

Batch vs Online Machine Learning — What's the Difference?

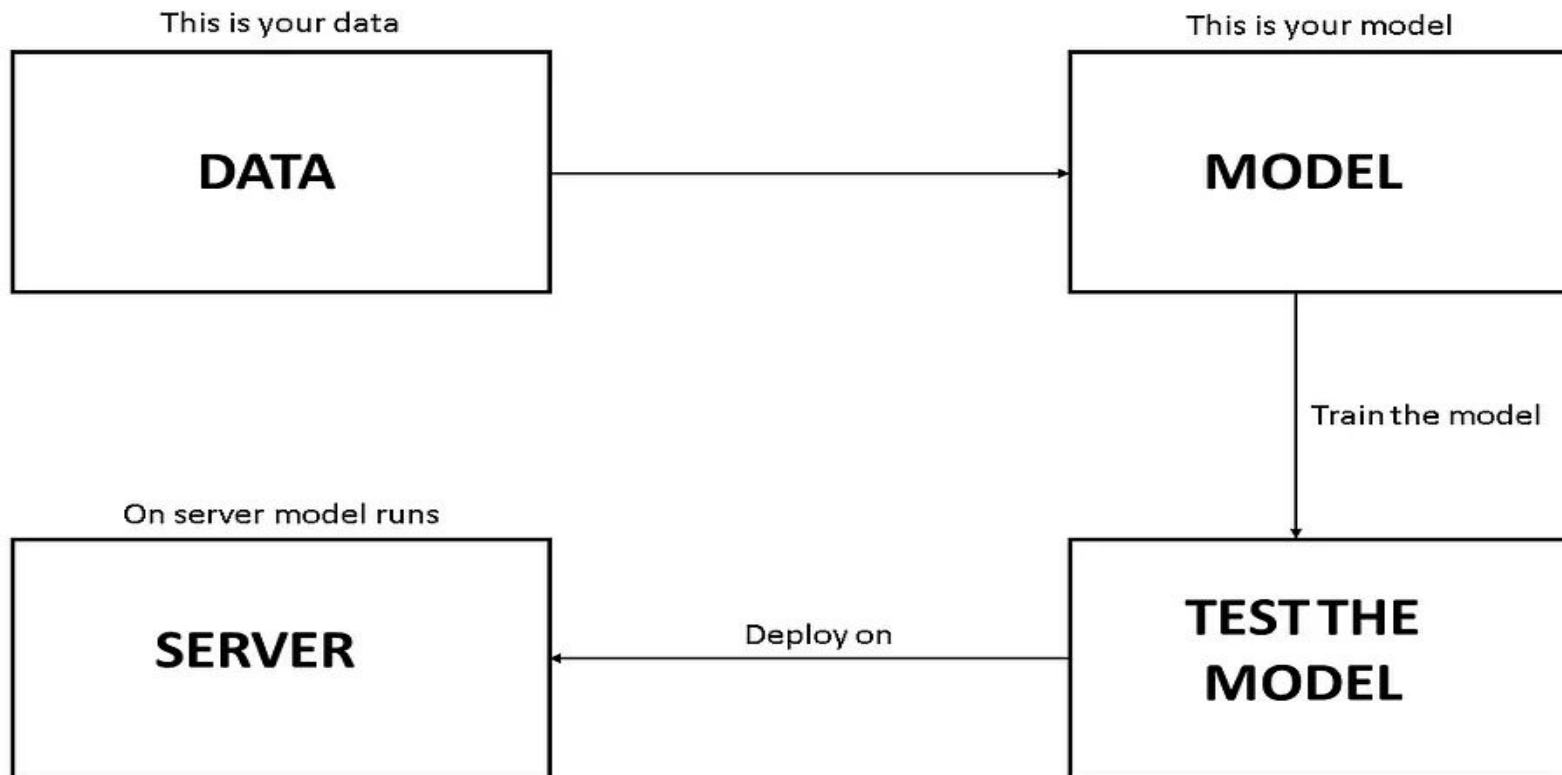
- Both batch learning and online learning raise concerns about how machine learning algorithms learn or train, especially in a production environment.
- In simple language production is a server on which your code is going to run

Batch learning:-

- It is a conventional way of training an ML model. In which you use all the data to train the model. You utilize all of your data.
- There is no incremental training. Incremental training is a learning technique where a model is updated and improved over a time by training it on a new data while retaining from previous learned patterns
- When you are working on a real-world problem. The data will be huge. Generally it is when you are going to train such a huge amount of data on a server. It would be costly and time consuming.

