



*Análise de Dados com Linguagem Python*

# Análise de Dados com Linguagem Python

## Weight of Evidence (WOE) e Information Value (IV)

Weight of Evidence (WOE) e Information Value (IV) são duas medidas usadas para analisar a força de uma variável em relação à variável alvo. Embora possa ser usada com variáveis numéricas e categóricas normalmente usamos com variáveis categóricas.

A fórmula para calcular o peso da evidência para qualquer característica é dada por:

$$\text{WOE} = \text{percentual de eventos} / \text{percentual de não eventos}$$

O peso da evidência (Weight of Evidence) indica o poder preditivo de uma única característica em relação a sua característica independente.

Se qualquer uma das categorias de um recurso tiver uma grande proporção de eventos em comparação com a proporção de não eventos, obteremos um alto valor de WOE que, por sua vez, diz que essa classe do recurso separa os eventos dos não eventos.

Por exemplo, considere a categoria C do recurso X. A proporção de eventos (0,16) é muito pequena em comparação com a proporção de não eventos (0,37). Isso implica que, se o valor do recurso X for C, é mais provável que o valor de destino seja 0 (não evento). O valor WOE apenas nos diz o quanto confiantes estamos de que o recurso nos ajudará a prever a probabilidade de um evento corretamente.

O valor WOE nos diz o poder preditivo de cada compartimento (classe) de um recurso. No entanto, um único valor que representa o poder preditivo de todo o recurso será útil na seleção de recursos. Isso é o IV (Information Value), dado pela fórmula:

$$\text{IV} = \text{sum}(\text{WOE} * (\text{percentual de eventos} - \text{percentual de não eventos}))$$

Interpretamos o resultado de IV assim:

Information Value, Poder de Previsão

- < 0.02, não deve ser usado para previsão
- 0.02 - 0.1, preditor fraco
- 0.1 - 0.3, preditor médio
- 0.3 - 0.5, preditor forte
- > 0.5, parece bom demais para ser verdade

O cálculo do WOE e IV nos ajuda a analisar vários pontos listados abaixo:

- O WOE ajuda a verificar a relação linear de um recurso com seu recurso dependente a ser usado no modelo.
- WOE é um bom método de transformação de variável para recursos contínuos e categóricos.
- WOE é melhor do que usar One-Hot-Encoding, pois esse método de transformação de variável não aumenta a complexidade do modelo.
- IV é uma boa medida do poder preditivo de um recurso e também ajuda a apontar o recurso que não é um bom preditor.

Embora WOE e IV sejam altamente úteis, sempre certifique-se de que sejam usados apenas quando a variável alvo é uma classe.