Data analytics is being used in each and every industry for improving the performance of the companies just like in the example of a cricket match where the captain and coach analyze the performance of the opponent team and then device strategies for winning the match. If we begin to list down the areas where data analytics is being used the list would be infinitely long, so some of the major industries include banking, finance, insurance, telecom, healthcare, aerospace, retailers, internet, e-commerce companies, etc.

Now, let us talk about the reasons as to why traditional analytics paved the way for machine learning.

**Limitations of traditional analytics:**

* **Involved Huge Efforts, Time Consuming**

The traditional data analytics involved huge efforts from the statisticians and it was very time consuming as it involved a lot of human intervention, where each job had to be done manually. This became a liability for the companies because of additional costs borne by the statisticians and, they failed to provide accurate results.

* **Static Data Model**

The models which these statisticians created were static in nature because of which the models needed to be revamped and re-calibrated periodically for making better predictions. This became an additional cost for the companies and they were struggling with the performance of the models.

* **Struggle to Manage the Exponential Growth in the Volume, Velocity, and Variability of Data**

With the exponential growth in the volume, velocity, and variability of data, traditional analytics struggled to manage and incorporate and integrate the data using their traditional methods. As a result of which their models struggled with the performance.

* **Lack of Skilled resources**

Lack of skilled resources was one of the major reasons for the downfall of traditional analytics. Companies found it difficult to get good resources and also, they lacked knowledge of advanced tools where data analytics could be done easily.

**Rise of Machine Learning:**

* **Human Intervention Reduced, Dynamic Model**

With the advent of machine learning, the challenges faced by traditional analytics were catered to. Machine Learning uses complex statistical methods and new-age tools for providing better and accurate solutions to the problems.

Using modern techniques and statistical methods we can create the capability for the machines to learn and make good and accurate predictions using complex machine learning algorithms. Modern machine learning algorithms help the companies to make better and accurate data-driven decisions where there are very few human efforts involved. The models created by these techniques are dynamic and cater to the changing variables.

For example, with the traditional analytics statisticians had to perform each step of their analysis using the available analytical tools manually. Also, the process was not streamlined and as a result of which, they had to spend too much time back and forth the initial steps. The models created were also not dynamic, suppose they created a model based on their available data, so after each month or once in a quarter they had to check the performance of the model based on the new data and make changes to their models. This proved to be an additional cost for the companies as it involved hiring a data analyst for updating and maintaining the model.

But with machine learning, the steps involved in creating the model were pre-defined and for each step, there were different algorithms. The models created based on the algorithms were more powerful and more accurate as they were built using complex statistical techniques. Human intervention was reduced a lot and as the process was streamlined, it saved a lot of time and money for the companies. Additionally, machine learning used new-age tools which made better predictions and accurate results in very less time.

* **Managing the Exponential Growth in the Volume, Velocity, and Variability of Data**

One more area where the traditional analytics faltered was working with volumes of data. With the increase in volume, velocity, and variability of data, the old tools (e.g. MiniTab, Matlab, IBM SPSS, etc) failed to manage the data. For example, if the data was structured the tools were able to manage and analyze the data to some extent (i.e. the volume of the data) but beyond that, the tools faltered. Also, for unstructured data, the tools had no answer at all, and there very few techniques or methods are known to the statisticians.

But with modern-day tools (e.g. Python, R, Knime, Julia, etc.) and techniques, we can manage the mammoth data, be it structured or unstructured, Machine Learning had a solution for each of these problems. This is what the evolution of data analytics is all about. Computing is getting better and better with time.

We had talked about the advantages of machine learning and how it is impacting our lives today, now let us talk about the challenges in machine learning.

**Limitations of Machine Learning:**

* **Large Data Requirement**

Fetching relevant data for training our model becomes a challenge. For making the model make better predictions, we have to train the model using enough data which is significant to the problem. In case there is not sufficient data available, the model build wouldn't be able to perform correctly.

* **Lack of Trained Resources**

Using different ML techniques, we get our results; but interpretation and understanding of the results vary from person to person. At times it might become a major challenge if we do not have a trained resource, even if he had applied the correct algorithm based on the person’s judgement and understanding the solution might not be correct. Additionally, the selection of the correct algorithm depends solely on the data scientist’s decision. If that person is not a trained resource, the solution might not be correct as the proper and correct algorithm was not selected. Companies do face the challenge of finding a good resource that is sound technically and statistically.

