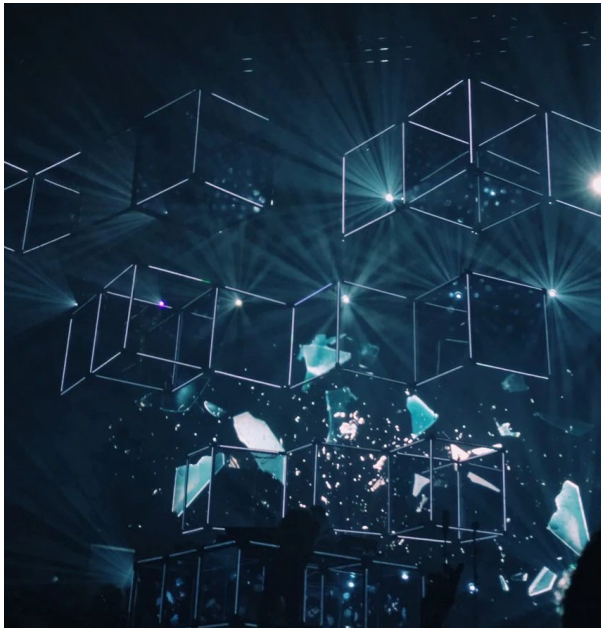
Introduction to feature engineering

DATA ANALYTICS | IRONHACK



Credit: Unsplash

Agenda

1. Feature engineering.
2. Feature extraction/generation.
3. Feature selection.

FEATURE ENGINEERING

- Is a process of transforming the given data into a form that is easier to interpret.

- All the techniques that we have used before including data transformations techniques, encoding categorical variables, scaling numerical features, imputing missing values, cleaning categorical columns, using regular expressions, using DateTime, and string functions are feature engineering techniques.

- The key idea is that we are manipulating the information that is available to us to be able to better understand it and improve the model.

FEATURE EXTRACTION/GENERATION

- It is the process of extracting relevant information from the existing available information.
- For example, you are provided with the date of birth of customers in the data. You are not interested in when they were born but more interested in their age.
- In that case, you can use the DateTime functions to calculate the age or extract other relevant information such as the year they were born, the month they were born.

FEATURE SELECTION

- It is the process of extracting relevant information from the existing available information.
- For example, you are provided with the date of birth of customers in the data. You are not interested in when they were born but more interested in their age.
- In that case, you can use the DateTime functions to calculate the age or extract other relevant information such as the year they were born, the month they were born.

FEATURE SELECTION

Some of the feature selection methods that we took a look at earlier include:

- Checking null values to drop a column
- Sense check to drop columns that are not significant
- Using heat maps to check multicollinearity for numerical variables
- Chi-square tests for categorical variables