



Licap

Formação Cientista de Dados



PUC Minas

Formação do Cientista de Dados

Entendimento e Modelagem conceitual do domínio de problema – Módulo Básico

Luis Enrique Zárate

Entendimento e Modelagem Conceitual do domínio de problema:

1. Re-definição precisa do problema
2. Fatos e Julgamentos
3. Análise Divergente – Convergente
4. Pró-Contra-e-Fixação
5. Diagramas de Causa-Efeito
6. Mapas conceituais.
7. Caracterização da base de dados por meio de atributos (descrição do domínio e do tipo de variável de cada atributo)

Entendimento do Domínio de Problema Licap

- O cientista de dados, com auxílio do especialista de domínio, deverá entender o domínio de problema e caracterizá-lo utilizando modelos de ontologia ou mapas conceituais.
- O objetivo desta etapa é identificar as características (variáveis ou futuros atributos) que possam compor a base de dados levando à descoberta de conhecimento útil e não óbvio.
- A experiência mostra que o conhecimento não óbvio é resultado muitas vezes de características consideradas ‘julgamentos’.

Entendimento do Domínio de Problema Licap

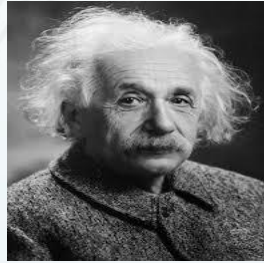
A descoberta de conhecimento em banco de dados baseado unicamente em informações de ‘fatos’ pode não levar a conhecimento relevante.

*Por exemplo, consideremos a busca de padrões para os acidentes de trânsito. Dar atenção somente aos **fatos** <imprudência, efeito do álcool, velocidade excessiva e falha mecânica> não traz conhecimento útil. Variáveis de **Julgamento**: <perfil dos condutores, perfil dos acompanhantes, etc> pode enriquecer a base de dados trazendo conhecimento não óbvio.*

O domínio de problema deve ser (E)xplorado, (C)ompreendido, (R)epresentado e (E)struturado antes de ser (M)odelado.



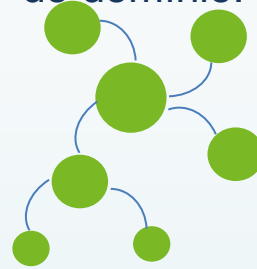
Explorado



Compreendido



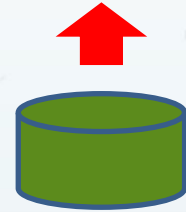
Entendimento
do domínio:



Representado



{N características}



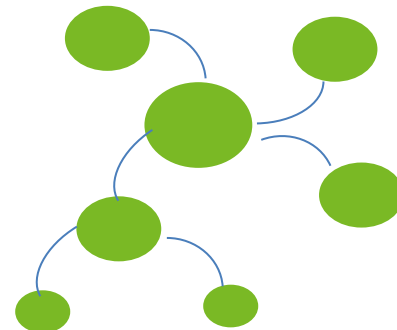
Estruturado

De acordo com Morgan D. Jones: *“a análise de problemas é feita através da estruturação do problema: isto significa, separar um problema dentro de seus elementos constituintes”.*

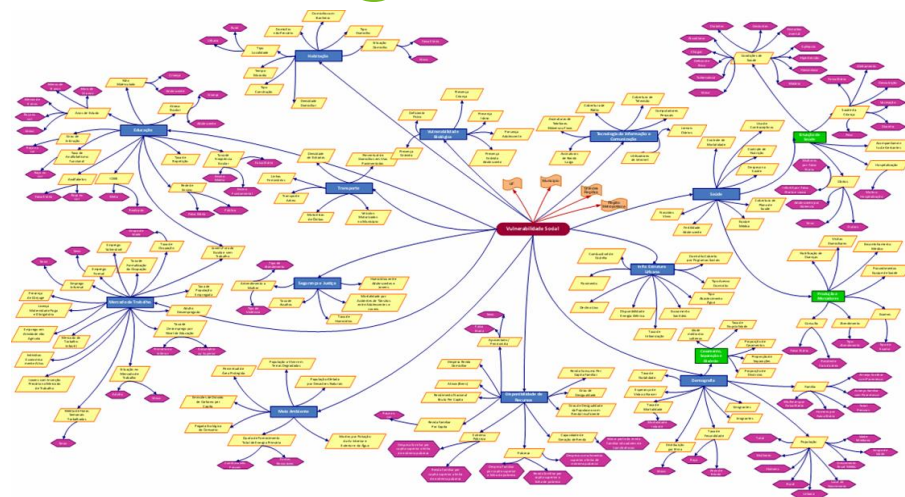
Técnicas para a Entendimento de Problemas



Entendimento do domínio:



- Re-definição precisa do problema
- Observação de Fatos e Julgamentos
- Análise Divergente - Convergente
- Pró-Contra-e-Fixação
- Mapas Cognitivos
- Resolução de Ambigüidades
- Diagramas de Causa-Efeito
- Mapas conceituais



1) Re-definição precisa do Problema

Para a definição precisa do problema é necessário procurar todas as informações que podem ser úteis.

