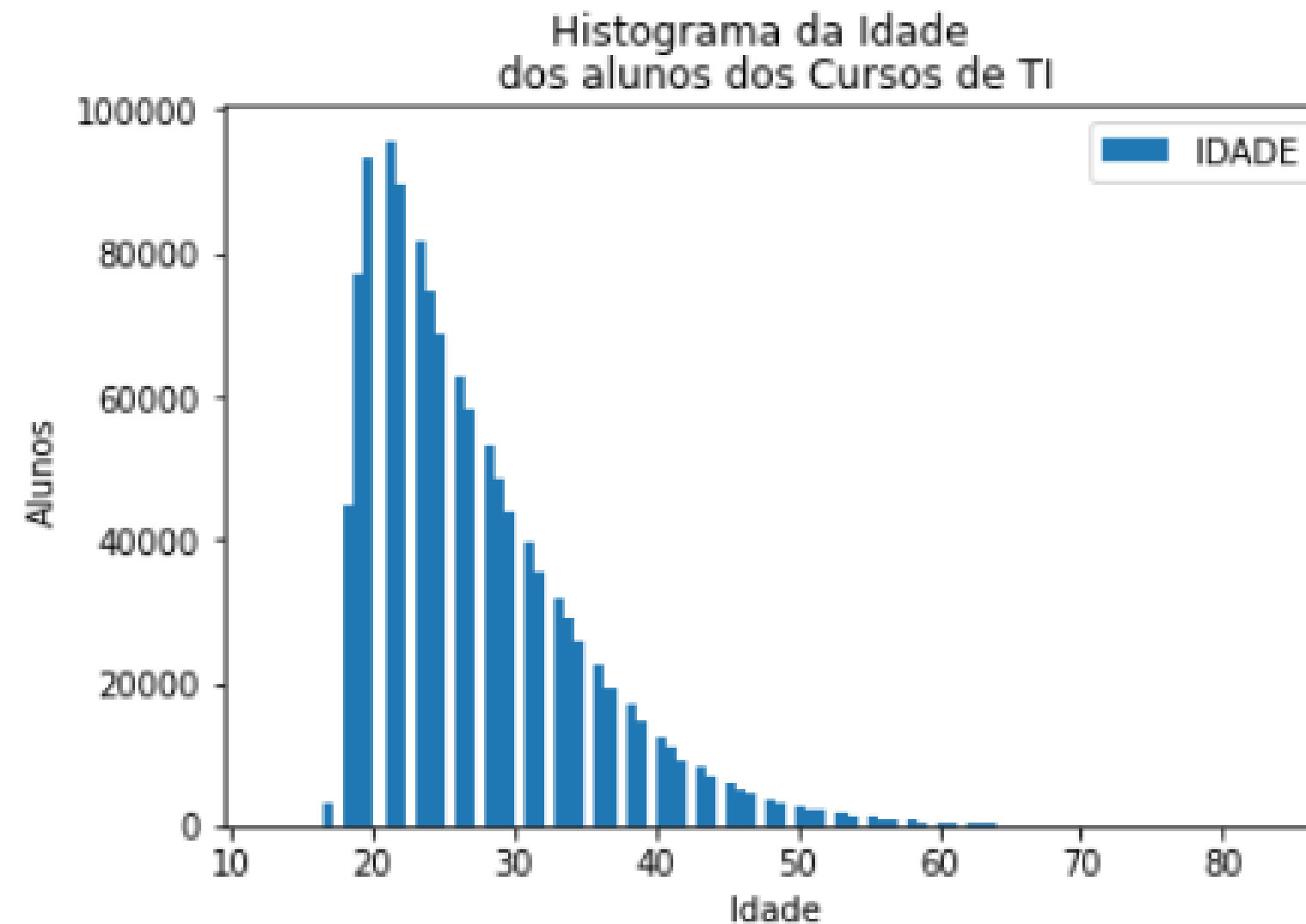
- Dataframe final conta com os dados de **1.217.117** alunos, sendo a idade mínima de **13** anos e a máxima de **84** anos.

- Medidas Centrais:

Sexo	Média	Mediana	Moda
Masculino	27 anos	25 anos	21 anos
Feminino	26 anos	25 anos	20 anos

Resultados e Discussões

- O Histograma da idade dos alunos:



75% dos alunos estão na região de 20 a 30 anos.

