



PROCESSAMENTO DE LINGUAGEM NATURAL

One-Hot-Encoding

One-Hot

Encoding

Técnica usada em machine learning para converter variáveis categóricas em vetores binários.

One-Hot Encoding

Modelo	Cor	Falha
M1	Vermelho	Sim
M1	Vermelho	Sim
M2	Verde	Nao
M3	Azul	Sim

Modelo	Cor	Vermelho	Verde	Azul	Falha
M1	Vermelho	1	0	0	Sim
M1	Vermelho	1	0	0	Sim
M2	Verde	0	1	0	Nao
M3	Azul	0	0	1	Sim

Vantagens & Desvantagens



Intuitiva e simples de implementar



Tamanho do vetor aumenta com o tamanho do vocabulário



Representação esparsa do vetor de características



Não entende fora do corpus



Não fornece informação semântica

Vantagens & Desvantagens

- + Intuitiva e simples de implementar
 - Tamanho do vetor aumenta com o tamanho do vocabulário
-
- Representação esparsa do vetor de características
 - Não entende fora do corpus
 - Não fornece informação semântica

Vantagens & Desvantagens

- + Intuitiva e simples de implementar
 - Tamanho do vetor aumenta com o tamanho do vocabulário
 - Representação esparsa do vetor de características
-
- Não entende fora do corpus
 - Não fornece informação semântica

Vantagens & Desvantagens

- + Intuitiva e simples de implementar
 - Tamanho do vetor aumenta com o tamanho do vocabulário
 - Representação esparsa do vetor de características
 - Não entende fora do corpus
-
- Não fornece informação semântica

Vantagens & Desvantagens

- + Intuitiva e simples de implementar
 - Tamanho do vetor aumenta com o tamanho do vocabulário
 - Representação esparsa do vetor de características
 - Não entende fora do corpus
 - Não fornece informação semântica
-


```
import pandas as pd
```

```
pd.get_dummies(df, columns=[col1,..colN])
```

pandas



PUC Minas
Virtual