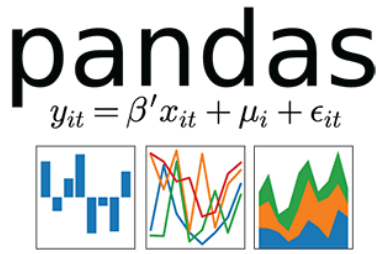
One-hot encoding

Python implementation

One-hot encoding with open source



pd.get_dummies()

- Does not store / learn seen categories.
- Does not handle unseen categories out-of-the-box.
- May return different number of dummies in train and test sets.

One-hot encoding with open source

Feature-Engine

Category Encoders

- OneHotEncoder()

- **fit()** ➔ learns categories in train set.
- **transform()** ➔ creates dummies for learned categories.
- Option to encode unseen categories as zeroes in all dummies, or raise an error.

One-hot encoding with open source

Feature-Engine

Category Encoders

- OneHotEncoder()
- Automatically detect categorical features.
- Option to indicate which features to encode.
  - Feature-engine, param variables.
  - Category encoders, param cols.
- Return dataframes with categorical variables encoded + rest of the variables

One-hot encoding with open source

- OneHotEncoder()
- Need ColumnTransformer() to select variables to encode.
- Returns Numpy arrays.
- Next release of Scikit-learn will have option to return dataframes.

Accompanying Jupyter Notebooks

1. Demo the method with the 3 libraries.