Inteligência Artificial

Preparação e Pré-processamento dos dados
Parte I

Prof. Dr. Ivan Carlos Alcântara de Oliveira

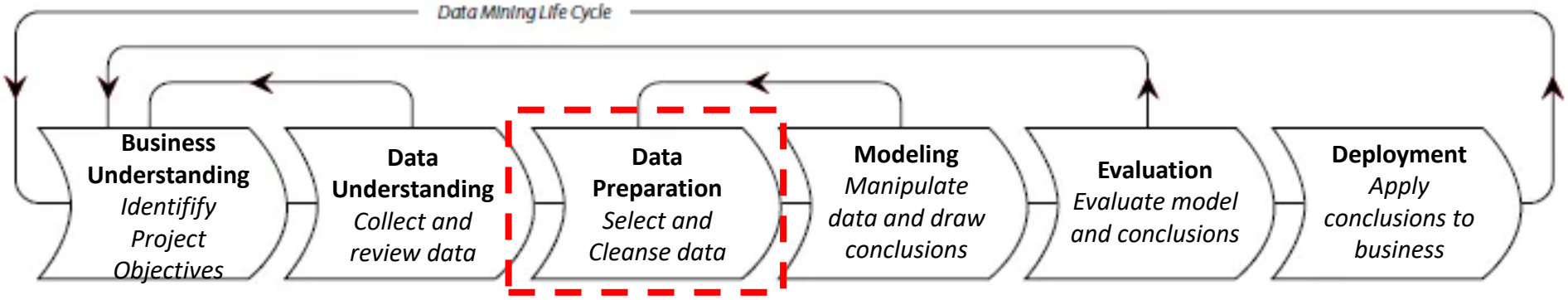
https://orcid.org/0000-0002-6020-7535

Preparação dos dados e Pré-Processamento

Ciclo de Vida de Projeto de Data Science

CRISP-DM

- **CRISP-DM (*Cross Industry Standard Process for Data Mining* -** Processo Padrão Interindústrias para Mineração de Dados) é um processo de fases bastante aceito na indústria para representar um ciclo completo de Ciência e Análise de Dados, incluído a aplicação de modelos de Aprendizado de Máquina.



Auxiliar na Identificação	Obtenção dos dados	Pré-processamento	Seleção do modelo	Validação do modelo	Aplica solução obtida ao negócio
Compreensão	Integração	Transformação de Variáveis	Cross-Validation	Otimização	Desenvolvimento de Aplicação Analítica
Pesquisas e Estudos	Análise Exploratória	Seleção de Atributos	Métricas de performance		Geração de relatório
Reunião com Especialistas		Redução de Dimensionalidade	Otimização		Construção de Dashborads etc.
Definição dos Atributos		Amostragem			
Definição medida de erro					

Ciência de Dados

Ciclo de Vida de Projeto de *Data Science*

3. **Preparação dos Dados**

Data Wrangling: pré-processamento, transformação, limpeza dos dados, etc.

Até aqui, aproximadamente 70% do tempo de um Cientista de Dados.

Preparação dos Dados

- Algoritmos de Aprendizado de Máquina aprendem a partir dos dados.
- O profissional de análise deve “alimentar” os algoritmos com dados consistentes, de acordo com o problema a resolver.
- Mesmo que o conjunto de dados (*dataset*) esteja bom, é preciso garantir que ele esteja na mesma escala, no mesmo formato e que os atributos (variáveis) mais significativos pertençam a ele.

Preparação dos Dados

- A etapa de pré-processamento de dados recebe muitas nomenclaturas: *Data Munging*, *Data Wrangling*, *Data Preparation*, *Data Processing* ou simplesmente “Preparação de dados”.

Preparação dos Dados

- O **desempenho** dos algoritmos de AM geralmente são **afetado**s pelo **estado dos dados**.
  - Conjunto de dados **podem apresentar diferentes características**, **dimensões** ou **formatos**.
  - Por exemplo, podem ser numéricos ou categóricos/simbólicos.

