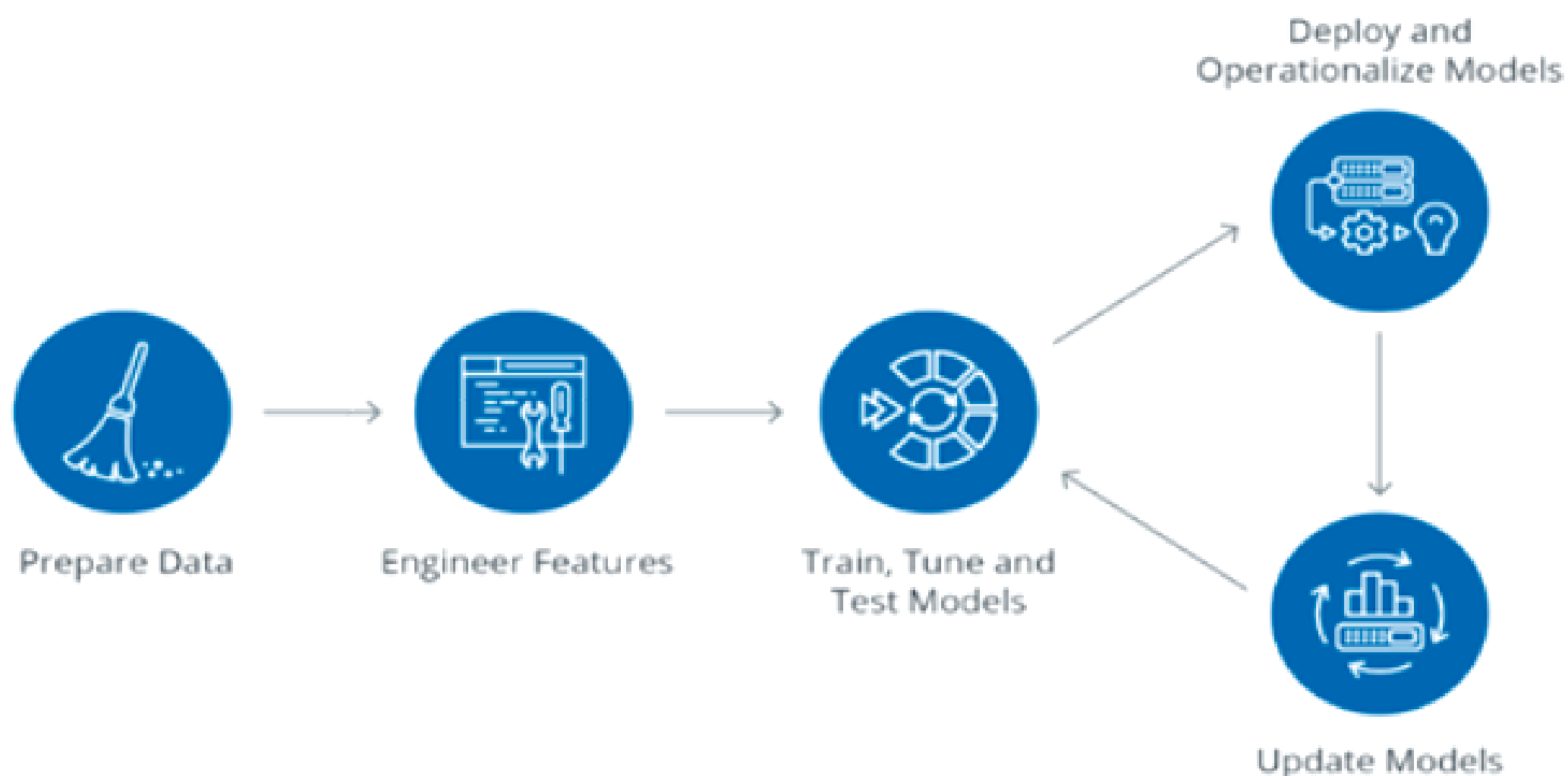


Feature Engineering for Machine Learning



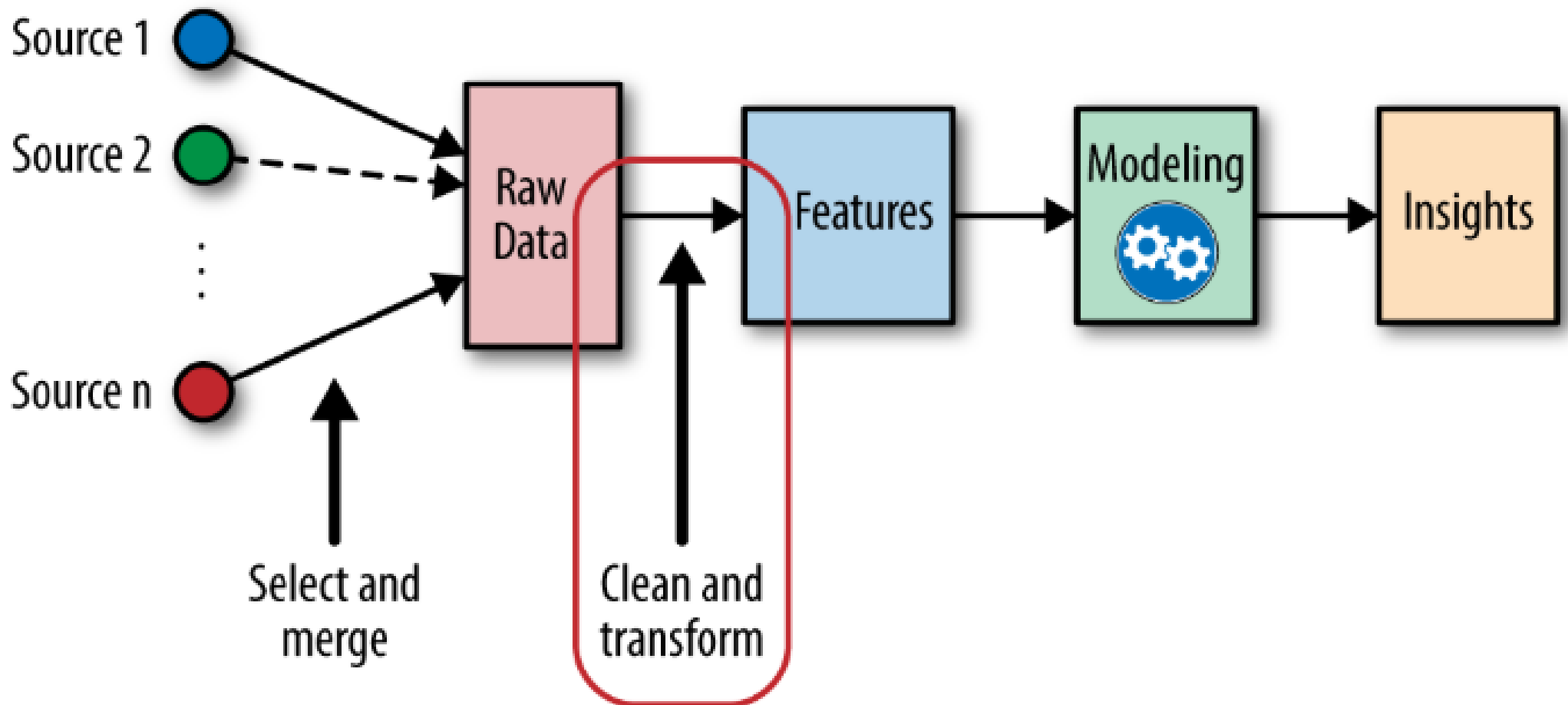
Imagine this: You've got a giant toy box and you want to find all the blue toys. Just by looking at the box you can't decide. You gotta get creative and figure out some special clues, like the shape or the texture or the sound they make.