In batch learning, you take up your entire data. You train your machine learning model on an offline system. Once the model is trained, you deploy it on the server.

This is the entire flow of batch learning =>



Problems with batch learning:-

Your model is static, which means it doesn't update or learn from new data. the problem with this approach is your business scenario is evolving

for example:-consider netflix,

- You made a recommendation system for Netflix. You considered all the movies and shows that are available today.
- But since Netflix adds new movies and TV shows weekly, your recommendation system should always grow in order to incorporate new movies.
- Your machine learning model needs to evolve constantly with new data.If you do not do this, what is the benefit of your recommendation system.

Another example is email spam classifier,

Your email spam classifier is up-to-date as of today's date. You put it on the server once, and after that, no training took place. You left it like that for one year.

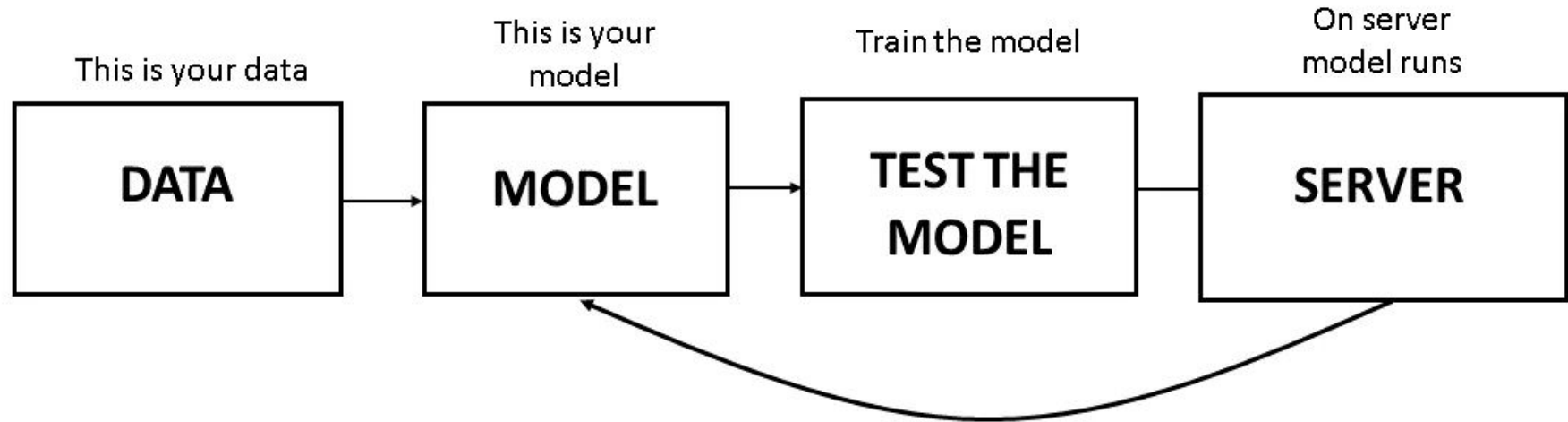
But in one year, marketers will find new techniques. To create spam emails that do not end up in the spam folder.

But in one year, marketers will find new techniques. To create spam emails that do not end up in the spam folder.

If you do not keep your system up-to-date, then what's the point? It will be obsolete.

The biggest problem with batch learning is that you need to retrain your model frequently; generally, people do this periodically. What they do is take the updated data, merge it with the old data, and retrain the model with the whole new data. Once the model is trained, you test it again and put it on the server.