Preparação dos Dados

- **Conjunto de dados podem também**:
  - Estar limpos ou **conter ruídos** e imperfeições
  - Com valores incorretos
  - **Inconsistentes**
  - **Duplicados ou ausentes**
  - Ter atributos independentes ou relacionados
  - **Apresentar poucos ou muitos objetos**
  - Ter um **número pequeno ou elevado de atributos**.

Preparação dos Dados

- **Técnicas de Preparação dos** dados são frequentemente **utilizadas** para **melhorar** a **qualidade dos dados** por meio da eliminação ou minimização desses problemas.

Preparação dos Dados

- Essa melhora pode:
  - **facilitar o uso de técnicas de AM**
  - **permitir a construção de modelos mais fiéis à distribuição real dos dados**
  - **Reduzir a complexidade computacional**
  - **Tornar mais fáceis e rápidos o ajuste de parâmetros** do modelo e seu posterior uso.
  - **Facilitar a interpretação dos padrões extraídos pelo modelo**

Preparação dos Dados

- Técnicas de Preparação dos dados também **podem tornar os dados mais adequados para** sua utilização por um determinado **algoritmo de AM**.
- **Por exemplo**, **alguns** algoritmos de AM **trabalham apenas com valores numéricos**.
- **Não existe** uma **ordem** fixa **para aplicação** das diferentes **técnicas** de **pré-processamento**.

Preparação dos Dados

Integração de Dados

- Quando **dados** para aplicação de AM estão **distribuídos em diferentes conjuntos de dados**, eles **devem ser integrados antes do início** do uso **da técnica**.

- Na integração, é **necessário identificar quais objetos estão presentes nos diferentes conjuntos a serem combinados.**

Preparação dos Dados

Integração de Dados

- A **existência de um conjunto grande de dados**, **tanto em** termos de **número de objetos** como de **atributos**, **não implica que** um **algoritmo de AM deve utilizar todo ele**.

- **Muitas vezes** é **mais eficiente utilizar parte do conjunto original**.

Preparação dos Dados

Integração de Dados

- Um **número elevado de atributos** pode **comprometer o desempenho do algoritmo**.
- Um **grande número de objetos** podem ocorrer **problemas de saturação de memória** e **aumento do tempo computacional** para **ajustar os parâmetros do modelo**.
- Técnicas de **amostragem** podem **minimizar esses problemas**.

Preparação dos Dados

Eliminação manual de Atributos

- Quando um **atributo não contribui** para a **estimativa** do valor do **atributo alvo**, ele **é** considerado **irrelevante**.

- O conjunto de atributos que formarão o conjunto de dados a ser analisado é **geralmente definido** de acordo com a experiência de **especialistas do domínio dos dados**.

Preparação dos Dados

Eliminação manual de Atributos

- Um **atributo** que **possui o mesmo valor para todos os objetos**, também **é irrelevante**, pois não contém informação que ajude a distinguir objetos.

- **Um atributo não precisa ter exatamente o mesmo valor para todos os objetos para ser considerado irrelevante**.

Preparação dos Dados

Eliminação manual de Atributos – Exemplo dos dados dos pacientes de um hospital – Especialistas decidem, alguns atributos não são relevantes para diagnóstico clínico

Id.	Nome	Idade	Sexo	Peso	Manchas	Temp.	# Int.	Est.	Diagnóstico
4201	João	28	M	79	Concentradas	38,0	2	SP	Doente
3217	Maria	18	F	67	Inexistentes	39,5	4	MG	Doente
4039	Luiz	49	M	92	Espalhadas	38,0	2	RS	Saudável
1920	José	18	M	43	Inexistentes	38,5	8	MG	Doente
4340	Cláudia	21	F	52	Uniformes	37,6	1	PE	Saudável
2301	Ana	22	F	72	Inexistentes	38,0	3	RJ	Doente
1322	Marta	19	F	87	Espalhadas	39,0	6	AM	Doente
3027	Paulo	34	M	67	Uniformes	38,4	2	GO	Saudável

Eliminação dos atributos Id., Nome e Est.

