3.   
According to given problem statement

'A chemist wants to find some interesting patterns in which patients are behaving upon administering the drug'

We shall break the above problem in following way:

1. Input: The Drug which has been prescribed to patients.
2. Output: Behaviour of different patients upon application of the given drugs.

Here we must understand that drugs given to set of patients are known i.e input to the model is known, but we do not have predefined groups of known behavior of patients upon application of known drugs.

i.e Labels are missing which is fundamental requirement of any classification problem.

So to implement this problem we have to study the behavior patterns among patients and draw similarity among available sets of pattern which is nothing but

Clustering .

So this problem can be termed as Unsupervised learning problem.

4.

Following factor must be considered while choosing the machine learning algorithm :

1. Understanding the Data:

* The size and type of data must be considered while choosing ML algo

One must check if data is numerical(continuous) or categorical. We use different algorithm for different type of data e.g Regression needs numerical data while classification needs categorical data.

* Spread of data , central tendency and correlation among data must be analysed to select appropriate algorithm
* Principal component Analysis must be performed for dimensionality reduction for some algorithms which are sensitive towards range of features.
* Appropriate measures must be undertaken to tackle Missing values in the data sample since some algorithm are sensitive to missing values .
* If not taken care of Outliers can affect the accuracy of models especially regression algorithms, Outliers must be normalized for such scenarios.

1. Feature Engineering:

* Number of features to be made out of existing data and the type of feature is very crucial in ML lifecycle.
* Model can perform to it’s fullest if appropriate feature is selected.

1. Categorize the problem

* **Categorize by the input:**If it is a labeled data, it’s a supervised learning problem. If it’s unlabeled data with the purpose of finding structure, it’s an unsupervised learning problem. If the solution implies to optimize an objective function by interacting with an environment, it’s a reinforcement learning problem.
* **Categorize by output:**If the output of the model is a number, it’s a regression problem. If the output of the model is a class, it’s a classification problem. If the output of the model is a set of input groups, it’s a clustering problem.

1. Requirement of the Application:

* There is always a tradeoff between accuracy , Run time and Utilization of computing resources.
* One can get 100% accuracy and minimal run time provided infinite computing resources which is hypothetical scenario.
* So algorithm must be selected according to application reqirement.
* E.g For problems were real time prediction is required like self driving cars, A algorithm must be selected which can bulid the model in real time with newly arrived data.

1. Once the model is decided it is advisable to test it with different sets of train data to analysed it’s behavior .
2. Scalability and complexity of the model must be considered .