Feature Engineering

Zeham Management Technologies BootCamp

by SDAIA

July 29th, 2024

Let’s practices

NOTEBOOK:

2 - Feature Engineering

LAB/Used_cars.ipynb

DataSets:

2 - Feature Engineering

LAB/Datasets/Used_cars.csv

Introduction to Feature Engineering

Let’s start together…

Agenda

- creating new feature from Raw data
- Transforming existing feature
- Selecting the most relevant feature

Feature Engineering

- **Raw Data:** This is the original dataset with various features and the matching target variable.
- **Feature Engineering:** This is the process in which important features from the raw data are selected, created, transformed, and dimensionally reduced.
- **Feature Transformation:** After features are selected, they are manipulated and transformed. This is a step where new features can be created, dimensions are reduced, and categorical values are handled.
- **Model Training:** In this stage, the data with engineered features is fit into the model as input. The model then learns patterns and relationships between the engineered features and the target variable. source

Feature Engineering

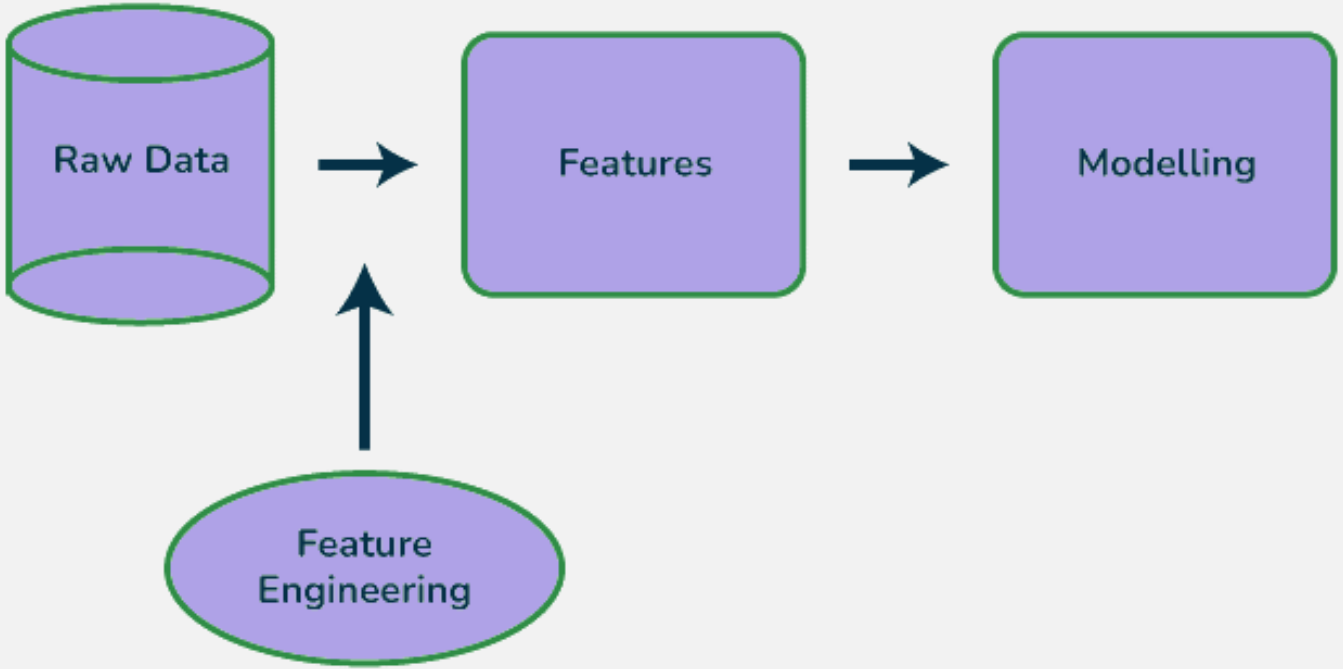
- **Model Evaluation:** This is a crucial step in Machine Learning where the model generalization and prediction accuracy are evaluated. The evaluation helps identify if the model has overfitting or underfitting issues.
- **Model Deployment:** This is the final step where the model is realized to perform real-world tasks. source

Feature Engineering

Feature engineering is the process of **transforming raw data into features that are suitable for machine learning models**. In other words, it is the process of selecting, extracting, and transforming the most relevant features from the available data to build more accurate and efficient machine learning models.

The success of machine learning models heavily depends on the quality of the features used to train them. Feature engineering involves a set of techniques that enable us to create new features by combining or transforming the existing ones. These techniques help to highlight the most important patterns and relationships in the data, which in turn helps the machine learning model to learn from the data more effectively. source

Feature Engineering



Feature Engineering

In the context of machine learning, a feature is an individual measurable property or characteristic of a data point that is used as input for a machine learning algorithm. Features can be numerical, categorical, or text-based, and they represent different aspects of the data that are relevant to the problem at hand. source

- For example, in a dataset of housing prices, features could include the number of bedrooms, the square footage, the location, and the age of the property. In a dataset of customer demographics, features could include age, gender, income level, and occupation.
- The choice and quality of features are critical in machine learning, as they can greatly impact the accuracy and performance of the model.