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
19	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Preparação dos Dados

Limpeza de Dados

- **Conjunto de dados podem apresentar dificuldades relacionadas à qualidade dos dados**.
- Por exemplo, dados:
  - **Incompletos**: ausência de valores para alguns atributos em parte dos dados.
  - **Inconsistentes**: que não combinam ou contradizem valores de outros atributos do mesmo objeto.
  - **Redundantes**: dois ou mais objetos apresentam os mesmos valores para todos os atributos ou dois ou mais atributos têm os mesmos valores para dois ou mais objetos.
  - **Ruidosos**: possuem erros ou valores diferentes do esperado.

Preparação dos Dados

Limpeza de Dados

- Dados com valores ausentes são de fácil detecção, mas dados inconsistentes, redundantes e ruidosos são mais difíceis.

- **Algumas técnicas de AM conseguem lidar bem com** algumas dessas **imperfeições nos dados**.

Preparação dos Dados

Limpeza de Dados

- Porém, **mesmo se a técnica for robusta** o suficiente para lidar com as imperfeições em um conjunto de dados, elas **pode**m **reduzir a qualidade das análises realizadas**.
- A **presença dessas deficiências pode resultar em estatísticas e análises incorretas**.
- Portanto, **todas as técnicas se beneficiam na melhora da qualidade dos dados**.

Preparação dos Dados

Limpeza de Dados – **Dados Incompletos**

- Um dos problemas em conjuntos de dados é a **ausência de valores para alguns atributos de alguns objetos**.

Preparação dos Dados

Limpeza de Dados – **Dados Incompletos – Exemplo: Dados de Pacientes de Hospital**

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
—	M	79	—	38,0	—	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	—	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
—	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Dados Incompletos

Preparação dos Dados

Limpeza de Dados – **Dados Incompletos - Alternativas**

a) Eliminar os objetos com valores ausentes.

Não indicada quando:

- poucos atributos do objeto possuem valores ausentes.
- o número de atributos com valores ausentes varia muito entre os objetos.
- o número de objetos restantes for pequeno.

Preparação dos Dados

Limpeza de Dados – **Dados Incompletos - Alternativas**

b) Definir e preencher manualmente os valores.

Não factível quando:

- o número de objetos ou atributos com valores ausentes for muito grande.

Preparação dos Dados

Limpeza de Dados – **Dados Incompletos - Alternativas**

c) **Empregar algoritmos de AM que lidam internamente com valores ausentes**. Por **exemplo**, algoritmos **indutores de árvore de decisão**.

d) **Utilizar algum método ou heurística para automatizar a definição de valores** para os atributos ausentes.

Preparação dos Dados

Limpeza de Dados – **Dados Incompletos - Alternativas – Definição automática de valores - Abordagens**

e) **Criar** para o atributo **um novo valor que indique** que o atributo possuía **um valor desconhecido**.

Problema:

- **Algoritmo indutor pode assumir que o valor desconhecido representa um conceito importante**.

Preparação dos Dados

Limpeza de Dados – **Dados Incompletos - Alternativas – Definição automática de valores – Abordagens – Estatística**

f) **Utilizar a média, mediana ou a moda[1]** (no caso de valor simbólico) **dos valores conhecidos para esse atributo**.

Observação:

- A imputação de valores ausentes pela **média pode levar à inconsistências**, como, por **exemplo**, um **paciente de 2 anos de idade** com **peso igual a 60 kilos**.

1 - Moda: valor mais frequente no conjunto de dados.

Preparação dos Dados

Limpeza de Dados **– Dados Incompletos - Alternativas – Definição automática de valores – Abordagens - Indutor**

g) **Empregar um indutor para estimar o valor do atributo**.

O valor a ser definido seria o atributo alvo e **os demais seriam os atributos de entrada**.

Vantagem do método:

Utiliza a informação presente nos demais atributos para inferir o valor do atributo ausente.

Preparação dos Dados

Limpeza de Dados – Dados Incompletos

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
—	M	79	—	38,0	—	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	—	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
—	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Média

Moda

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
27	M	79	**Inexistentes**	38,0	4	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	**F**	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
27	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Preparação dos Dados

Limpeza de Dados **– Dados Inconsistentes**

- **São aqueles que possuem valores conflitantes em seus atributos**.