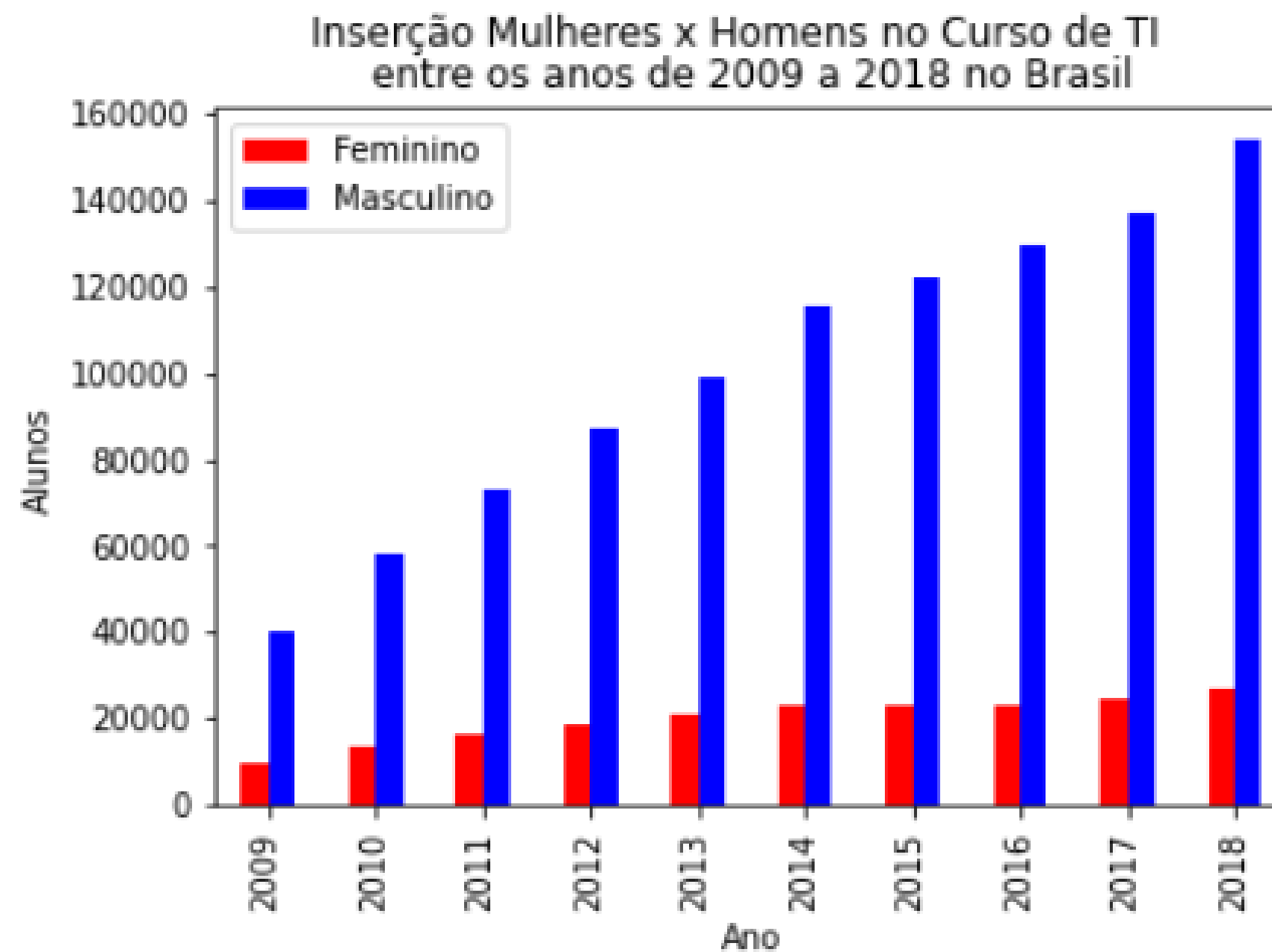
Resultados e Discussões

- Recorte de Gênero:

Ano/Sexo	Feminino	Masculino	Feminino (%)	Masculino (%)
2009	10164	40309	20.14	79.86
2010	13887	58140	19.28	80.72
2011	16777	73014	18.68	81.32
2012	19052	86917	17.98	82.02
2013	20876	98899	17.43	82.57
2014	23483	115072	16.95	83.05
2015	23559	121969	16.19	83.81
2016	53639	129352	15.45	84.55
2017	24430	136797	15.15	84.85
2018	27247	153534	15.07	84.93

Resultados e Discussões

- Recorte de Gênero:



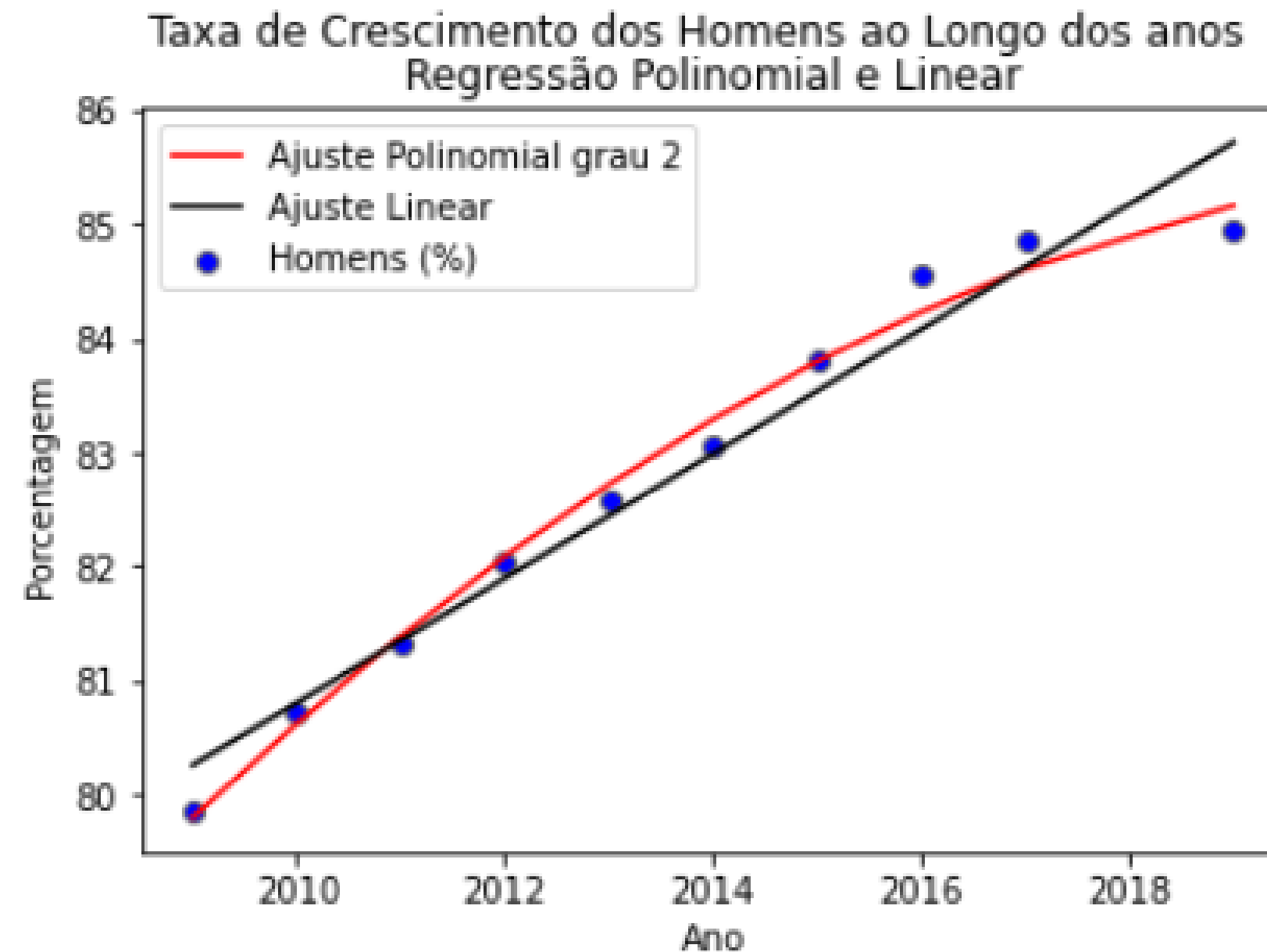
Resultados e Discussões

- Correlação e a covariância em relação a taxa de porcentagem do crescimento da inserção por gênero ao longo dos anos:

Ano/Gênero	Feminino	Masculino
Correlação	-0.979503	0.979503
Covariância	-5.595556	5.595556

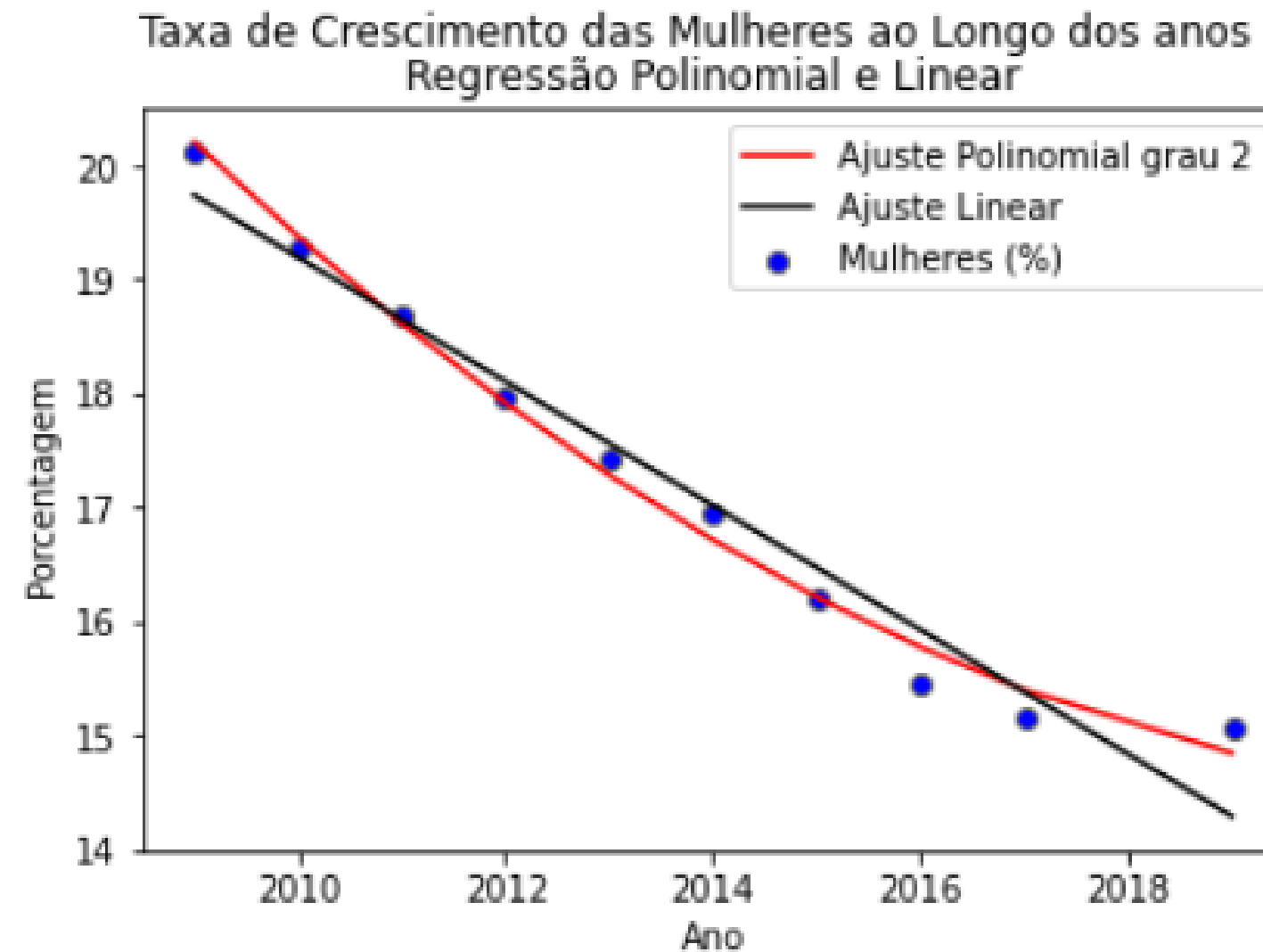
Resultados e Discussões

- Regressão para verificar se uma variável independente (Ano), tem relação com uma variável dependente (Número de alunos por sexo e por ano):



