OneModel Academy

One AI

Feature Engineering

Machine Learning

Hayley Bresina
Client Enablement Specialist

20 May 2024

Topics Covered

- What a feature is & the different types of features
- What feature engineering is & its significance in machine learning
- The process & techniques of feature engineering
  - Understanding the data
  - Data cleaning
  - Feature selection, transformation, & creation
  - Refining & improving features
- Using Exploratory Data Analysis (EDA) for feature exploration & engineering

Learning Outcomes

You will:

- Have the ability to clearly explain what a feature is & distinguish between the different types
- Understand the importance of feature engineering & its process & techniques such as data cleaning, feature selection, and transformation
- Be confident in conducting exploratory data analysis to understand dataset characteristics & inform feature engineering processes

Feature Overview

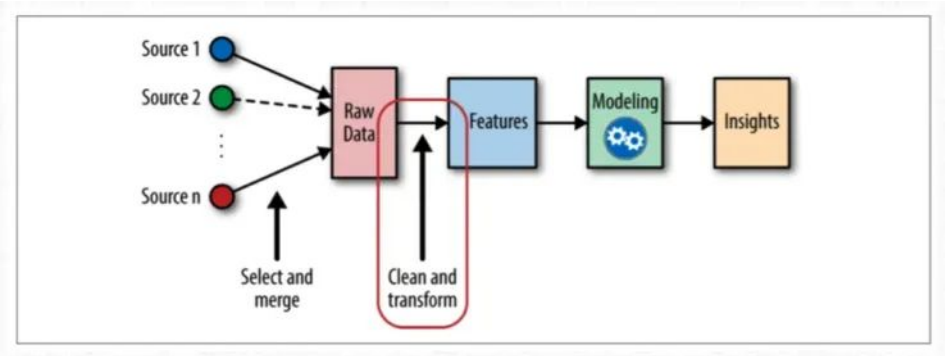
Features

- **Feature:** individual measurable property or characteristic of the data used as input for a machine learning model to make predictions
  - Also known as input variables, predictors/predictor variables, columns, or attributes
- Types of Features
  - Numerical features (includes dates)
  - Categorical features
    - Ordinal or nominal
    - Binary
  - Generated / derived features

Feature Engineering Overview

Feature Engineering

- **Feature engineering:** process of extracting & transforming raw data into a format that is consumable by ML models



  - Improves performance
  - Improves generalization on unseen, new data
  - Results in more interpretable models

Feature Engineering Process & Techniques

Process & Techniques for Feature Engineering

1. **Understanding the data**
   - Developing an understanding of the model dataset & problem domain
   - Techniques:
     - Regular communication with stakeholders
     - Data exploration (EDA)

2. **Data Cleaning**
   - Involves handling missing & constant values, checking for correlations, & detecting data leakage
   - Techniques:
     - Null-filling (imputation)
     - Removing outliers
     - Correcting data errors

Process & Techniques

3. **Feature Selection**
    - Selecting the most relevant features; discarding irrelevant or reductant features
        - Reduces dimensionality & prevents overfitting
        - Less is more!
    - **Techniques**:
        - Univariate feature selection
        - Recursive feature elimination (rfe)
        - Feature importance

Process & Techniques

4. **Feature Transformation**
    - Transforming the features to make them more suitable to be read by the ML algorithm

    - **Techniques**:
        - Scaling
        - One Hot Encoding
        - Log Transformation
        - Power Transformation

Process & Techniques

5. **Feature Creation**
    - Creating new features from existing ones to capture additional information or relationships in the data
    - **Techniques**:
        - Generative attributes
        - Binning
        - Interaction terms
        - Polynomial features
6. **Iterate & Improve**
    - Iterative process involving improving features based on feedback from performance & fine-tuning the feature pipeline

EDA & Feature Engineeing

Exploratory Data Analysis & Feature Engineering

- **Exploratory Data Analysis** (EDA) is used to analyze model datasets & summarize their main characteristics, often employing data visualization methods
  - One AI creates an EDA report for every model iteration successfully run
  - Provides rich exploration into variables & details on feature transformation, missing data, correlations
  - Should be used throughout the feature engineering process

OneModel Academy

Thanks for watching!

OneAI