Exemplo: Falha na linha de produção de uma fábrica
Informações adicionais

- Quais elementos constituem a falha
- Como a falha foi detectada ou medida
- Que componentes do problema devem ser observados: equipamentos, pessoal, meio ambiente, etc.



Confiança Analítica:

Os problemas podem ser categorizados pelo papel que os **FATOS** e os **JULGAMENTOS** têm haver no análise do problema.

Fato = Relevante; Julgamento = Fracamente Relevante; Forte Julgamento = Irrelevante

FATOS		JULGAMENTOS	
Confiança cai		Erro	
Simplista Existe somente uma resposta	Determinística Somente uma resposta e deve ser usada a formula correta	Aleatória Várias respostas e todas podem ser identificadas	Indeterminado Várias respostas suposições e não todas podem ser identificadas

2) Identificando Fatos e Julgamentos

Fatos:

Circunstância que causam diretamente o problema

Julgamentos:

Observações a serem disputados ou decididos

Exemplo 1- Acidente de trânsito

Imprudência

Efeito de álcool

Excessiva velocidade

Falha mecânica



Qual é o perfil do motorista de cada veículo?

Quais são os perfis dos acompanhantes dos veículos?

Quais eram as condições meteorológicas durante o acidente?

Exemplo 2- Devolução de Produtos

Defeito na envoltura

Fora da validade

Produto estragado

Fora de especificação



Quem fabrica a envoltura?

Quem comercializa o produto?

Como o produto foi estocado?

Como o produto é embalado?

Como o produto é transportado?

3) Análise Divergente / Convergente

Divergência:

Consiste em direcionar nossa mente em diferentes direções de um simples aspecto procurando novas evidências.

Convergência:

Consiste em direcionar a nossa atenção, focalizando nossa mente sobre um simples aspecto do problema.

Ambas são necessárias para a efetiva solução de problemas. A divergência abre a mente para criar alternativas e a convergência peneira alternativas fracas e fortalece as fortes.

Passos:

- 1) **Divergente:** Brainstorm
- 2) **Convergente:** Examinar e agrupar cada idéia
- 3) **Divergente:** Voltar ao ***passo 1*** para cada grupo de idéias

Exemplo: Problemas na produção de soja

Divergente: Problemas na qualidade da semente, no plantio, no cultivo, no clima, na maquinaria utilizada, etc.

Convergente: Problema na qualidade da semente

Divergente: melhoramento genético da semente, tecnica de preparação da semente, estocagem da semente, etc.

➡ 4) Pró-contra-e-fixação



Esta técnica é fundamentada na compulsão humana de ser críticos, especialmente quando algo é novo ou fora do convencional.

