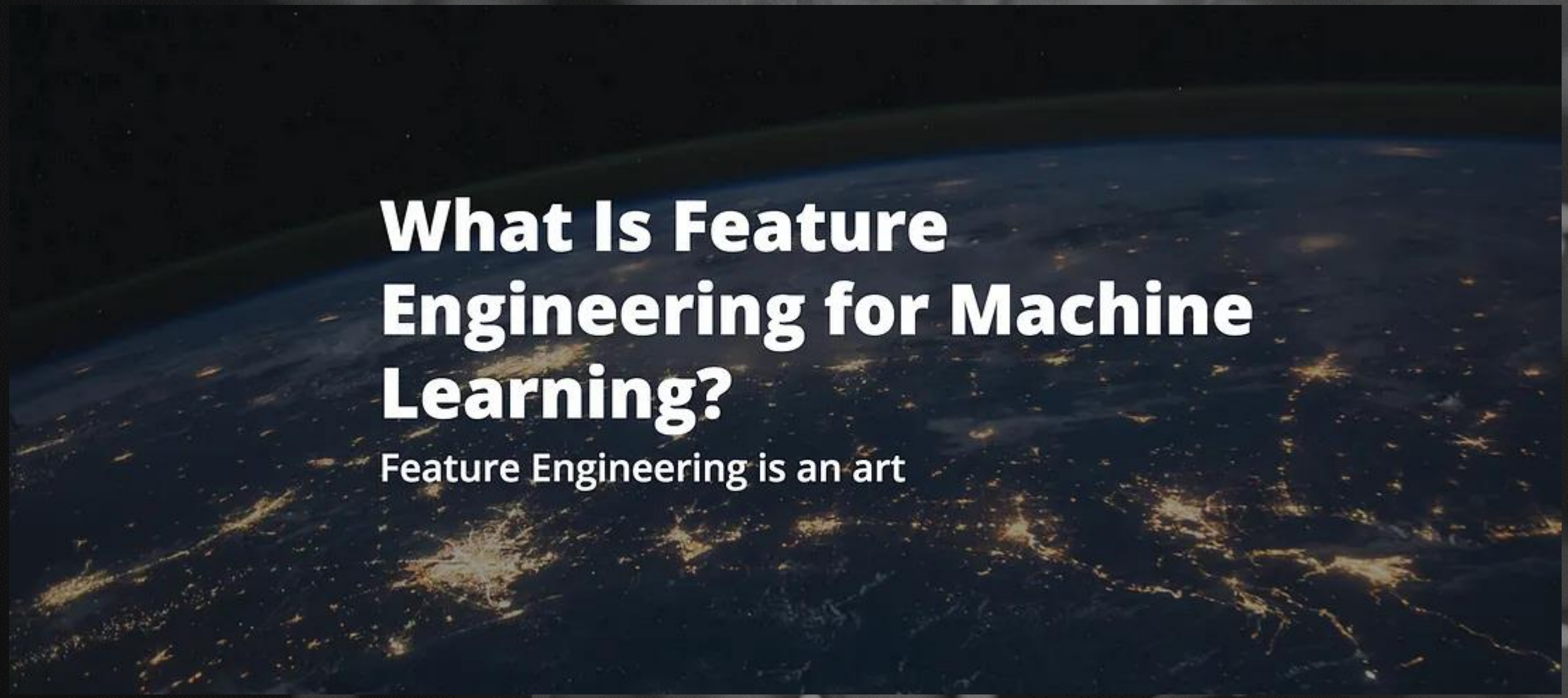

What Is Feature Engineering for Machine Learning?
Feature Engineering is an art

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

FEATURE ENGINEERING

- *Feature engineering is the process of using domain knowledge of the data to create features that make machine learning algorithms work.*

- *If feature engineering is done correctly, it increases the predictive power of machine learning algorithms by creating features from raw data that help facilitate the machine learning process.*

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

PROBLEM STATEMENT: ANTICIPATING OR PREDICTING THE CAREER PROGRESS OF EVERY INDIVIDUAL

In today's competitive job market, individuals seek ways to predict and enhance their career success. Employers also aim to identify promising candidates based on various factors such as age, current job, skillset, interests, responsibilities, and qualifications, these are generally called features and we undergo some sequence of steps that need to follow to make features more accurate to improve the accuracy of the machine learning model to predict the career success. This is called Feature Engineering.

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

FEATURE ENGINEERING

For Internships/Workshops: Mail hr@brainovision.in, Contact: +91 7416422509

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

Feature engineering is the most important art in machine learning which creates a huge difference between a good model and a bad model.