- Podem ser **resultado do processo de integração de dados** de fontes ou tabelas diferentes ou da presença de ruídos nos dados.

- Por **exemplo**: diferentes conjuntos de dados podem usar **escalas diferentes para uma mesma medida** (metros e centímetros)

Preparação dos Dados

Limpeza de Dados **– Dados Inconsistentes**

- **Algoritmos simples podem verificar automaticamente** se relacionamentos existentes entre atributos são violados.

- **Quando o conjunto não é muito grande**, dados inconsistentes **podem ser ajustados/removidos manualmente**.

Preparação dos Dados

Limpeza de Dados **– Dados Inconsistentes**

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
25	F	83000g	Inexistentes	38,5	4	Doente
19	F	87	Espalhadas	39,0	6	Doente
22	F	72	Inexistentes	38,0	3	Saudável

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
18	F	67	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
25	F	83	Inexistentes	38,5	4	Doente
19	F	87	Espalhadas	39,0	6	Doente
22	F	72	Inexistentes	38,0	3	Saudável

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes**

- Um **conjunto de dados pode possuir tanto objetos como atributos redundantes**.

- Um **objeto** é **redundante** quando ele é **muito semelhante a um outro objeto do mesmo conjunto de dados**.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes**

- Um **atributo é redundante** quando **seu valor para todos os objetos pode ser deduzido a partir do valor de um ou mais atributos**.

- **Problemas na coleta, na entrada, no armazenamento, na integração ou na transmissão de dados podem gerar objetos ou atributos redundantes**.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes - Objetos**

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
18	**F**	**67**	**Inexistentes**	**39,5**	**4**	**Doente**
49	M	92	Espalhadas	38,0	2	Saudável
18	**F**	**67**	**Inexistentes**	**39,5**	**4**	**Doente**
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	37,6	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
19	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes - Objetos**

- **Objetos redundantes** em um conjunto de dados **participam mais de uma vez no processo de ajuste de parâmetros de um modelo**.
- **Então, contribui mais que os outros objetos para a definição do modelo final**, **dando a falsa impressão** que esse objeto é **mais importante** que os demais.
- Como **resultado**, o **tempo necessário para a indução de um modelo pode aumentar**.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes - Objetos**

- Então, é **desejável a eliminação de redundâncias**, que pode ser feita em dois **passos**.

  1. **Identificação de objetos redundantes**.
  2. **Eliminação das redundâncias encontradas**.

- No **passo 2 ocorre a eliminação de objetos semelhantes ou a combinação dos valores dos atributos desses objetos**.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes - Objetos**

- Essa **eliminação** é geralmente **feita no final do processo de limpeza**.

- A **não eliminação** pode **faz**er **o algoritmo de AM atribuir** uma **importância maior** ao objeto repetido do que os demais.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes - Atributos**

- Um atributo é **considerado redundante se seu valor puder ser estimado a partir de pelo menos um dos seus atributos**.

- Isso ocorre quando dois ou mais atributos têm a mesma informação preditiva.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes - Atributos**

- Por exemplo:
  - **os atributos “idade” e “data de nascimento”**. Nesse caso, **a idade pode ser obtida facilmente pela data de nascimento**.
  - dois **atributos Quantidade de Venda e Valor da Venda e um outro atributo Venda Total**. Aqui, **o atributo venda total pode ser obtido pelo produto dos dois anteriores**.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes - Atributos**

- Um **atributo redundante pode supervalorizar um dado aspecto dos dados**, **por estar presente mais de uma vez**, **ou tornar mais lento o processo de indução**, **devido a maior quantidade de atributos a serem analisados** pelo algoritmo de AM.

- **Geralmente, são eliminados por técnicas de seleção de atributos**.

- A sua **redundância** está **relacionada a** sua **correlação** com **um ou mais atributos** do conjunto de dados.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes - Atributos**

- **Dois ou mais atributos estão correlacionados quando apresentam um perfil de variação semelhante** para os diferentes objetos.

- **Quanto mais correlacionados** os atributos, **maior o grau de redundância**.