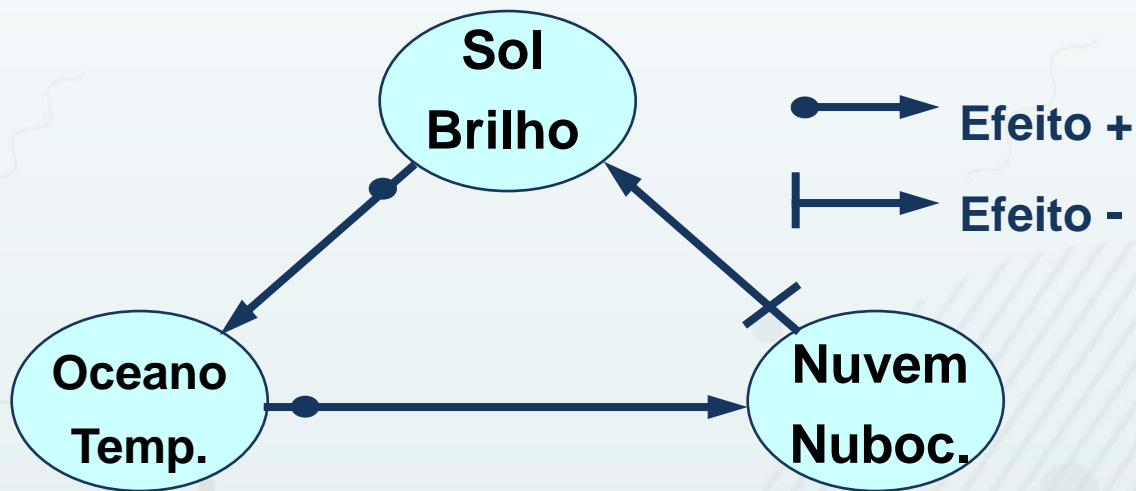
- **Passos:**

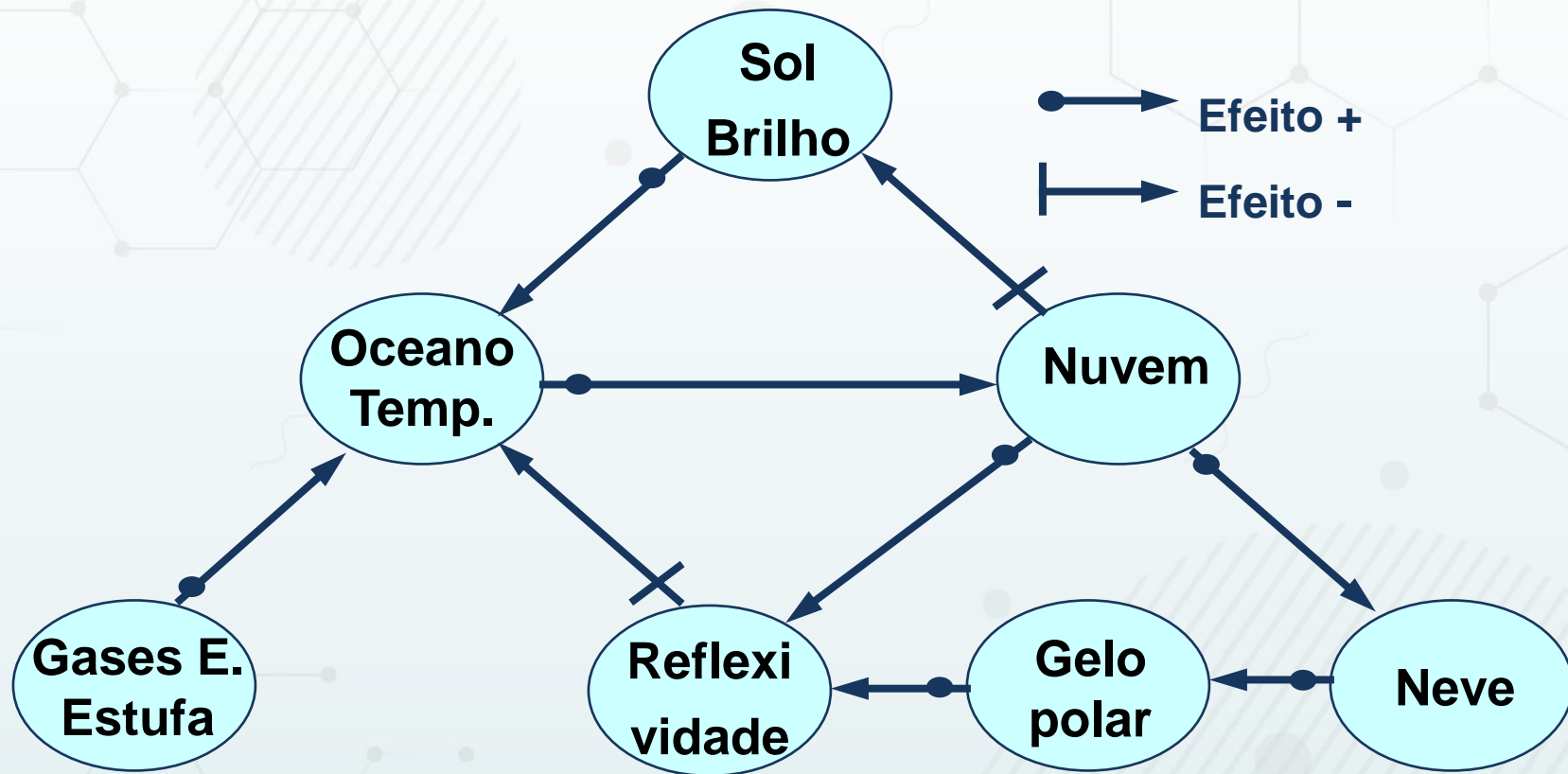
- **1) Listar todos os prós:** aspectos positivos, benefícios, méritos e vantagens.
- **2) Listar todos os contrários:** razões que não permitem que determinada situação aconteça (pensamento divergente).
- **3) Revisar e consolidar os contrários:** pensamento convergente.
- **4) Neutralizar aspectos contrários quanto possíveis**

- **Exemplo:** O que leva à compra de um veículo
 - **1) Listar todos os prós:** preço, promoção, consumo de combustível, revenda, conforto, segurança, etc..
 - **2) Listar todos os contrários:** importado, visado, raro serviço de oficina, manutenção cara, mecânica especializada, etc..
 - **3) Revisar e consolidar os contrários:** importado, **visado**, raro serviço de oficina, **manutenção cara**, mecânica especializada, etc..

