



Unyleya
EDUCACIONAL



Os valores NaN

O que são valores NaN

O termo NaN (Not a Number - Não é um número) é utilizado na computação para identificar valores irrepresentáveis ou a ausência de valores.

Um NaN não pode ser considerado como infinito e não pode ser utilizado para a realização de operações matemáticas.

O que são valores NaN

O ideal é que os valores NaN sejam tratados e removidos de dentro do conjunto antes de se iniciar o treinamento.

Vamos exemplificar esse tratamento de dados considerando que está sendo utilizado objetos das biblioteca Pandas, um Dataframe.

Removendo dados

A função `dropna` permite realizar a remoção de dados conforme a especificação passada para a função.

```
df.dropna(axis=0, how='any', thresh=None, subset=None, inplace=False)
```

- **axis:** `0` - para remover linhas com NaN; `1` - para remover toda coluna com valores NaN.
- **how:** `any` - se precisa de somente um elemento nulo para que a coluna seja removida; `all` - se todos os valores precisam ser NaN para a remoção da coluna;
- **thresh:** é um número inteiro para informar um valor mínimo para a coluna ser totalmente removida;
- **subset:** é definido um array para ser inserido onde a remoção está ocorrendo;
- **inplace:** `True` - para indicar que as alterações devem ser feitas diretamente no DataFrame.

Removendo dados

A função *isnull* serve para retornar a existência de valores nulos por atributo. Utilizando a função junto com o *sum* será retornado um valor inteiro para cada atributo indicando a quantidade de valores nulos encontrados nele.

```
df.isnull().sum()
```

Os valores NaN

Removendo dados



É possível atualizar o DataFrame de forma direta, onde os valores NaN do atributo idade será preenchido com o valor 999.

```
df.update(df['Idade'].fillna(999))
```



Obrigada!

Ana Laurentino