Resultados e Discussões

- Regressão para verificar se uma variável independente (Ano), tem relação com uma variável dependente (Número de alunas por sexo e por ano):



Resultados e Discussões

Validação do Modelo:

- Os valores de **r-quadrado** , também conhecido como, coeficiente de determinação e indica o poder preditivo da reta de regressão.
- Em ambas regressões linear e polinomial foi de 0,959 e 0,989, respectivamente. Nesse caso afirma-se que 96% e 99%, respectivamente das variações do ano podem ser explicadas com base na variação do número de alunos inseridos nos cursos.
- Indicando que ambas as regressões apresentam ótimo modelo para predizer a taxa de crescimento dos homens e das mulheres nos anos de 2008 a 2019.

Resultados e Discussões

Validação do Modelo:

- As métricas da raiz quadrática média, também conhecida como **rmse** (root mean squared error) e o erro absoluto médio, conhecido como mae (mean absolut error), também foram utilizadas para validar o modelo. Ambas as medidas expressam o erro médio do modelo preditivo, em relação aos dados originais.
- Para a regressão linear, os valores encontrados de rmse e de mae foram 0,34 e 0,26, respectivamente, para ambos os gêneros.
- Para a regressão polinomial os valores foram 0,18 e 0,15, respectivamente, para ambos os gêneros. Novamente a regressão polinomial apresentou valores menores de erro em relação a regressão linear.

Conclusão

- A análise empreendida aqui busca analisar a participação das mulheres nos cursos da tecnologia da informação a partir da categoria gênero, realizando para isso uma análise descritiva dos dados.
- Tal análise evidencia o comportamento e o fluxo da inserção dos alunos ao longo dos anos. Observa-se que o número de mulheres nos cursos da área de TI, está cada vez mais diminuindo. Análises similares foram encontrados em estudos do curso de Ciências da Computação de Moreira, Silva e Carvalho[6] na UFPB e de Santos[1], no IME-USP.

Conclusão

- As administrações educacionais devem usar esses dados com a intenção de criar métodos para evitar a redução da procura dos cursos por mulheres e torná-los mais atraentes e inclusivos para o público feminino.
- Na tomada dessas decisões, é necessário criar indicadores e destacar os pontos relevantes para a coordenação do curso.
- Para trabalhos futuros, recomenda-se que se utilizem fatores relacionados ao tema, para estabelecer um modelo preditivo com maior consistência e maior precisão, verificando assim a relação da diminuição da inserção das mulheres nos cursos de TI, a outros fatores relacionados à área.

Referências Bibliográficas

- [1] SANTOS, C. M. Por que as mulheres 'desapareceram' dos cursos de computação?. Jornal da USP, 2018. Disponível em: <https://jornal.usp.br/universidade/por-que-as-mulheres-desapareceram-dos-cursos-de-computacao>. Acesso em 10/08/2020.
- [2] ABBATE, J. The Pleasure Paradox: Bridging the Gap between Popular Images of Computing and Women's Historical Experiences. In: SOUZA, T. P de. Seminário Internacional Fazendo Gênero. Congresso, Florianópolis, 2017. Disponível em: <https://bit.ly/37oKgyJ> Acesso em 01/06/2020.
- [3] HAYES, C. C. Computer Science: The incredible shrinking Woman. In: SOUZA, T. P de. Seminário Internacional Fazendo Gênero. Congresso, Florianópolis, 2017. Disponível em: <https://bit.ly/37oKgyJ> Acesso em 01/06/2020.
- [4] YOCTOO, 2019, in: SILVA, José; OLIVEIRA, Letícia; SILVA, André. Meninas na Computação: uma análise inicial da participação das mulheres nos cursos de Sistemas de Informação do estado de Alagoas. In: Anais do XXVII Workshop sobre Educação em Computação. SBC, 2019. p. 444-452.

Referências Bibliográficas

[5] FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37-37, 1996.

[6] MOREIRA, J. A.; SILVA, R. M.; CARVALHO, M. E. P. Cenários Prospectivos: Uma Visão do Futuro da Presença Feminina em Cursos de Ciência da Computação de uma Instituição de Ensino Superior. In: *Anais do XXVI Workshop sobre Educação em Computação*. SBC, 2018.

Dúvidas

Email

arcursino@gmail.com

GitHub

www.github.com/arcursino