

## Data Science | 30 Days of Machine Learning | Day - 21

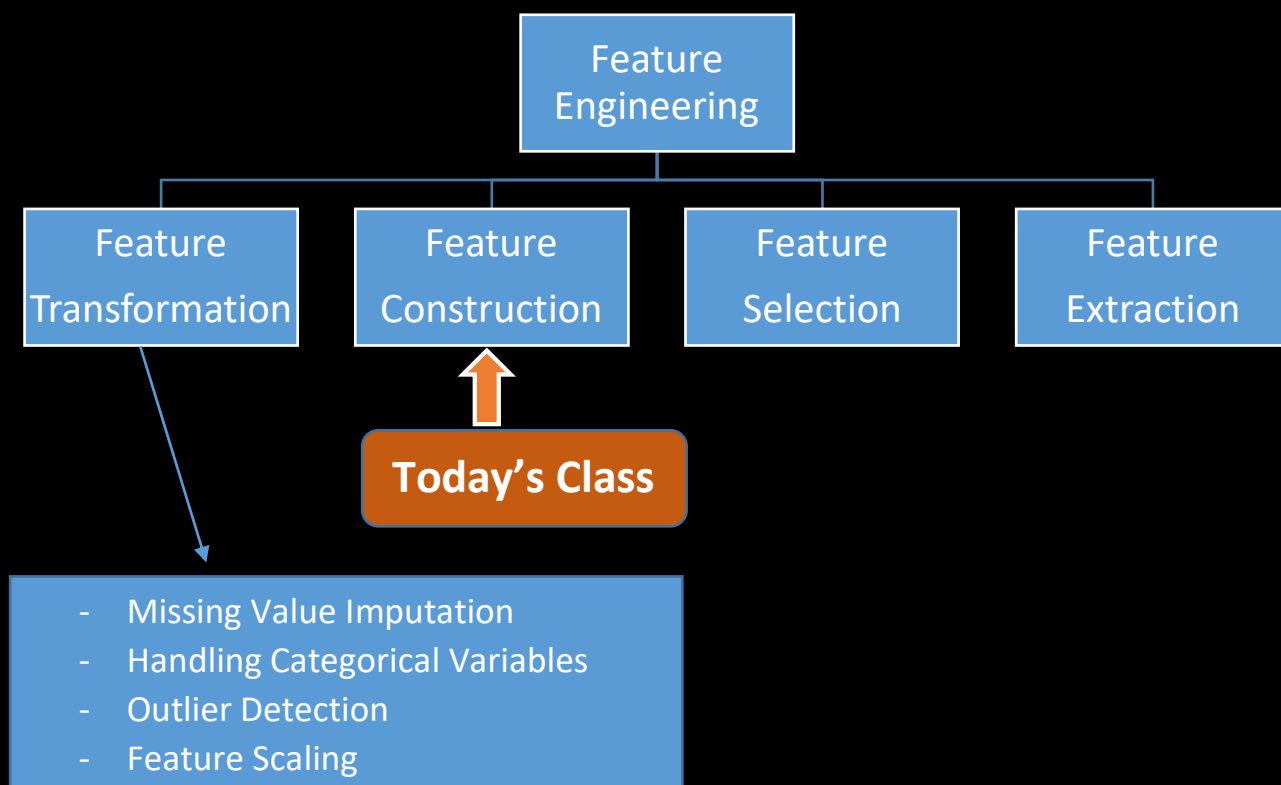
Educator Name: Nishant Dhote  
Support Team: +91-7880-113-112

### ----Today Topics | Day 21----

#### Feature Construction & Feature Splitting

----

- What is feature construction in machine learning?
  - What is feature splitting in machine learning?
  - What is a purpose of feature construction and splitting?
  - Technique used for feature construction & feature splitting?
  - Uses and advantages for construction and splitting?
  - What result shown in the output?
- Dataset Link GitHub: [https://github.com/TheiScale/30\\_Days\\_Machine\\_Learning/](https://github.com/TheiScale/30_Days_Machine_Learning/)



## What is feature construction in machine learning?

**Feature Construction:** - Feature construction is a process which builds intermediate features from the original descriptors in a dataset. The aim is to build more efficient features for a ML task.

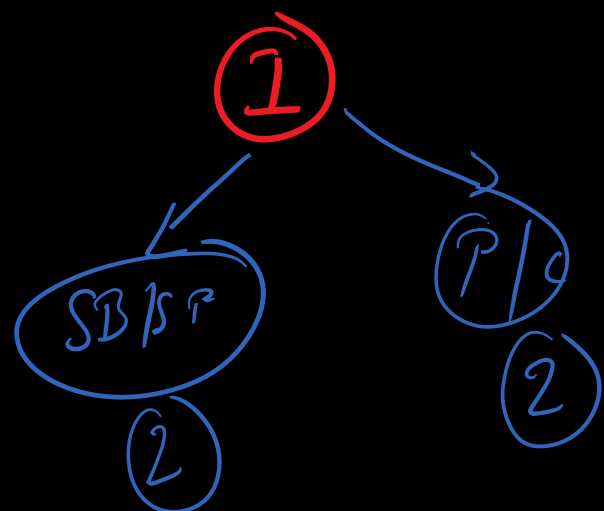
Example: In Titanic Dataset ✓

S. No.	Sibling/Spouse	Parents/Child	FAMILY
1	1	0	2
2	1	0	2
3	0	1	2
4	1	1	4

new

→ Sib/Spouse → Par/child

	Age	Pclass	SibSp	Parch	Family_size
0	22.0	3	1	0	2
1	38.0	1	1	0	2
2	26.0	3	0	0	1
3	35.0	1	1	0	2
4	35.0	3	0	0	1