STEPS THAT ARE INVOLVED WHILE SOLVING ANY PROBLEM IN MACHINE LEARNING ARE AS FOLLOWS:

Feature Creation
Feature Selection
Feature Transformation
Feature Encoding
Feature Extraction

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

FEATURE CREATION

- *Creating features involves creating new variables which will be most helpful for our model.*

- *These artificial features are then used by the algorithm in order to improve its performance, or in other words getting better results.*

EXAMPLE:

These are the following features we have gathered for the problem statement discussed above:

Name: Name of the person

Father Name: Father name of the person

Contact: Point of contact

Email Id: Email id to contact

Address : Communication Address

Age: Age of the individual.

Height: Height of an individual

Weight: Weight of an individual

Current Job: Title or position of the individual's current employment.

Skillset: A list of skills possessed by the individual.

Interests: Areas of interest or hobbies.

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

FEW MORE FEATURES

Responsibilities: Duties or responsibilities associated with the individual's current job.

Qualification: Educational qualifications achieved by the individual.

Years of Experience: Experienced individuals having diverse skill set needed to the industry.

Education Level: Include variables such as highest degree obtained certifications that might be relevant to the career.

Networking Skills: Assess the individual's ability to build and maintain professional relationships.

Adaptability: Include features that reflect an individual's ability to adapt to changes.

Leadership Experience: If the role involves leadership, include features related to past leadership experience.

Industry-specific Skills: Depending on the career, include specific skills that are highly valued in that industry.

Work-Life Balance: Consider factors related to work-life balance, as individuals with a healthy work-life balance may perform better in the long run.

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

PREDICTION FEATURES

We are going to predict outcome as the following:

Career_Success: It is output regression

Career: It is the classification

What is Feature Selection in Machine Learning?

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

FEATURE SELECTION

- *Feature Selection is the method of reducing the input variable to your model by using only relevant data and getting rid of noise in data.*

- *It is the process of automatically choosing relevant features for your machine learning model based on the type of problem you are trying to solve.*

FEATURES SELECTED:

Inputs:

Age: Age of the individual.

Height: Height of an individual

Weight: Weight of an individual

Current_Job: Title or position of the individual's current employment.

Skillset: A list of skills possessed by the individual.

Interests: Areas of interest or hobbies.

Responsibilities: Duties or responsibilities associated with the individual's current job.

Qualification: Educational qualifications achieved by the individual.

Outputs:

Career_Success: It is output regression

Career: It is the classification

FEATURE TRANSFORMATION

- It refers to creates new features using the existing features.

- Feature transformation refers to the process of applying a mathematical operation or function to the existing features in order to change their representation.

Example: Scaling features (e.g., normalization or standardization), logarithmic transformation, polynomial transformation, or applying other mathematical functions to the features.

For Internships/Workshops: Mail hr@brainovision.in, Contact: +91 7416422509

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

EXAMPLE:

Salary: 10000, 20000,80000 etc.,

Distance: 1000m, 5000m

Like this we have ml to litres so on…..

FEATURES AFTER TRANFORMATION:

Inputs:

Age: Age of the individual.

Height: Height of an individual

Weight: Weight of an individual

Current_Job: Title or position of the individual's current employment.

Skillset: A list of skills possessed by the individual.

Interests: Areas of interest or hobbies.

Responsibilities: Duties or responsibilities associated with the individual's current job.

Qualification: Educational qualifications achieved by the individual.

Outputs:

Career_Success: It is output regression

Career: It is the classification

BRAIN O VISION

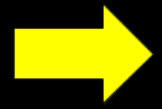
Brain O Vision Solutions India Pvt. Ltd.

FEATURE ENCODING

Transforming the categorical values of the relevant features into numerical ones. This process is called feature encoding.

Current_Job	Current_Job
Data Scientist	1
Software Engineer	2
Software Programmer	3
Financial Analyst	4
Accountant	5

Color
Red
Red
Yellow
Green
Yellow



Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1

BRAIN O VISION
Brain O Vision Solutions India Pvt. Ltd.

FEATURE EXTRACTION

- *Feature extraction helps to reduce the amount of redundant data from the data set.*
- *It yields better results by applying machine learning directly to the raw data.*
- *Example: Suppose we have height and weight features – bmi*

$$BMI = weight(kg)/height(m)^2$$

FEATURES AFTER EXTRACTION:

Inputs:

Age: Age of the individual.

Bmi: $\frac{weight(kg)}{height(m)^2}$

Current_Job: Title or position of the individual's current employment.

Skillset: A list of skills possessed by the individual.

Interests: Areas of interest or hobbies.

Responsibilities: Duties or responsibilities associated with the individual's current job.

Qualification: Educational qualifications achieved by the individual.

Outputs:

Career_Success: It is output regression

Career: It is the classification

Note: Here bmis done not reflect on the output career_success, we can remove it.

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

DATASET PREPARATION

Dataset preparation is a crucial step in the data science workflow. It involves collecting, cleaning, and organizing the data to make it suitable for analysis. Here are the key steps involved in dataset preparation:

Data Collection
Data Cleaning
Data Exploration
Data Transformation
Feature Extraction

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

DATASET COLLECTION

Age	Current_Job	Skillset	Interests	Responsibilites	Qualification	Career_ success	Class
24	Data Scientist	Programming	Coding	Upskilling	B.Tech		
23	Software Engineer	Administration	Data and Analytics	Reviewing	B.Tech		
18	Software Programmer	Cybersecurity	Networking	Upskilling	M.Tech		
23	Financial Analyst	Database Management	Project Management	Reviewing	MCA		
19	Accountant	Problem-Solving	Automation	Quality	BCA		
18	Accountant	Problem-Solving	Networking	Reviewing	DEGREE		
22	Financial Analyst	Database Management	Automation	Quality	MCA		

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

DATA CLEANING AND PREPROCESSING

It is a process of Handling missing data or removing unnecessary data etc.,

- *Example:* Inputing missing values in a dataset of customer feedback scores by using the mean or median values.
- Removing unnecessary rows for which maximum data is not found etc..

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

DATA TRANSFORMATION:

Age: 18-25, Current_Job:1-5, Skillset: 1-5, Interests: 1-10,Responsibilites: 1-5, Qualification: 1-5

Age	Current_Job	Skillset	Interests	Responsibilites	Qualification	Career_success	Class
24	3	1	3	3	2		
23	4	5	10	2	1		
18	4	2	10	1	2		

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

DATA EXPLORATION:

Data exploration is a crucial initial step in the data science process. It involves analyzing and visualizing the dataset to gain insights, understand patterns, and identify potential challenges.

```
import pandas as pd
data = pd.DataFrame(sample_data)
print(data.shape)  # Number of rows and columns
print(data.info())  # Data types and non-null counts
print(data.describe())  # Basic statistics like mean,median so on.....
correlation_matrix = data.corr()
print(correlation_matrix)
```

Age	Current_Job	Skillset	Interests	Responsibilites	Qualification	Career_success	Class
24	3	1	3	3	2		1
23	4	5	10	2	1		1
18	4	2	10	1	2	6	0

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

FEATURE EXTRACTION:

Creating new features Career_Success and Class

Age: 18-25, Current_Job:1-5, Skillset: 1-5, Interests: 1-10,Responsibilites: 1-5, Qualification: 1-5

Weights: Age: 0.2, Current_Job:0.1, Skillset: 0.3, Interests: 0.1,Responsibilites: 0.1, Qualification: 0.1 ➔ 1

Threshold: tr: (0.2*w1,0.1*w2,0.3*w3,0.1*w4,0.2*w5,0.1*w6)*0.7(cutoff)

Age	Current_Job	Skillset	Interests	Responsibilites	Qualification	Career_success	Class
24	3	1	3	3	2	6.2	1
23	4	5	10	2	1	2.3	0
18	4	2	10	1	2		

CODE TO POPULATE OR GENERATE THE DATA

```
import random
import pandas as pd
[w1,w2,w3,w4,w5,w6] = [0.2,0.1,0.3,0.1,0.2,0.1]
vals = []
tr = (25*w1+5*w2+5*w3+10*w4+3*w5+5*w6)*0.7(cutoff)
for i in range(10000):
    x1 = random.randint(18, 25)
    x2 = random.randint(1, 5)
    x3 = random.randint(1, 5)
    x4 = random.randint(1, 10)
    x5 = random.randint(1, 3)
    x6 = random.randint(1, 5)
    eq = w1*x1+w2*x2+w3*x3+w4*x4+w5*x5+w6*x6
    if eq > tr:
        cls = 1
    else:
        cls = 0
    vals.append([x1,x2,x3,x4,x5,x6,eq,cls])
df = pd.DataFrame(vals,columns=['age','current_job','skillset','interests'
,'responsibilities','qualification','career_success','class'])
df.to_csv('career_success.csv',index=False)
```

BRAIN O VISION

Brain O Vision Solutions India Pvt. Ltd.

THANK YOU