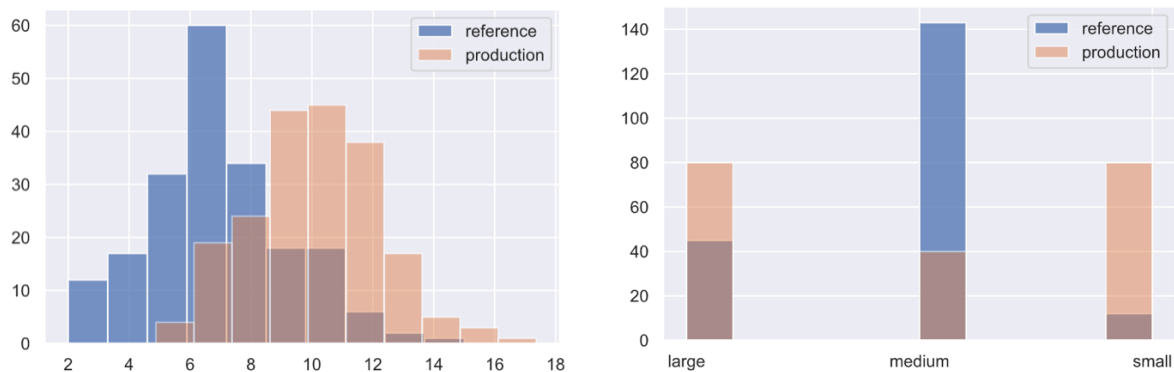


Data Drift

Data Drift is the difference between training data and test data. This happens after a period of time from model deployed to the production.

If we are using live streaming data after a period of time there are chances that the distribution, patterns and format of the new data may change, which causes data drift. For which the model's accuracy will get impacted.



Types of Data Drift:

1. Target Drift:

In binary class classification problem, change in the values of output label.

Eg: Output label(Training data): Male, Female

Output label(Test data): Male, Female, Others

2. Concept Drift:

In predictive model if the input and output data relationship is changed over time, this is known as concept drift. In other words if statistical properties of the target variable which model is trying to predict, changes over time.

eg: Covariance (Training data): 58%

Covariance (Test data): 2%

Connect with me: <https://www.linkedin.com/in/sai-subhasish-rout>

Follow me: <https://github.com/saisubhasish>