5) Mapas Cognitivos

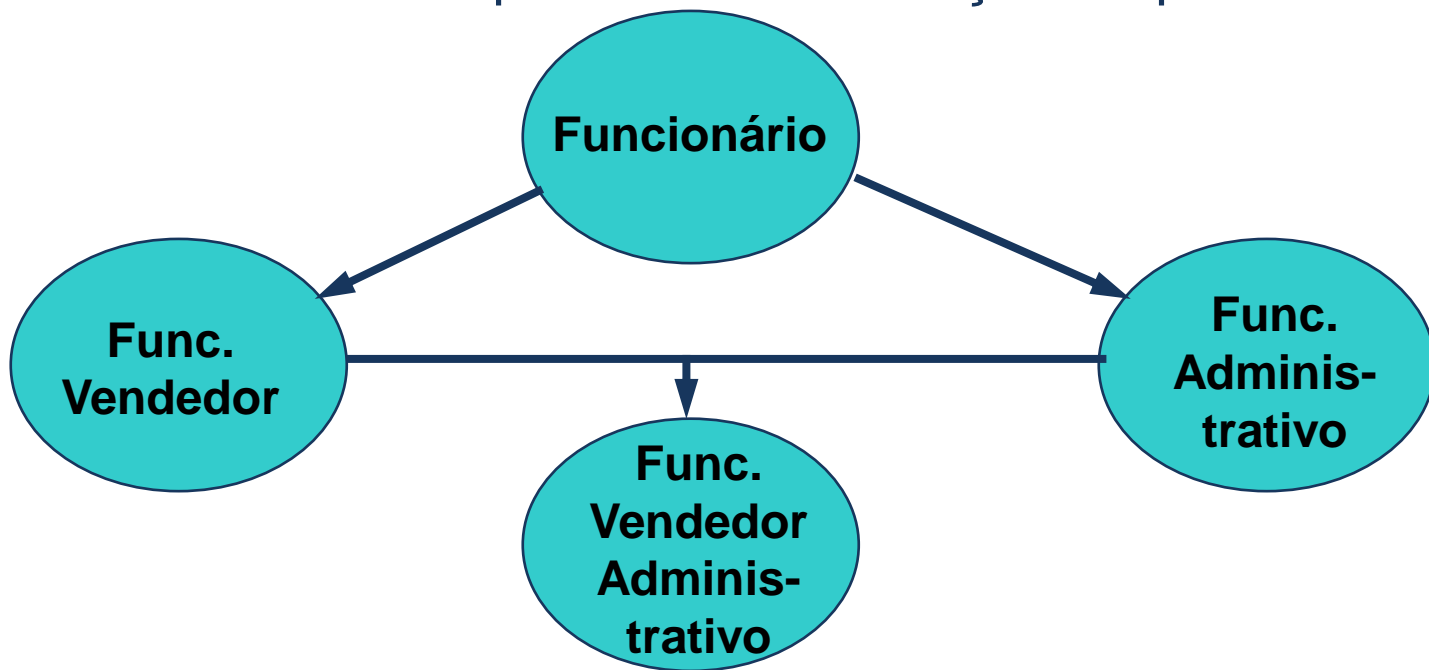
São usados quando os problemas são difíceis de serem entendidos ou alguns fatos são difíceis de serem inseridos na estrutura do problema.





