

FEATURE ENGINEERING

The process of creating new features from the existing features in a dataset. The performance of a predictive model is heavily dependent on the quality of the features in the dataset used to train that model creating new features that help in providing ,more information to the model about the target variable leads to improved performance.

Featuretools is an open source library for performing automated feature engineering.

It has 3 major components:

- i. Entities: considered a representation of a Pandas DataFrame.
- ii. Deep Feature Synthesis: a feature engineering method, enables creation of new features from a single or multiple DataFrames.
- iii. Feature Primitive : an operation applied to data to create a feature. They fall into to categories, Aggregation(mean,min,max or standard deviation) and Transformation (eg. extracting a the day from dates or difference between two columns)