**3. Define problem category for below problem statement “A chemist wants to find some interesting patterns in which patients are behaving upon administering the drug”.**

* This problem statement comes under “Unsupervised Learning” where algorithm used to draw inferences from datasets consisting of input data without labeled responses.
* We grouped it under “Unsupervised” because the here in this problem statement outcomes are based on software analysis of relations or patterns between patients and drugs without user provided any specific datasets.
* Here, we as a Data Scientists need to use techniques to determine which drugs are related to which particular patients and will group them into classes.
* Thus, we can use exploratory data analysis to find hidden patterns between patients behaving upon different drugs.

**4. How will you select suitable machine learning algorithm for a problem statement**

* In order to choose a suitable machine Learning Algorithm, we need to observe the data we have and according to that we need to classify the statement, there are basically four types of approach based on our data,
* Supervised learning
* Unsupervised learning
* Semi-supervised learning
* Reinforcement learning

**Supervised learning --->** Supervised learning is the task of inferring a function from labeled training data. By fitting to the labeled training set, we want to find the most optimal model parameters to predict unknown labels on test set. If the label is a real number, we call the task regression. If the label is from the limited number of values, where these values are unordered, then it’s classification.

**Unsupervised learning --->** In unsupervised learning we have less information about objects, it’s possible to observe some similarities between groups of objects and include them in appropriate clusters. Some objects can differ hugely from all clusters, in this way we assume these objects to be anomalies.

**Semi-supervised learning --->** Semi-supervised learning tasks include supervised and unsupervised learning. They use labeled and unlabeled data. That is a great opportunity for those who can’t afford labeling their data. The method allows us to significantly improve accuracy, because we can use unlabeled data in the train set with a small amount of labeled data.

**Reinforcement learning --->** In Reinforcement learning, we don’t have labeled or unlabeled datasets here. RL is an area of machine learning concerned with how software agents ought to take actions in some environment to maximize some notion of cumulative reward.

5**.Define one problem statement for Education industry?**

* From our perspective we can take an example of problem statement for Education Industry – “What effect does discipline have on creativity among school students”.