* **Reliance on the Results Obtained by the ML techniques**

We tend to rely on the results obtained by the ML techniques more compared to our judgement and experience. Suppose we build a model using different variables as selected by the ML algorithm, and the algorithm doesn’t consider a variable that is according to the data scientist critical for the model. So, such kind of situations does occur if we rely completely on ML  
techniques.

* **Treatment of Missing Fields**

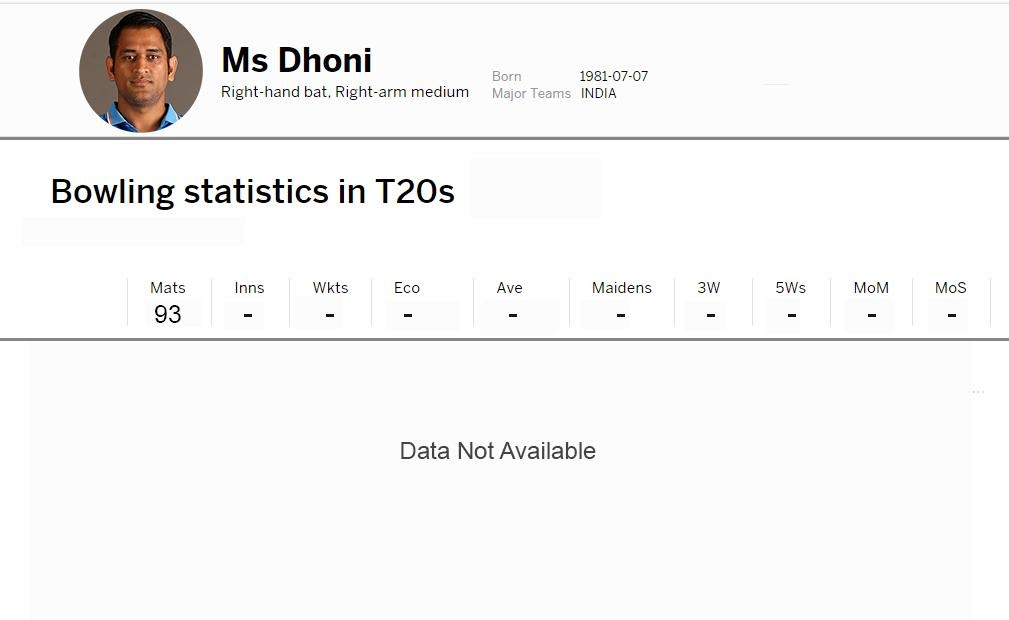
Lastly, when we have missing fields in our data, we use machine learning to replace those missing values with some alternate values. Imputing those missing values at times might bias our data and hence impact our models and in turn the results. The conclusions and inferences might change if we use the missing data, so we treat them using ML techniques selected by the data scientists. The selection of the imputing technique depends on the data scientist's judgement and becomes a disadvantage for the model.

We have talked about missing data but, what do we mean by that?

**Missing data:**

Whenever the data is not available or not present for any fields in the data, we say that the data is missing for that field. It is sometimes represented by a “Blank” or “N/A” or “n.a.” (Not Available) or even “-”. This is can be the case for both structured and unstructured data.

Suppose for example, when we look at the bowling statistics for M.S. Dhoni in T20 internationals, we get missing data. The screenshot is given below for your reference. The data is not available since Dhoni hasn’t bowled in any of the T20 matches, but when we structure that data we get data missing for Dhoni.



There might be n number of reasons as to why some of the values for a variable is missing, it might be unavailability of data or the data was not captured, etc. But as a data scientist, when we work on building a model using the datasets we always have to replace those missing values with some other value in such a way that it won’t impact or have very little impact on our model. There are different techniques available for imputing or replacing those missing values which we are going to discuss later.

**After ML, what next?**

As we are making progress in data science and computing, new-age technologies are making a breakthrough. Researchers are constantly working on next-level techniques where there is minimalistic human intervention involved and better and more accurate results. Deep learning and now AI is a breakthrough in the world of data science. It is becoming better and better where deep learning-based techniques are reducing the time and costs for the data analysis. The progress can be summarized in the diagram given below.