4 - Big Family  
2 - Small  
1 - Single

## What is Feature Splitting?

### Feature Splitting:

Feature splitting is a technique in machine learning that involves breaking down a single feature into multiple features. The purpose of this technique is to generate more informative features that can provide greater insight into the relationships between input variables and the target variable.

Example: Splitting

S. No.	Name	Split 1	Split 2
1	Mr. Nishant	Mr.	Nishant
2	Mrs. Avantika	Mrs.	Avantika
3	Dr. Reddy	Dr.	Reddy
4	Miss. Jyoti	Miss	Jyoti

	Title	Name
0	Mr	Braund, Mr. Owen Harris
1	Mrs	Cumings, Mrs. John Bradley (Florence Briggs Th...
2	Miss	Heikkinen, Miss. Laina
3	Mrs	Futrelle, Mrs. Jacques Heath (Lily May Peel)
4	Mr	Allen, Mr. William Henry
...	...	...
886	Rev	Montvila, Rev. Juozas
887	Miss	Graham, Miss. Margaret Edith
888	Miss	Johnston, Miss. Catherine Helen "Carrie"
889	Mr	Behr, Mr. Karl Howell
890	Mr	Dooley, Mr. Patrick

891 rows × 2 columns

→ screenshot  
code ✓

## Difference Between Feature Construction & Feature Splitting

### Purpose | Technique | Uses | Result

	Feature Construction	Feature Splitting
Definition	Creating new features from existing ones	Dividing a single feature into multiple features
Purpose	Capture important patterns in the data not captured by original features	Capture different aspects of the original feature relevant to the ML task
Technique	Transforming, combining, or scaling existing features	Extracting different aspects of the original feature
Example	Creating a new feature called body mass index (BMI) by combining height and weight	Splitting a date feature into day, month, and year features
Result	Increases the number of features	Increases the number of features
Impact on model performance	May improve model performance by capturing additional patterns in the data	May improve model performance by providing more information to the model
When to use	When there is a need to create additional features to improve model performance	When there is a need to capture different aspects of a single feature for the ML task

## <Start Coding>

### #Import Library

```
import numpy as np
import pandas as pd
```

```
from sklearn.model_selection import cross_val_score
from sklearn.linear_model import LogisticRegression
```

```
import seaborn as sns
```

### #Import Dataset

```
df =
pd.read_csv('train.csv')[['Age', 'Pclass', 'SibSp', 'Pa
rch', 'Survived']]
```

```
----
```

```
df.head()
```

### #Drop NA Value

```
df.dropna(inplace=True)
```

```
----
```

```
df.head()
```

```
#Separate X and Y
```

```
X = df.iloc[:,0:4]  
y = df.iloc[:, -1]
```

```
----
```

```
X.head()
```

```
#Check Accuracy for Logistic Regression
```

```
np.mean(cross_val_score(LogisticRegression(),X,y,sco  
ring='accuracy',cv=20))
```

```
#Applying Feature Construction
```

```
#Create New Column
```

```
X['Family_size'] = X['SibSp'] + X['Parch'] + 1
```

```
----
```

```
x.head()
```

```
#Apply New Function
```

```
def myfunc(num):  
    if num == 1:  
        #alone  
        return 0  
    elif num >1 and num <=4:  
        # small family  
        return 1  
    else:  
        # large family  
        return 2
```

```
----  
myfunc(4)  
  
#Apply M Function  
  
X['Family_type'] = X['Family_size'].apply(myfunc)  
  
----  
  
X.head()  
  
#Drop unwanted columns  
  
X.drop(columns=['SibSp', 'Parch', 'Family_size'], inp  
lace=True)  
  
----  
  
X.head()  
  
#Review Accuracy after Feature Construction  
  
np.mean(cross_val_score(LogisticRegression(), X, y, s  
coring='accuracy', cv=20))
```

```
=====
```

```
#Feature Splitting
```

```
#Review Import Data
```

```
df = pd.read_csv('train.csv')
```

```
----
```

```
df.head()
```

```
#Use Name Column
```

```
df['Name']
```

```
#Separate Salutation
```

```
df['Title'] = df['Name'].str.split(', ',  
expand=True)[1].str.split('.', expand=True)[0]
```

```
----
```

```
df['Name'].str.split(', ',  
expand=True)[1].str.split('.', expand=True)[0]
```

```
----
```

```
df[['Title', 'Name']]
```

```
#Review Analysis after Feature Splitting
```

```
np.mean(cross_val_score(LogisticRegression(), X, y, s  
coring='accuracy', cv=20))
```