

4. Geração e Extração de características

1. Binning:
 1. Reduz a complexidade das características
 2. Melhora a performance do modelo
 3. Pode ser usado em atributos numéricos ou categóricos
 4. Categorias "outros" para baixa cardinalidade

Binning

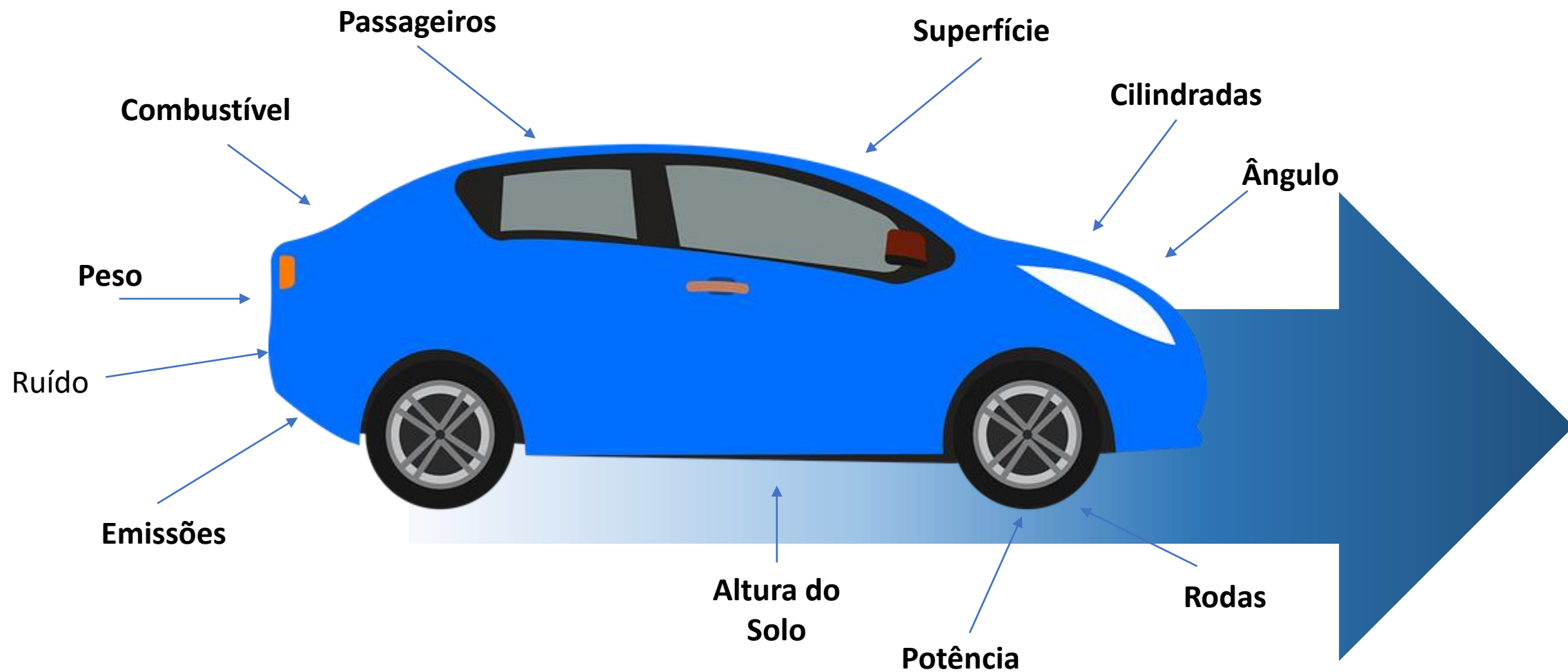
Eletrodomésticos	12
carro novo	234
carro usado	103
educação	50
mobilia/equipamento	181
negócios	97
obras	22
outros	12
qualificação	9
radio/tv	280



