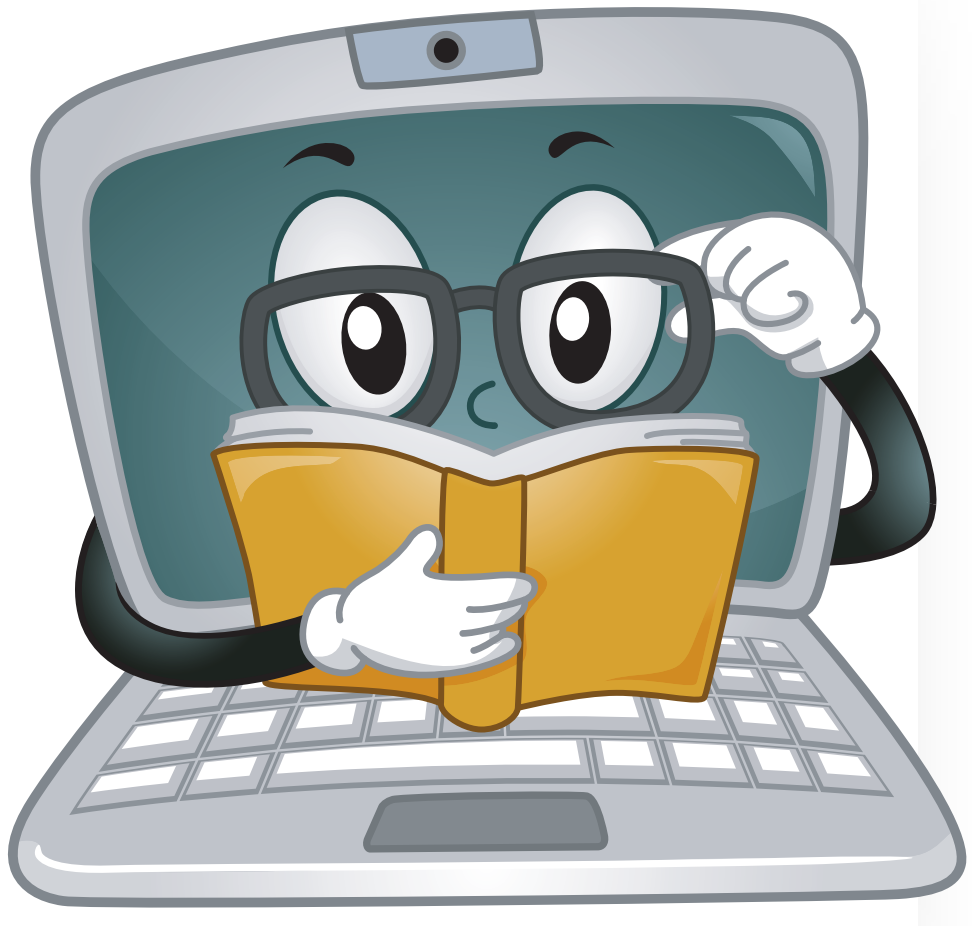
LET THE DATA CONFESS

Understand | Learn | Code | Implement

What is Data Encoding

-letthedataconfess



It is a way of representing or converting categorical data into numerical representation so that the algorithms can easily interpret it.

https://www.letthedataconfess.com/

LET THE DATA CONFESS
Understand | Learn | Code | Implement

Types of Encoding

1. One Hot Encoding
2. Label Encoding
3. Count of Frequency Encoding
4. Target Guided Encoding
5. Hash Encoding
6. Binary Encoding
7. Dummy Encoding

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

One Hot Encoding

In this type of encoding technique for each category we will create a new feature or column separately and wherever that category is present it will be represented by 1 and rest all with 0.

https://www.letthedataconfess.com/

LET THE DATA CONFESS
Understand | Learn | Code | Implement

When to use One Hot Encoding ??

It is a type of an encoding where each category will be assigned a number ranging till the length of that category.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

Label Encoding

It is a type of an encoding where each category will be assigned a number ranging till the length of that category.

https://www.letthedataconfess.com/

LET THE DATA CONFESS
Understand | Learn | Code | Implement

When to use Label Encoding ??

It can be used when there is a specific order of categories or it's an ordinal variable.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

Count of Frequency Encoding

This type of encoding is done based on the count of occurrence of the categorical feature. If category has more count then the value of the count will be assigned.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

When to use Count of Frequency Encoding??

It can be used when the variables are nominal or there is no rank among categories

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

Target Guided Encoding

In this type, mean value is calculated for each category based on target variable, so if the value of mean is high with respect to target then higher integer value will be assigned to that category. If the mean value is low then a lower integer value will be assigned.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

When to use Target Guided Encoding??

It can be used when the order of category is necessary or its an ordinal variable.

https://www.letthedataconfess.com/

LET THE DATA CONFESS
Understand | Learn | Code | Implement

Hash Encoding

In this encoding for every category an integer value will be generated using a hashing function so each category will be mapped to an integer value. Here the category is the key and the generated integer is the pair value. Also chances are 2 different categories may have same integer value causing collision.

https://www.letthedataconfess.com/

When to use Hash Encoding??

When the order of categories is not important or it's a nominal variable.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

6 Binary Encoding

In this type, for each category a new feature is created just like one hot encoding. They are converted to ordinal variables, but instead of assigning 1, binary values are assigned.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

When to use Binary Encoding??

When the order of variables is ordinal or where the order is important.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

Dummy Encoding

It is the same as one hot encoding.

The only difference is the last feature or column will be dropped, suppose there are 5 features the last feature will be dropped and 4 features will remain.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

When to use Dummy Encoding??

When the variables are nominal or there is no specific order in categories.

https://www.letthedataconfess.com/

LET THE DATA CONFESS

Understand | Learn | Code | Implement

Facebook

Instagram

Linkedin

Telegram

Twitter

http://

https://www.letthedataconfess.com/