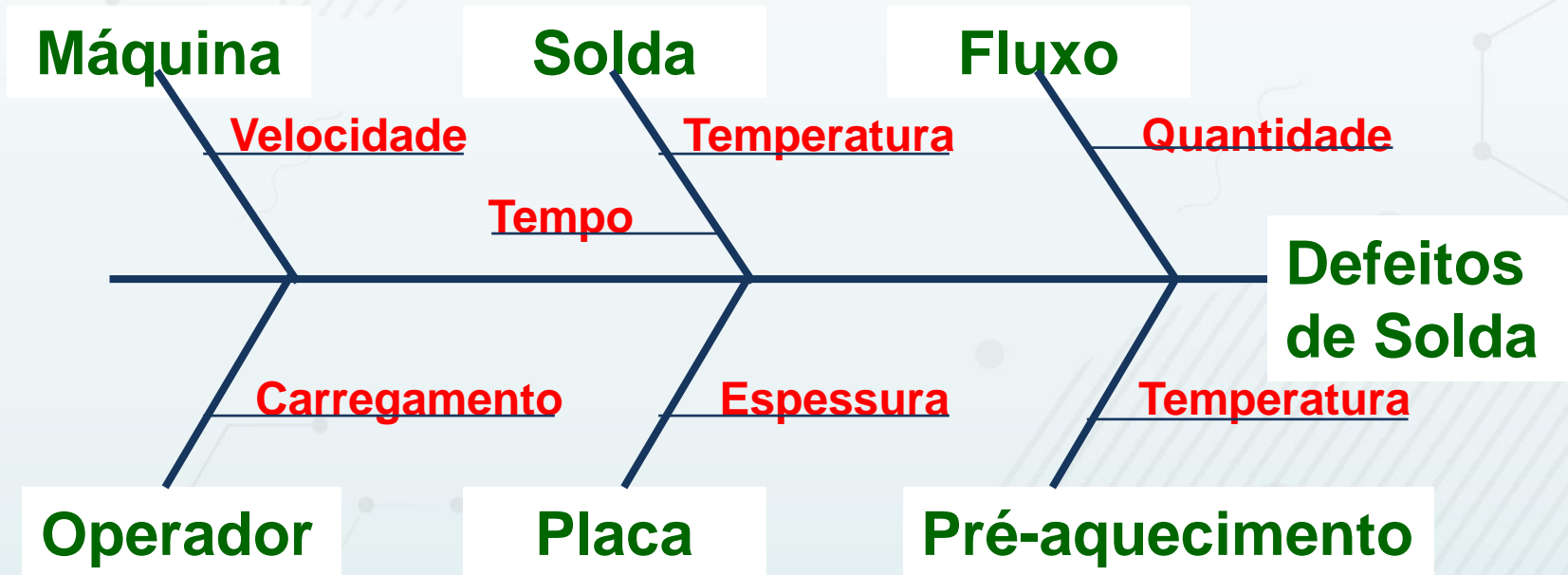
6) Resolução de Ambigüidades

A resolução de ambigüidades é útil para assegurar que onde existe uma interpretação com duplo sentido exista uma nova interpretação que pode ser considerada para a caracterização do problema.