That's where **feature engineering** comes in for machine learning.

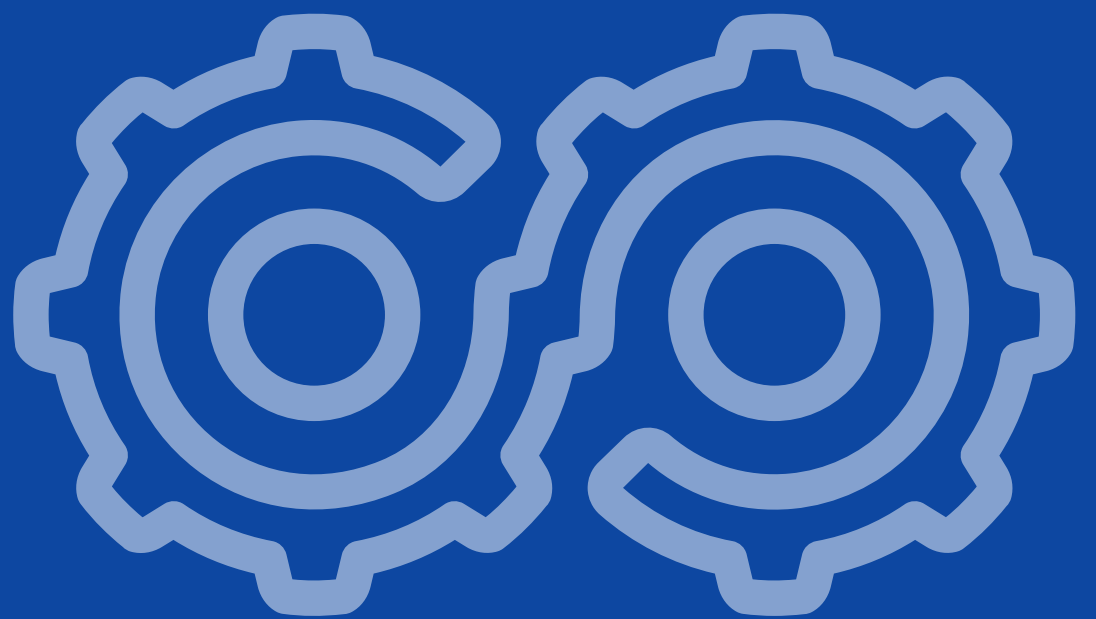


What is Feature Engineering?

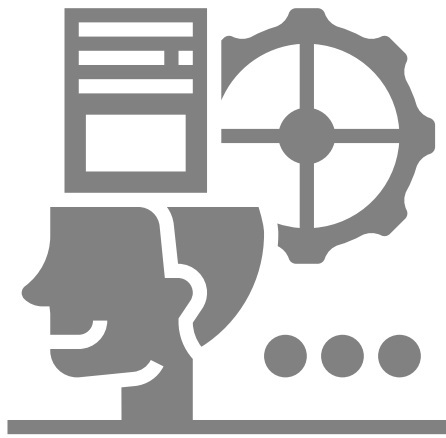
Feature Engineering is the process of **extracting** and **organizing** the important **features** from raw data in such a way that it fits the purpose of the machine learning model.



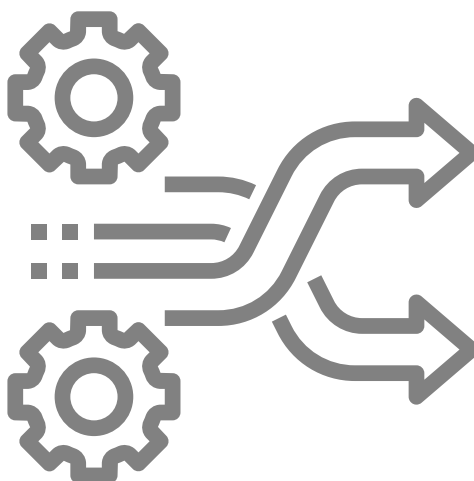
Process of Feature Engineering in Machine Learning



Feature Creation: Feature creation is finding the most useful variables to be used in a predictive model. It involves mixing existing features using addition, subtraction, and ratio to create new features with great flexibility.