Why do we need feature engineering in Machine learning

- **Improve User Experience:** The primary reason we engineer features is to enhance the user experience of a product or service. By adding new features, we can make the product more intuitive, efficient, and user-friendly, which can increase user satisfaction and engagement.
- **Competitive Advantage:** Another reason we engineer features is to gain a competitive advantage in the marketplace. By offering unique and innovative features, we can differentiate our product from competitors and attract more customers.
- **Meet Customer Needs:** We engineer features to meet the evolving needs of customers. By analysing user feedback, market trends, and customer behaviour, we can identify areas where new features could enhance the product's value and meet customer needs. source

Why do we need feature engineering in Machine learning

- **Increase Revenue:** Features can also be engineered to generate more revenue. For example, a new feature that streamlines the checkout process can increase sales, or a feature that provides additional functionality could lead to more upsells or cross-sells.
- **Future-Proofing:** Engineering features can also be done to future-proof a product or service. By anticipating future trends and potential customer needs, we can develop features that ensure the product remains relevant and useful in the long term. source

Feature Engineering Process

The main goal of feature engineering is to provide the model with important features of data that will help the model learn and make accurate predictions.

Feature engineering is an iterative process involving experimentation, model evaluation, and refinement to find the best feature set. the process are :

source

Creating new feature from Raw data

Feature Engineering Process

- **Feature Creation**
- Feature Transformation
- Feature Extraction
- Feature Selection

Feature Creation

Feature creation is the process of developing new features from existing ones to capture complicated relationships and patterns in data.
Let’s discuss the ways in which features are created.

- **Interaction features:** You can create new features by capturing the combined influence of two or more features.
- **Arithmetic operations:** New features can be created by doing simple calculations using arithmetic operations. For example, you can calculate the sum of values from existing features. That way, you’ll have created a new feature with the sum of the original values.
- **Feature binning:** You can create new features by separating continuous features into discrete bins or intervals. source

Feature Creation

- **Polynomial features:** You can create features by raising values in existing features to power. This technique is beneficial for linear algorithms because it captures nonlinear correlations between characteristics and target variables.

- **Aggregate features:** You can create new features by using functions to summarize or aggregate data from numerous entries per entity. For example, you can calculate the average of values across different entities.

- **Feature scaling and normalizing:** New features can be created by modifying the distribution of features for specific algorithms. Techniques like z-score normalization or min-max scaling are used to perform feature scaling and normalization. source

Creating new feature from Raw data

This is an example of creating new feature :

	Gender	Height	Weight	Index	Height_M
0	Male	174	96	4	1.74
1	Male	189	87	2	1.89
2	Female	185	110	4	1.85
3	Female	195	104	3	1.95
4	Male	149	61	3	1.49

	Gender	Height	Weight	Index	Height_M	BMI
0	Male	174	96	4	1.74	31.708284
1	Male	189	87	2	1.89	24.355421
2	Female	185	110	4	1.85	32.140248
3	Female	195	104	3	1.95	27.350427
4	Male	149	61	3	1.49	27.476240

Let’s practice

Dataset :

500_Person_Gender_Height_Weight_Index.csv

Transforming existing feature

Feature Engineering Process

- Feature Creation
- **Feature Transformation**
- Feature Extraction
- Feature Selection

Feature Transformation

Feature transformation is a step-in feature engineering that involves utilizing mathematical approaches to change the values of features to improve the performance of machine learning models.

It serves numerous functions, including enhancing algorithm convergence during training, matching data distributions with the assumptions of particular algorithms, and permitting a better fit between the model and the data. source

Method for Feature Transformation

- **Normalization:** Features can be transformed by normalizing them. This can be done by decreasing the range of feature values to a preset range, typically between 0 and 1.
- **Standardization:** Features can be standardized by converting feature values to have a mean of 0 and a standard deviation, ensuring that all features contribute equally.
- **Logarithmic scaling:** Features can be transformed by taking the logarithm of feature values, allowing it to handle skewed data distributions and reduce the impact of big data sets.

source

Method for Feature Transformation

In feature transformation, other mathematical operations such as square root, exponentiation, and division can also be applied.

source

Feature Engineering Process

- Feature Creation
- Feature Transformation
- **Feature Extraction**
- Feature Selection

Feature Extraction

Feature extraction is an important step in feature engineering since it aims to minimize the dimensionality of data by translating it into a lower-dimensional representation while maintaining useful information.
This technique is useful for high-dimensional data or compact representations that retain crucial data features. Feature extraction techniques reduce complexity and improve visualization. Reducing dimensions may result in interpretability loss. What determines if the feature can be reduced is the nature of the data, the problem, and the trade-offs.

Selecting the most relevant feature

Feature Engineering Process

- Feature Creation
- Feature Transformation
- Feature Extraction
- **Feature Selection**

Types of feature selection

- **Filter methods:** Examine the importance of features independently of any given model, employing statistical measures to rank or score features. Wrapper approaches include
- **Wrapper methods:** Involve training and evaluating the model multiple times with different feature subsets
- **Embedded methods:** Incorporate feature selection into the model-building process.

Features in a dataset are not equally valuable. Selecting unnecessary or redundant features might result in overfitting, increased computational cost, and decreased model interpretability.

Let’s practices

NOTEBOOK:

2 - Feature Engineering

LAB/Creating new Feature from Raw Data.ipynb

DataSets:

2 - Feature Engineering

LAB/Datasets/500_Person_Gender_Height_Weight_Index.csv

Thank you!

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority