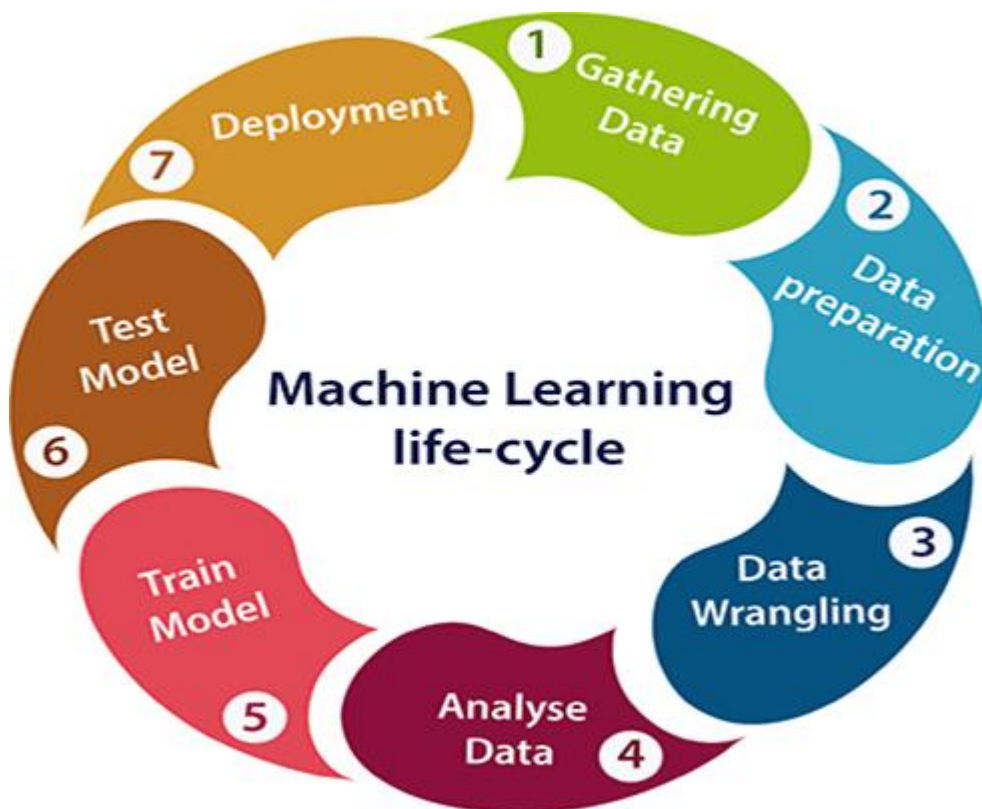# Machine Learning Development Life Cycle

Planning: The planning step is essential for ensuring that the machine learning project is successful. During this step, the following questions should be answered:

○ What is the business problem that the machine learning model will solve?
○ What data is available to train the model?
○ What machine learning algorithms and techniques are appropriate for the problem?
○ What are the success criteria for the project?

Gathering data: The next step is to gather the data that will be used to train the machine learning model. This data can be collected from a variety of sources, such as:

- Internal data from the organization
- External data from public sources
- Data generated by sensors or other devices

It is important to ensure that the data is accurate, relevant, and complete. The data should also be cleaned to remove any errors or outliers.

Data preparation: The data preparation step is critical for ensuring that the machine learning model is trained on accurate and relevant data. During this step, the following tasks should be performed:

- Clean the data to remove errors and outliers.
- Transform the data into a format that is compatible with the machine learning algorithm.
- Label the data if it is not already labeled.

Data wrangling: This step involves cleaning and transforming the data in a more complex way, such as by identifying and removing outliers, imputing missing values, and creating new features.

Data analysis: Before training the model, it is important to analyze the data to gain insights into the problem that the model is trying to solve. This involves the following tasks:

- Exploratory data analysis (EDA): This involves using statistical methods to explore the data and identify patterns and trends.
- Feature engineering: This involves creating new features from the existing data to improve the performance of the machine learning model.

Train model: The model development step involves selecting the right machine learning algorithm, training the model on the prepared data, and evaluating the model's performance. During this step, the following tasks should be performed:

- Select the machine learning algorithm that is most appropriate for the problem.
- Train the model on the prepared data.
- Evaluate the model's performance using a variety of metrics.

Test model: This step involves evaluating the performance of the machine learning model on a holdout dataset. The holdout dataset is a set of data that was not used

to train the model. This step helps to ensure that the model is not overfitting the training data.

Deployment: The model deployment step involves making the machine learning model available to users so that it can be used to make predictions. During this step, the following tasks should be performed:

- ○ Choose a deployment platform.
- ○ Deploy the model to the platform.
- ○ Monitor the model's performance.

Monitoring and maintenance: The monitoring and maintenance step involves monitoring the performance of the machine learning model and making changes as needed to ensure that it continues to perform well. During this step, the following tasks should be performed:

- ○ Monitor the model's performance using a variety of metrics.
- ○ Make changes to the model if its performance is not satisfactory.
- ○ Update the model with new data as it becomes available.

The MLDLC is an iterative process, and it is important to go back and forth between the different steps as needed. By following the MLDLC, organizations can increase the chances of success for their machine learning projects.