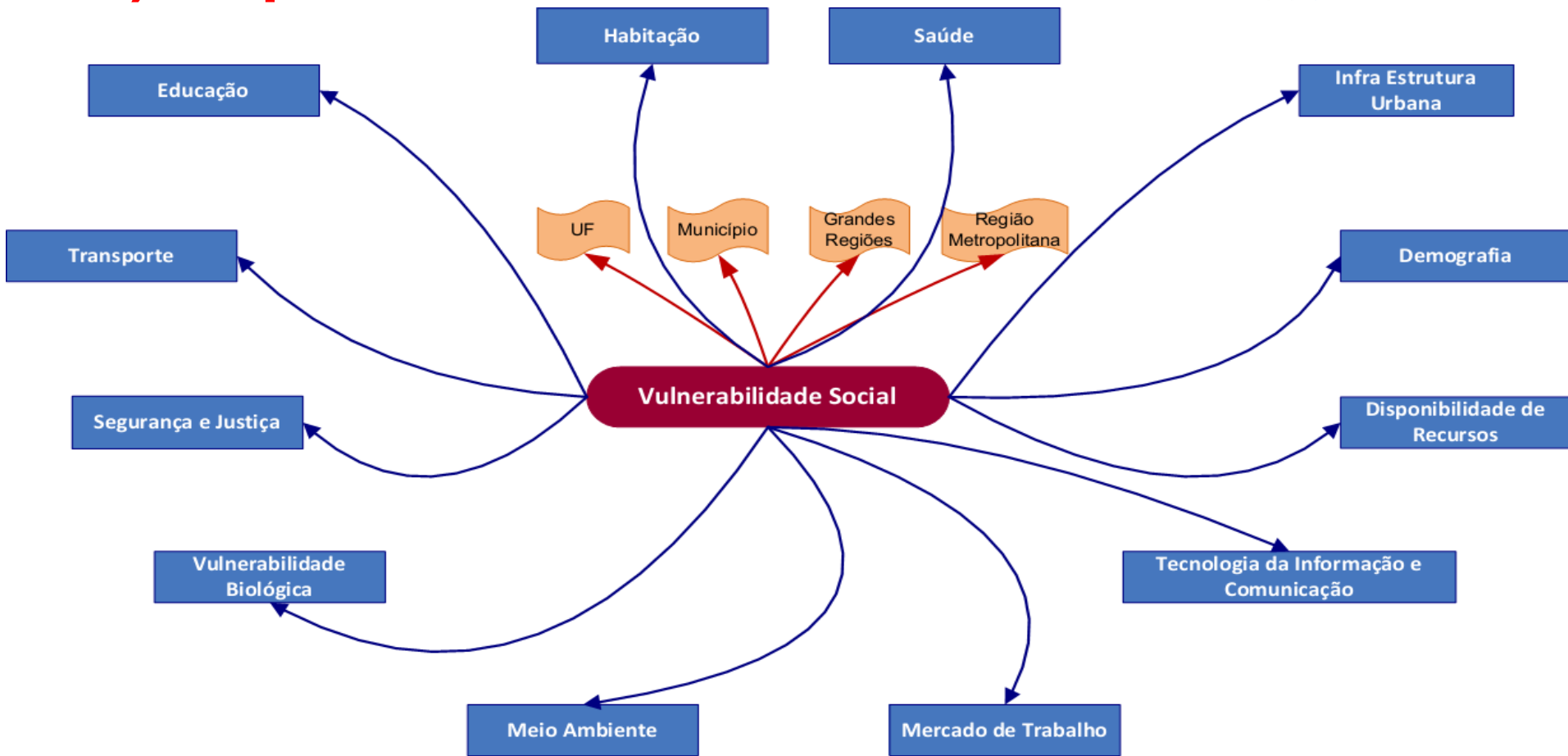


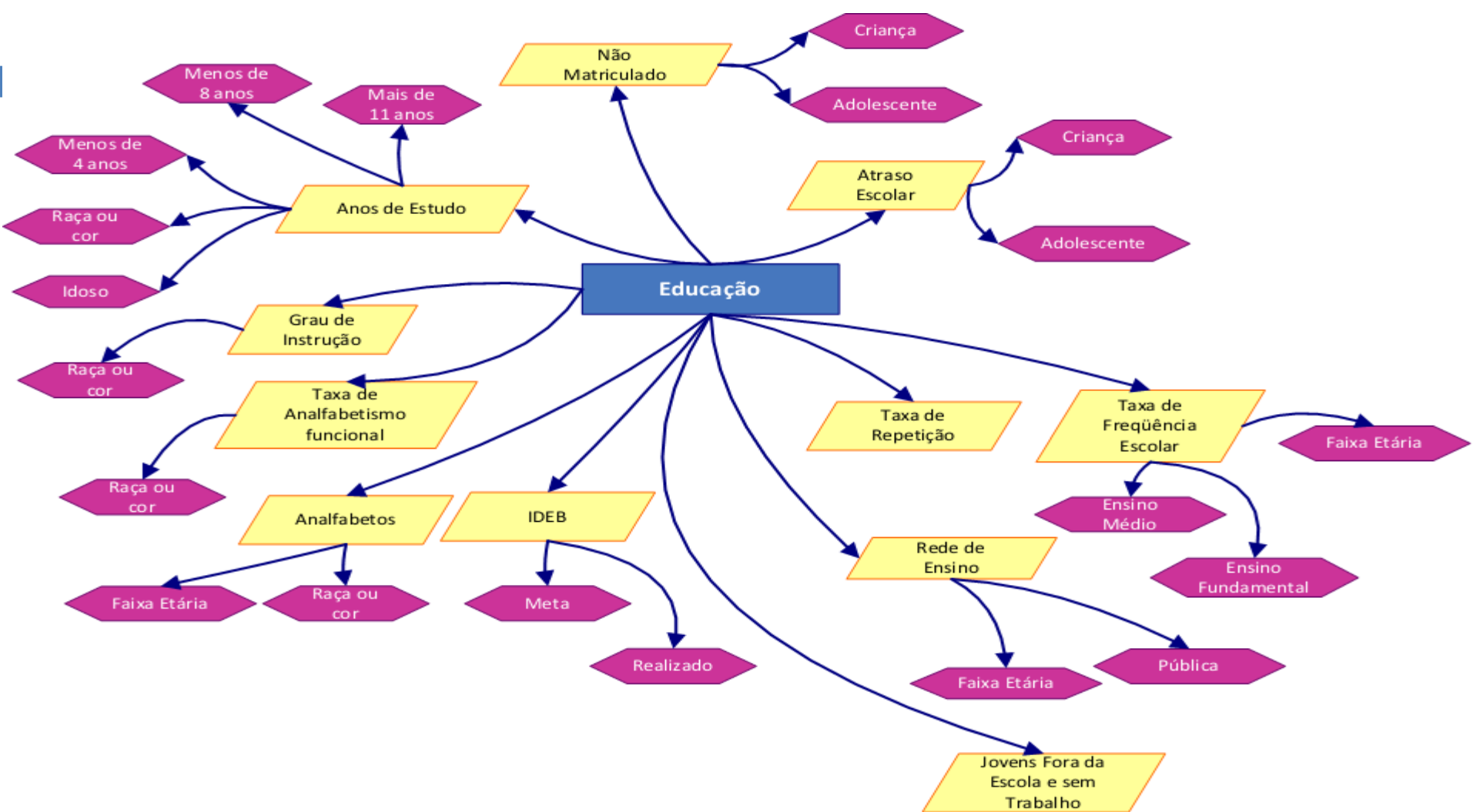
7) Diagramas de Causa-Efeito

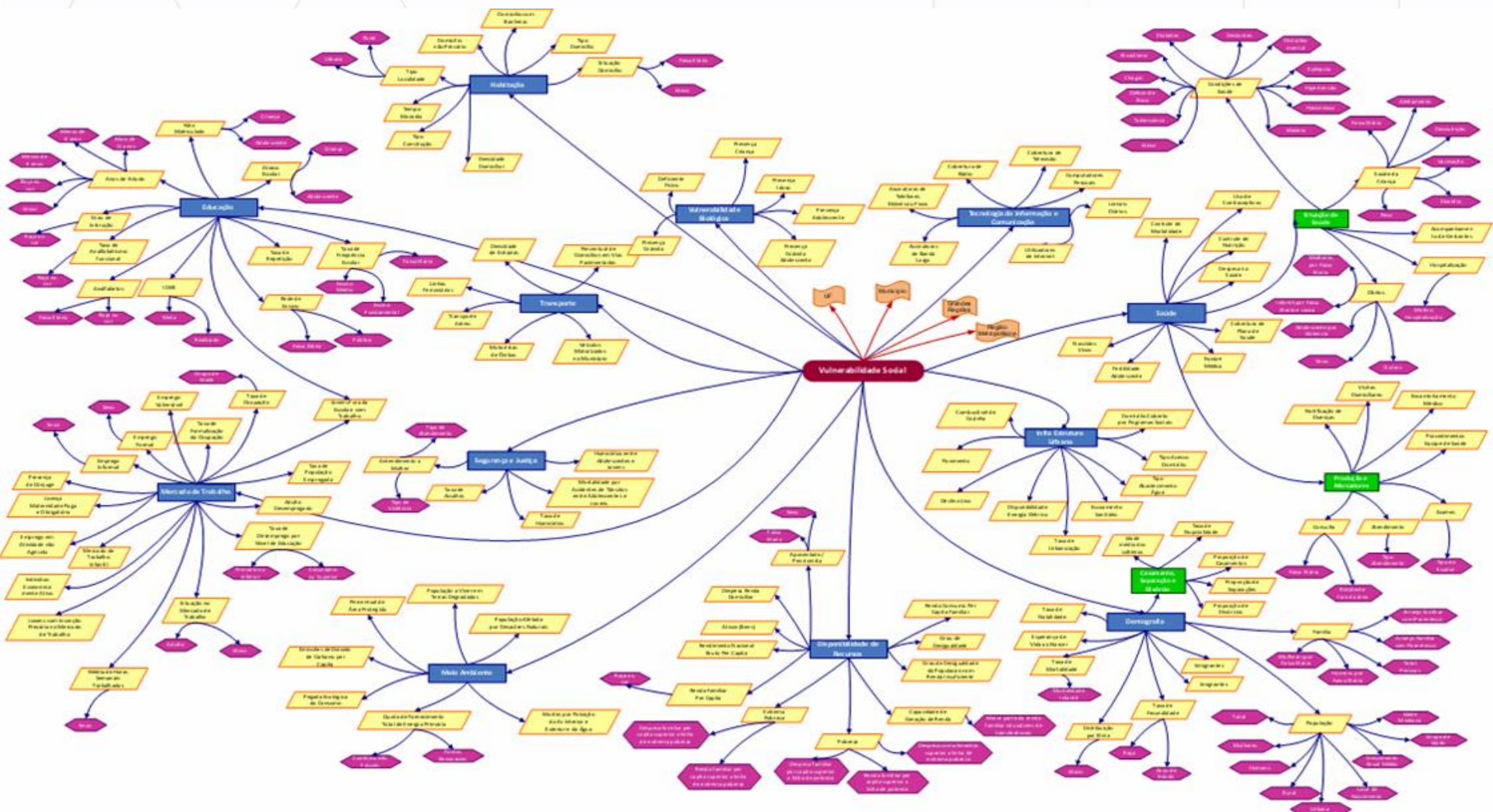
Diagrama de Ishikawa o de causa-efeito



8) Mapa conceitual



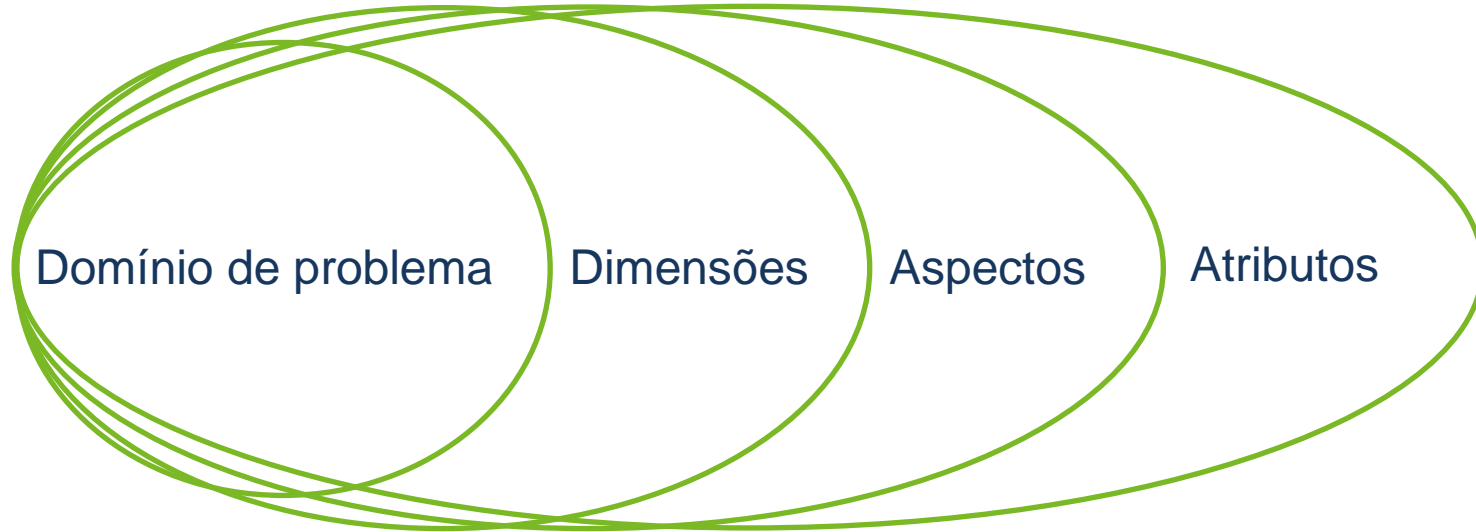








Modelagem Conceitual do domínio de Problema





Licap

Prática 3 – Para o problema considerado utilize a técnica dos mapas conceituais para representar as principais dimensões e aspectos que podem caracterizar o domínio de problema.



PUC Minas

Caracterização do domínio por meio de atributos



- Após o entendimento do domínio do problema é necessário identificar os atributos para compor o conjunto de dados para construção dos modelos.
- Cada atributo deve ser avaliada pelo analista de domínio e cientista de dados e selecionada de acordo com a sua relevância em relação ao problema.
- Deve ser identificado a fonte (base de dados) que poderá fornecer esses dados.
