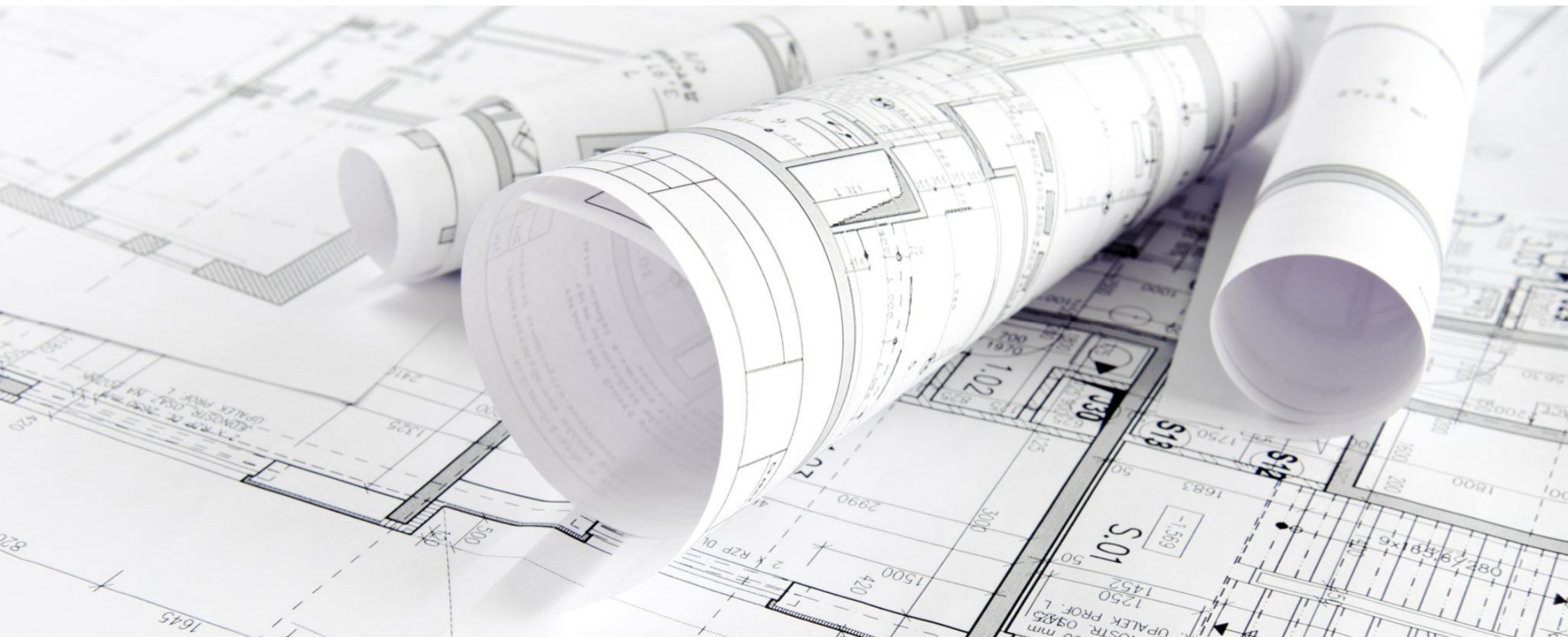


Engenharia de Atributos

ATRIBUTO:

Características do negócio que está sendo modelado



Engenharia de Atributos

Manipular características de forma a melhorar o modelo

Engenharia de Atributos



1. Pré-processamento

Tratamento de valores faltantes
Tratamento de outliers



2. Codificação de Categorias**



3. Dimensionamento de Características**



4. Geração e Extração de características

Binning
PCA
Decomposição de datas



5. Seleção de Características Importantes

1. Pré-processamento



Valores faltantes

- A maioria dos classificadores não tolera
- Investigar se não existe uma causa "natural"
- Pode-se excluir instâncias
- Pode-se completa-las com
 - Valores Numéricos:
 - Mediana
 - Valores Categóricos:
 - Moda
 - "Outros"

1. Pré-processamento



Tratamento de outliers

Não existe "regra geral"

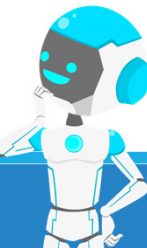
Regra conhecida:

- Exemplo: idade, teto salarial etc.
- 3 Desvios Padrão da média



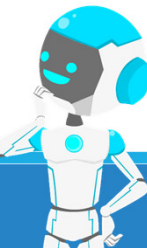
O que fazer?

Remover
"Normalizar"

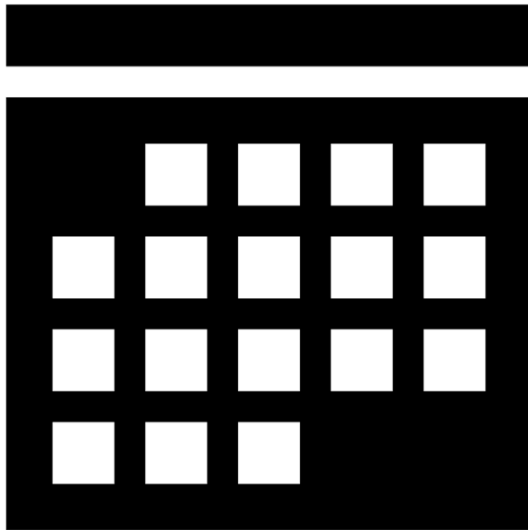


4. Geração e Extração de características

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



4. Geração 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"

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