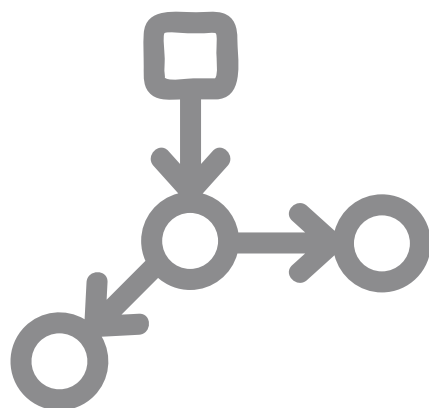
Transformations: The transformation step adjusts predictor variables to improve model accuracy and performance by ensuring flexibility, consistent scaling, and avoiding computational errors.



Feature Extraction: Feature extraction is an automated feature engineering process that generates new variables by extracting them from the raw data. The main aim of this step is to reduce the volume of data so that it can be easily used and managed for data modelling.



Feature Selection: Feature selection is a way of selecting the subset of the most relevant features from the original features set by removing the redundant, irrelevant, or noisy features.



Feature Engineering Techniques



1. Imputation

Imputation in machine learning refers to the process of filling in missing data with estimated values. This is done to ensure that the data is complete and can be used for analysis and modeling.



2. Handling Outliers

Outliers are those data points that are significantly deviated from the rest of the dataset. They have a greatly impact on the analysis or model, and can lead to incorrect results if not handled properly.



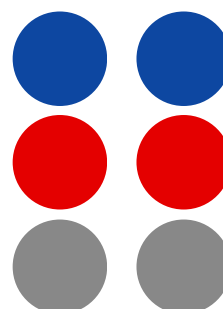
3. Log transform

Log transform makes skewed data easier to handle, and they approximate normal distributions after transformation. By normalizing magnitude differences, it also reduces the effects of outliers on the data.



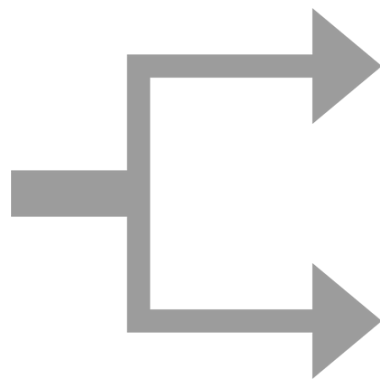
4. Binning

Binning is a process of grouping continuous data into discrete intervals or categories. This is useful for simplifying complex datasets or for reducing the impact of outliers.



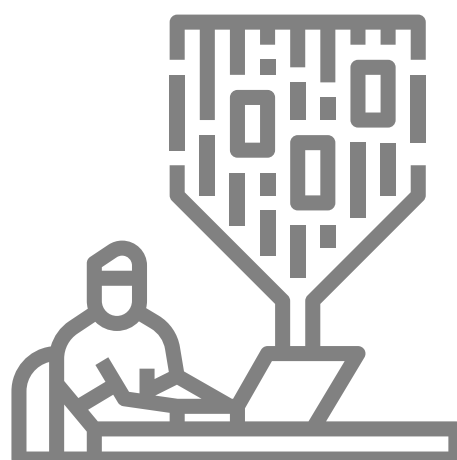
5. Feature Split

Feature split is the process of splitting features intimately into two or more parts and performing to make new features. This technique helps the algorithms to better understand and learn the patterns in the dataset.



6. Encoding

It is a technique that converts the categorical data in a form so that they can be easily understood by ML algorithms and hence can make a good prediction.



That's a wrap.
Was this post
Helpful?

Follow us for more!