- É importante indicar, com auxílio do especialista de domínio se o atributo é: Muito Relevante; Relevante; Pouco Relevante; e Irrelevante (Fatos ou Julgamentos) em relação ao domínio do problema.

Exemplo - Acidente de trânsito

Imprudência

Efeito de álcool

Excessiva velocidade

Falha mecânica



Qual é o perfil do motorista de cada veículo?

Quais são os perfis dos acompanhantes dos veículos?

Quais eram as condições climáticas durante o acidente?

Atributos: {

FATOS

Dados pessoais: Idade, sexo, ocupação dos envolvidos;
Tipo de manobra prévia ao acidente dos envolvidos;
Manobra declarada por testemunhas para os envolvidos;
Nível alcoólico dos envolvidos;
Velocidade no momento do acidente dos veículos;
Tipo de falha mecânica dos veículos;

Tempo de carteira dos envolvidos;

JULGAMENTOS

Dados pessoais dos acompanhantes: idade, sexo, ocupação;
Condições climáticas durante o acidente;
Pontos na carteira dos motoristas;

...}

Documentando as variáveis:



Atributos: {

FATOS

Dados pessoais: Idade, sexo, ocupação dos envolvidos;
Tipo de manobra prévia ao acidente dos envolvidos;
Manobra declarada por testemunhas para os envolvidos;
Nível alcoólico dos envolvidos;
Velocidade no momento do acidente dos veículos;
Tipo de falha mecânica dos veículos;

Tempo de carteira dos envolvidos;

JULGAMENTOS

Dados pessoais dos acompanhantes: idade, sexo, ocupação;
Condições climáticas durante o acidente;
Pontos na carteira dos motoristas;
...}

- A etapa de entendimento e modelagem conceitual é considerada por alguns cientistas de dados como de **Grande Relevância** para a extração de conhecimento consistente, útil e não óbvio.
- A descrição do domínio por meio de atributos **permite avaliar a representatividade** da base de dados a ser utilizada.
- Pela caracterização por meio de atributos é possível **colocar restrições** ao conhecimento obtido quando a base de dados não é completamente representativa.



Licap

Prática 4 – Para o problema identificado na etapa anterior, caracterize o domínio de problema por meio de variáveis.



PUC Minas



Licap

Formação Cientista de Dados

Obrigado!



PUC Minas