

Machine learning live cycle

Data collection _ integration :

Data collection is the process of gathering and measuring information from countless different sources. In order to use the data we collect to develop practical artificial intelligence (AI) and machine learning solutions, it must be collected and stored in a way that makes sense for the business problem at hand.

(Integration of multiple databases, data cubes, files, or notes)

Data cleansing :

Data cleansing or data cleaning is the process of detecting and correcting corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

● **Data cleaning**

- Fill in missing values
- smooth noisy data
- identify or remove outliers
- resolve inconsistencies

Feature extraction and selection (Dimensionality Reduction) :

Feature selection is for filtering irrelevant or redundant features from your data set. The key difference between feature selection and extraction is that feature selection keeps a subset of the original features while feature extraction creates brand new ones.

Model training :

Training a model simply means learning (determining) good values for all the weights and the bias from labeled examples.

(Use data that has been cleaned up and feature engineering to train a model)

Model Evaluation :

Model evaluation is the process of using different evaluation metrics to understand a machine learning model's performance, as well as its strengths and weaknesses. Model evaluation is important to assess the efficacy of a model during initial research phases, and it also plays a role in model monitoring

Model Deployment :

Deployment is the method by which you integrate a machine learning model into an existing production environment to make practical business decisions based on data