Feature Engineering

Thursday, 5 February 2026 11:48 PM

What is Feature Engineering?

Feature Engineering is the process of using domain knowledge to extract, transform, or create new variables (features) from raw data that make machine learning algorithms work better.

While algorithms learn patterns, they cannot understand context or nuance inherently. Feature engineering translates raw data into a language the model can understand more effectively.

The Feature Engineering Workflow

1. **Understand the Data:** Analyze the raw data and identify what information is useful.
2. **Clean the Data:** Handle missing values, outliers, and incorrect data types.
3. **Transform Data:** Change data formats to be suitable for algorithms (e.g., scaling numerical data).
4. **Create New Features:** Combine existing features or extract information to create more predictive variables.
5. **Select Features:** Choose the most relevant features to prevent overfitting and improve speed.

Common Techniques with Examples

Technique	Description	Example
Imputation	Handling missing data.	Replacing missing Age values with the median age.
Encoding	Converting categorical data to numbers.	Turning [Red, Blue, Green] into [0, 1, 2].
Scaling	Normalizing numerical data to a similar range.	Scaling Income (0-100k) and Age (0-100) to a scale of 0-1.
Interaction	Combining two features to create a new one.	Creating House Price Per Square Foot from Total Price and Square Footage.
Extraction	Pulling specific information out of raw data.	Extracting Day of Week from a timestamp (2023-10-27 10:00:00).

Why is it Important?

- **Improves Model Accuracy:** High-quality features make the underlying patterns clearer to the model.
- **Reduces Complexity:** Good features allow you to use simpler, more interpretable algorithms while maintaining high performance.
- **Handles Raw Data Limits:** Algorithms cannot directly use raw data like text, images, or timestamps; feature engineering bridges this gap.