Basically this process which happens in cycle Repeats again and again in the given period. This period could be 24 hr period, weekly or monthly or 6 monthly

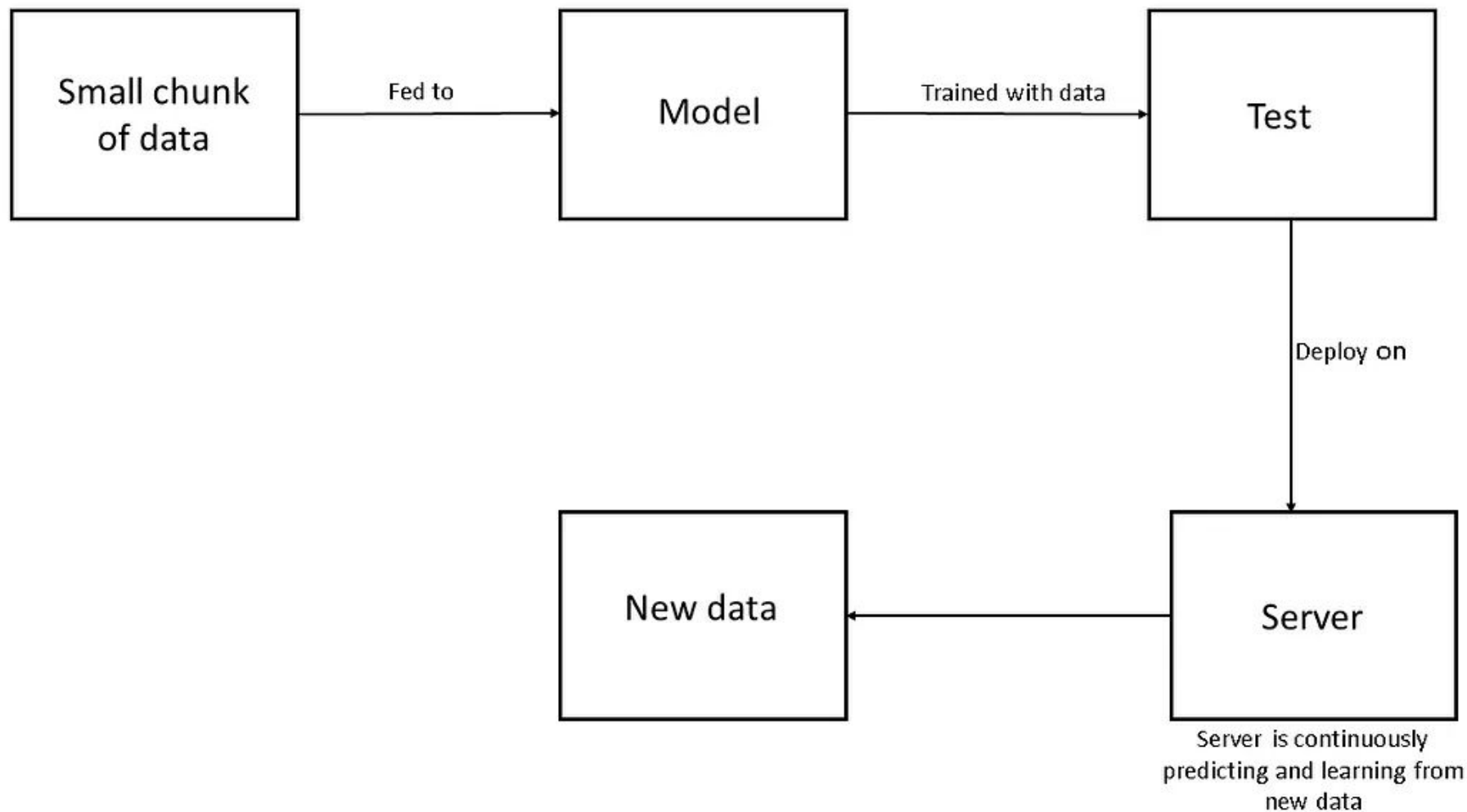


Online learning:-

Online learning is quite unlike batch learning, which is done incrementally.