- **Se a correlação ocorrer entre um atributo de entrada e um atributo rótulo, esse atributo de entrada terá uma grande influência na predição do valor do atributo rótulo**.

Preparação dos Dados

Limpeza de Dados – **Dados Redundantes – Atributos – Exemplo: O atributo # Vis. Indica quantas vezes um paciente esteve no hospital, igual ao # Int.**

Idade	Sexo	Peso	Manchas	Temp.	# Int.	# Vis.	Diagnóstico
28	M	79	Concentradas	38,0	2	**2**	Doente
18	F	67	Inexistentes	39,5	4	**4**	Doente
49	M	92	Espalhadas	38,0	2	**2**	Saudável
18	M	43	Inexistentes	38,5	8	**8**	Doente
21	F	52	Uniformes	37,6	1	**1**	Saudável
22	F	72	Inexistentes	38,0	3	**3**	Doente
19	F	87	Espalhadas	39,0	6	**6**	Doente
34	M	67	Uniformes	38,4	2	**2**	Saudável

Preparação dos Dados

Limpeza de Dados – **Dados com Ruídos**

- São **dados que** contêm objetos que, **aparentemente, não pertencem à distribuição que gerou os dados analisados**.

- **Ruído pode ser definido como uma variância ou erro aleatório no valor gerado ou medido para um atributo**.

Preparação dos Dados

Limpeza de Dados **– Dados com Ruídos**

- Dados com ruídos **podem levar a um superajuste do modelo utilizado**, fazendo o algoritmo se ater as especificidades em vez da distribuição verdadeira que gerou os dados.

- **Por outro lado**, **a sua eliminação pode levar à perda de informação importante** e fazer com que algumas regiões do espaço de atributos não sejam consideradas no processo de indução de hipóteses.

Preparação dos Dados

Limpeza de Dados – **Dados com Ruídos**

- **Nem sempre é possível ter certeza de que um valor é** ou não resultado da presença de **ruído**, mas **apenas ter** uma **indicação** ou indício **de que** um dado valor para um **atributo pode ter sido gerado com ruído**.

- **Exemplos** são a presença de *outliers*, ou seja, **dados** fora dos limites aceitáveis, e **muito diferentes dos demais valores observados**.

Preparação dos Dados

Limpeza de Dados – **Dados com Ruídos - Exemplo**

Idade	Sexo	Peso	Manchas	Temp.	# Int.	Diagnóstico
28	M	79	Concentradas	38,0	2	Doente
18	F	300	Inexistentes	39,5	4	Doente
49	M	92	Espalhadas	38,0	2	Saudável
18	M	43	Inexistentes	38,5	8	Doente
21	F	52	Uniformes	60,0	1	Saudável
22	F	72	Inexistentes	38,0	3	Doente
19	F	87	Espalhadas	39,0	6	Doente
34	M	67	Uniformes	38,4	2	Saudável

Possível ruído

Preparação dos Dados

Limpeza de Dados – Dados com Ruídos - Técnicas

- Além de **técnicas estatísticas**, **nem sempre realistas ou eficientes**, **outras técnicas de pré-processamento** que podem ser utilizadas **são**:
  - **Técnicas de encestamento: suavizar o valor de um atributo**. Por exemplo:

    1º os atributos são ordenados;

    2º divididos em faixas ou cestas, cada uma com o mesmo número de valores;

    3º Os valores da mesma cesta são substituídos, por exemplo, pela média ou mediana dos valores presentes na cesta.

Preparação dos Dados

Limpeza de Dados – Dados com Ruídos - Técnicas

- **Técnicas baseadas em regressão ou classificação**: Utilizam uma função de regressão para, dado um valor com ruído, estimar seu valor verdadeiro ou, se o valor for simbólico, uma técnica de classificação pode ser utilizada.

Pré-processamento de Dados

Dados Desbalanceados

- Em vários conjuntos de dados reais, **o número de objetos varia para as diferentes classes**.

- Isso é comum em aplicações em que dados de um subconjunto de classes aparecem com uma frequência maior que os dados das demais classes.
- **Por exemplo**, considere que o **total de pacientes do conjunto de dados do hospital** tenha **80% com uma dada doença** e **20% saudáveis**.
- Então, a classe com **pacientes doentes seria majoritária** e a de **saudáveis minoritária**.

