

How much does the data tell us is enough?

###Artificial intelligence connect to data

After reading the article “The Revolution Hasn’t Happened Yet” by Michael Jordan. I try to think about what Artificial intelligence means for me. Most of us will think this is a way to improve human life, Artificial intelligence is better than humans, smarter and faster than the human brain. It can do more than humans, such as analyze the data and give a conclusion with the data. But we forget the most important thing about AI which is AI was produced by the human. The way the AI runs depends on the code and model that is given by humans. Sometimes it is like a baby, the human was the teacher that taught it how to study and use what kind of way to learn. The AI will learn and improve the knowledge base on the way of human teaching. So for AI, machine learning was the most important thing than teaching.

Artificial Intelligence is a technology designed to enable machines to perform tasks that resemble human intelligence. Machine Learning, on the other hand, is a branch of AI that focuses on enabling machines to learn from experience without explicit programming. Think of AI as a broad concept, and machine learning as one of the methods to achieve this concept. Machine learning achieves tasks by allowing computers to learn and adapt from data rather than through explicit programming. By providing large amounts of data and corresponding algorithms, machine learning models can automatically identify patterns and trends, making predictions or decisions when processing new data. This ability for automatic learning enables machine learning to find wide applications across various fields.

The author mentions that humans ignore machine learning but think AI is the important thing. What is the difference between them? We can image there are two humans, one of them has a brain and the other one with no brain. We called it A and B. A can learn and study new things, he or she will practice and use the skill the next time and also will learn the new thing based on the last knowledge. Then the B with no brain, only can listen to the order. If you want him or her to clean the floor, he only can do the cleaning. But he or she will not discover a new skill in cleaning the floor. The whole difference between them is the brain, so for the AI, Machine learning was the brain of them.

In the article, the author mentions that his wife was pregnant. After the doctor checked and said “There was a geneticist in the room, and she pointed out some white spots around the heart of the fetus. “Those are markers for Down syndrome,” she noted, “and your risk has now

gone up to one in 20.” ” Jordan,2019) Then the author wondered why we can say the white spots are mean syndrome. He goes to search the data, he finds out that in the last century. The syndrome baby comes out with also with a white spot, then it tells the white spot will increase the syndrome. But it ignore the most important things, Did that white spot come from the body or was the machine broken? This story tells us that the data is important to analyze the information. But sometimes we should think more about where the data come from, by the conclusion by analyzing data can be used in any condition or not.

At the end of the AI topic, by the story I tell in the last. We should set up an AI with a brain. It can analyze the data and let us use it. Also, it should have machine learning to discover new skills. In everyday life, we witness the widespread application of machine learning. For instance, on social media, machine learning algorithms recommend content based on user's interests and behaviours. In the healthcare sector, machine learning is used for disease prediction and diagnosis. In autonomous vehicles, machine learning enables cars to adapt to different traffic conditions. These examples illustrate how machine learning enhances and simplifies many aspects of our daily lives. In the before, people only think the AI is connected to computers and math. But in the future, people will use AI to do many things analyze data then advise people. It should combine health and geography.....

Clean data is not struggle work but it was a storybook

“In fact, “cleaning” is just a spectrum of reusable data transformations on the path towards doing a full data analysis.” Randy.Au 2020). Data analysis sometimes is boring because we use the same code and output the clean data. The way to clean the data is the same, but what is the difference in the clean data?

For most data workers, data analysis is connected with the data clean. What can clean data bring us, for clean data people can use it to make the conclusions and also find out some results of one thing. Just like the article Jordan wrote. If we have enough clean data, we can find out what percentage of the white spot showing up will bring the syndrome. The clean data is just like a book, it will tell you the story. You can find anything that you want in the story, and after you read the whole book you can easily conclude.

“Wikipedia's post on data cleaning does a decent summary of the big important qualities of data quality: Validity, Accuracy, Completeness, Consistency, and Uniformity. It's also got a section on “process” that's dry and academic (in a negative way) and won't help you clean any data at all.”(Randy.Au 2020). Sometimes, for the data worker, we will define what is good data analysis. We have a rubric for it, you should have each point to earn more points to let your data become better. But why we should give so many rules for clean data? As we said, clean data or we can say data analysis is like a book. We should write the book with a clear mind and let the book tell our story. The only thing to write the book is with a pen, for the data analysis pen is like the data authenticity

The author challenges the perception of data cleaning as a mechanical and non-analytical task, asserting that it is an integral part of data analysis. The act of cleaning data is described as the preferential transformation of data for better interpretability, making it an essential aspect of the analytical process. The post concludes with the assertion that cleaning data is a prerequisite for a solid understanding of the underlying data, emphasizing the importance of personally undertaking the cleaning process to avoid potential risks and challenges associated with misinterpretation or reliance on others' cleaning efforts.