- “Have you ever heard that many companies promote their products in such a way that you will use our products as much. The performance of our product will go on increasing”
- Means your performance in terms of usability increase
- So, have you heard this then those Companies are talking about “online learning”

- So, what you do in online learning is that. You feed data to your model in small batches, sequentially. These batches are called mini batches.
- After, each batch of training, your model gets better. since these batches are small chunks of data. so you can perform this training on server (in production)
- That's why it is called online learning means your model is getting trained when your model is on server



example of online learning

chat bots:-

- All of you used chatbots. The chatbots by famous companies like Google Assistant, Alexa, and Siri are good. examples of online learning
- Because when it is deployed on the server at the time of deployment, it is also doing prediction as chat with new data. and at the same time learning from that new data.

Youtube:-

when you scroll on YouTube and come back to the feed by clicking a particular video. After watching the video, the feed changes. It automatically gets modified according to the video. that has been watched

This is again an example of online learning.

Learning rate:-

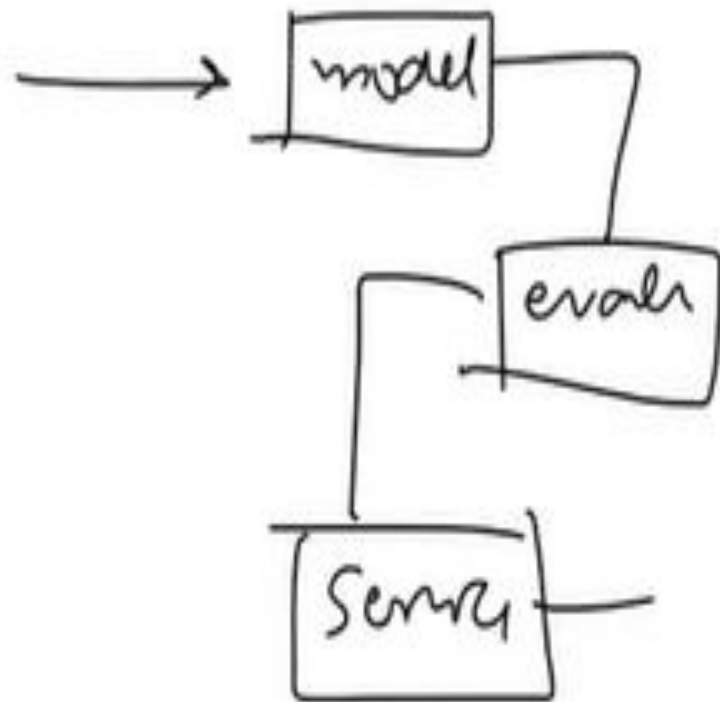
- There is a concept in online learning called learning rate. The learning rate determines how frequently your model is going to be trained.
- You don't train your model on every incoming piece of data; by doing this, your model drastically changes, and forgets the old learned things.
- Generally, you would not like to forget everything old and start learning new things. And you don't even want to learn slowly, under this is also bad

so correct learning rate has to be kept

Setting the correct learning rate for business in online learning is a difficult task if you do not do it properly. your model. may misbehave

Out of Core Learning

Out-of-core learning refers to a set of algorithms designed to handle data that exceeds the capacity of a machine's primary memory (RAM). These algorithms work by reading small mini-batches of data from the disk, processing them, and then discarding them to free up memory for the next batch



Disadvantages:-

- Tricky to Use
- Risky

Offline Learning	Features	Online Learning
Less complex as model is constant	Complexity	Dynamic complexity as the model keeps evolving over time
Fewer computations, single time batch-based training	Computational Power	Continuous data ingestions result in consequent model refinement computations
Easier to implement	Use in Production	Difficult to implement and manage
Image Classification or anything related to Machine Learning - where data patterns remains constant without sudden concept drifts	Applications	Used in finance, economics, health where new data patterns are constantly emerging
Industry proven tools. E.g. Sci-kit, TensorFlow, Pytorch, Keras, Spark Mlib	Tools	Active research/New project tools: E.g. MOA, SAMOA, scikit-multiflow, streamDM