Pré-processamento de Dados

Dados Desbalanceados

- Por esse fato, **é aceitável que a acurácia preditiva de um classificador** para um **conjunto de dados desbalanceados deve ser maior**, atribuindo todo novo objeto **à classe majoritária**.

- **Vários algoritmos de AM** tem seu **desempenho prejudicado na presença de dados desbalanceados**, tendendo favorecer a classificação de novos dados na classe majoritária.

Pré-processamento de Dados

Dados Desbalanceados

- **Técnicas que procuram balancear artificialmente o conjunto de dados podem ser utilizadas** para lidar com esse problema.
- As principais são:
  - **Redefinir o tamanho do conjunto de dados**
  - **Utilizar diferentes custos** de classificação **para as diferentes classes**
  - **Induzir um modelo para uma classe**

Pré-processamento de Dados

Dados Desbalanceados – **Redefinir o tamanho do conjunto de dados**

- Nesse caso, **podem ocorrer tanto o acréscimo de objetos à classe minoritária**, **como** a **eliminação de objetos da classe majoritária**.
- No entanto, ao **acrescentar os dados**, **pode induzir um modelo inadequado para os dados**, ocorrendo *overfitting*, quando o modelo é superajustado aos dados de treinamento.
- Ao **eliminar objetos da classe majoritária**, é possível que **dados de grande importância para a indução do modelo correto sejam perdidos**, ocorrendo *underfitting*, quando o modelo não se ajusta aos dados de treinamento.

Pré-processamento de Dados

Dados Desbalanceados – **Utilizar diferentes custos de classificação para as classes diferentes**

- Apresenta como **dificuldade a definição** desses **custos**.
- **Dificuldade de incorporar a consideração de diferentes custos em alguns algoritmos de AM**.
- Por exemplo, se o número de exemplos da classe majoritária for o dobro da minoritária, um erro de classificação da classe minoritária, pode equivaler à ocorrência de dois erros de classificação para um exemplo da classe majoritária.

Pré-processamento de Dados

Dados Desbalanceados **– Induzir um modelo para uma classe**

- A **classe minoritária ou a majoritária** (**ou ambas** as classes) **são aprendidas separadamente**.
- Nesse caso, pode ser utilizado algoritmo de classificação para uma classe apenas. **Esses algoritmos são treinados utilizando apenas exemplos da classe positiva**.
- A classe positiva pode ser, por exemplo, a classe minoritária.

Bibliografia

BÁSICA:

- AGGARWAL, Charu C. **Artificial Intelligence: A Textbook**. New York: Springer: 2021.
- CHOLLET, François. **Deep Learning with Python, 2ed.** Shelter Island: Manning, 2021.
- GÉRON, Aurélien. **Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems**, 2 ed**.** Sebastopol: O'Reilly, 2019.

COMPLEMENTAR:

- GOODFELLOW, Ian; BENGIO, Yoshua, COURVILLE, Aaron. **Deep Learning**. Cambridge: MIT Press, 2016.
- RASCHKA, Sebastian; MIRJALILI, Vahid. **Python Machine Learning**. 3 ed. Birmingham: Packt, 2017.
- RUSSEL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3 ed. Upper Saddle River: Pearson, 2010.
- TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining**. 2 ed. Upper Saddle River: Pearson, 2018.
- VANDERPLAS, Jake. **Python Data Science Handbook**. Sebastopol: O'Reilly, 2017.

ADICIONAIS:

- FACELI, Katti et al. **Inteligência artificial: uma abordagem de aprendizado de máquina**. 2ª Ed. Rio de Janeiro: LTC- Livros Técnicos e Científicos, 2021.
- LUGER, George F. **Inteligência Artificial** - 6ª ed. São Paulo: Pearson Education do Brasil, 2015.
- RUSSELL, Stuart; NORVIG, Peter. **Inteligência artificial: Uma Abordagem Moderna** - 4ª. Ed. GEN LTC, 2022.