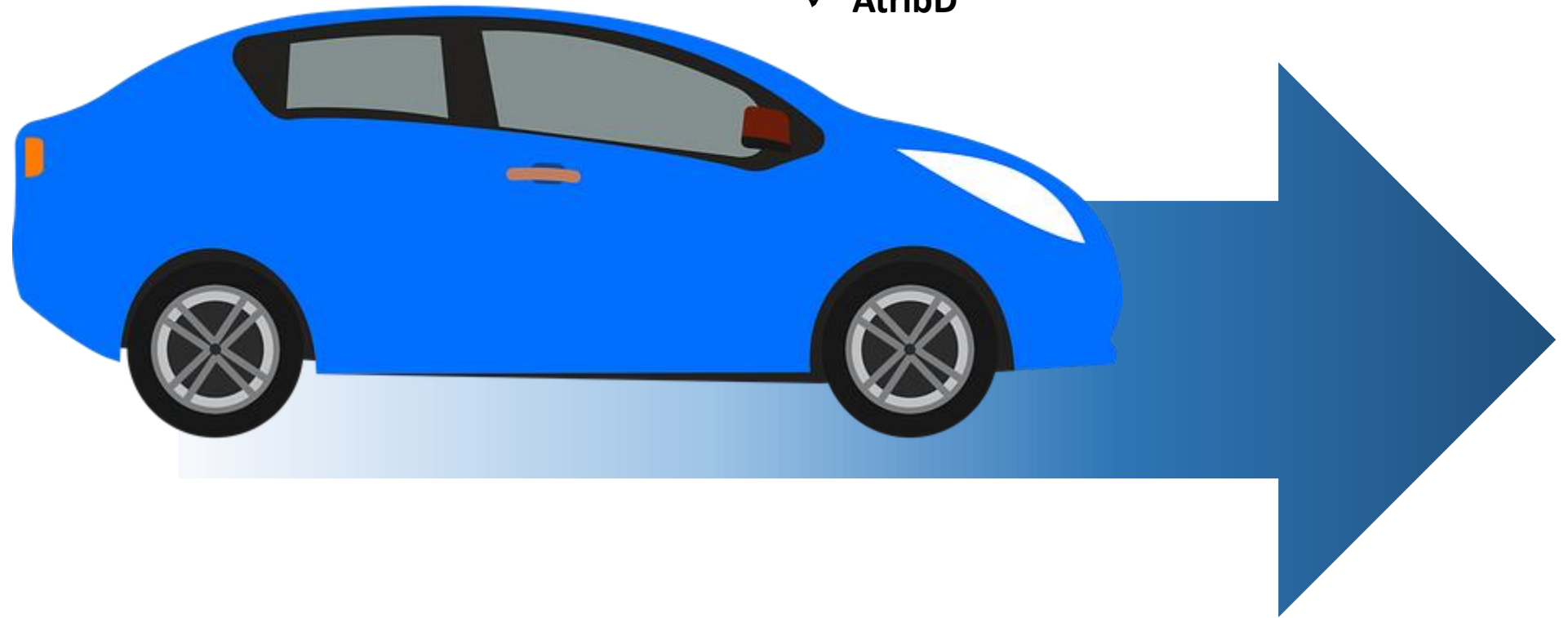
carro novo	234
carro usado	103
mobilia/equipamento	181
negócios	97
obras	22
outros	83
radio/tv	280

4. Geração e Extração de características: PCA

- Alta dimensionalidade:
 - Menor capacidade de generalização
- PCA: Redução de Dimensionalidade
- Cria atributos sintéticos, sem compreensão funcional
- Estes novos atributos buscam manter as características importantes dos dados
- Representação dos atributos originais: projeção
- Não permite avaliar importância de atributos e não mais representam o negocio analisado



- ✓ AtribA
- ✓ AtribB
- ✓ AtribC
- ✓ AtribD



PCA

- Redução da Dimensionalidade
- Treinamento do Modelo
- Testes e previsões devem ser aplicados com dados com o mesmo processo

4. Geração e Extração de Características

1. Exemplo: data
 1. Para o modelo é um "texto"
 2. Extraíndo "Mês", "Ano", "Feriado", "Dia da Semana" o modelo pode descobrir "padrões"

5. Seleção de Características Importantes

- Definir quais características, entre as "naturais" e as produzidas, são mais importantes para a performance do modelo