

Feature Engineering

Introduction

Feature engineering is the process of selecting, creating, and transforming features from raw data to improve the performance of machine learning models. Good features help models learn better patterns, while poor features can reduce accuracy.

What is a Feature?

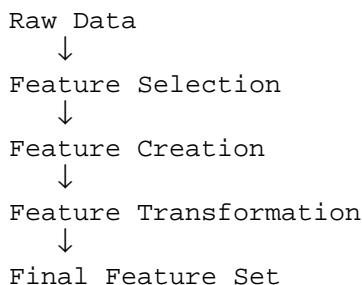
A feature is an individual measurable property or characteristic of the data. In simple terms, features are the input variables used by a machine learning model.

Example: In house price prediction, features can be size, location, number of rooms, and age of the house.

Why Feature Engineering is Important

- 1 Improves model accuracy
- 2 Reduces model complexity
- 3 Helps models learn meaningful patterns
- 4 Plays a key role in real-world ML success

Feature Engineering Workflow



Feature Selection

Feature selection involves choosing the most relevant features and removing irrelevant or redundant ones.

- 1 Filter methods (correlation, variance)
- 2 Wrapper methods (forward selection)
- 3 Embedded methods (Lasso, Decision Tree)

Feature Creation

Feature creation involves generating new features from existing data to better represent the underlying problem.

Example: Creating 'price per square foot' from price and area.

Feature Transformation

Feature transformation modifies features to make them more suitable for machine learning algorithms.

- 1 Normalization and standardization

- 2 Log transformation
- 3 Encoding categorical variables

Handling Categorical Features

- 1 Label Encoding
- 2 One-Hot Encoding
- 3 Target Encoding

City → Delhi:0, Mumbai:1, Pune:2

Feature Scaling in Feature Engineering

Feature scaling ensures that all features contribute equally to model learning.

- 1 Min-Max Scaling
- 2 Standardization (Z-score)

Real-Life Example

In a customer churn prediction system, raw data such as login count, subscription duration, and payment history are converted into meaningful features to help the model predict churn accurately.

Summary

Feature engineering is a critical step that directly impacts the performance of machine learning models. Well-designed features often matter more than choosing complex algorithms.