The Physics of Data. Part II

The Physics of Data - Part II | Alfonso R. Reyes

Few days ago I published a post where I referred to the disappointing results we get in some data Science and machine Learning projects due to poor quality of the data. Here, when I say "data" I am referring to real-world data; data that originates from sensors and instruments in the field.

Data with sources in natural environments such as remote locations: the dessert, the jungle, the mountains, and the ocean are affected by innumerable factors. Picture those sensors for a moment, and the harsh conditions - external and internally - under which they perform their measurements. Second, think about the frequency at which you get that data. In the real world you don't get readings every second. Third, periodically you perform a **physical control test** to verify that your sensor measurements are within range. You wish you could do more of these control tests but they are costly, they consume resources and time, and tend to affect your production cycle.

Big data is not only a misnomer; it's been misleading

Remember the first time you heard about "big data" and the prediction miracles you read about it. The term Big Data" originated from the immense accumulation of input coming from the internet browsers and your clicks. Billions of users, millions of computer devices, and quadrillions of interactions tend to produce huge amounts of data. But this is not the outside world; it is a universe of interconnected devices that never stops producing data. The world of *big data* is not the real world!

The data science and machine learning of *big data* environments are ideal for any kind of algorithm, and it's been a mistake to extrapolate the ways and methods of that *ideal world* to the *real world* that produces resources, materials, goods, and energy.

It is no wonder that the world of big data achieves those high rates of accuracy in the predictions of their models. There are no forces of nature impinging on the data stream.

The real world is a universe of small and imperfect data

The world that produces resources, goods, and energy is not necessarily abundant in "big data". There might be some pockets of it and represents a tiny fraction of the data that move the industries.

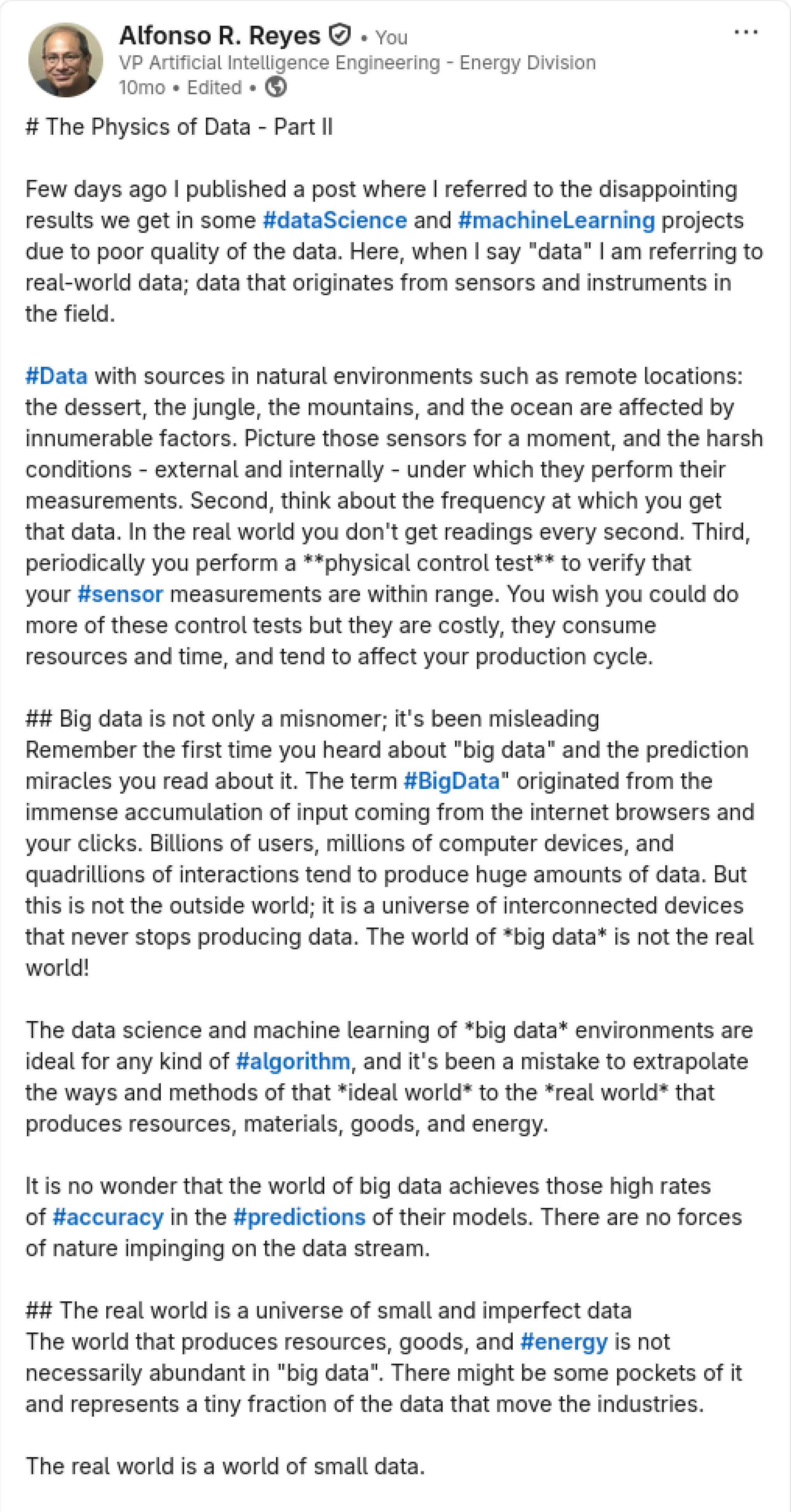
The real world is a world of **small data**.

Real world data without context is meaningless

Picture any situation where you had to work with field measurements. They are always measuring something that will tell you something about a process: a pressure, a temperature, a position, a status, a mass rate. But their readings will only have meaning if you know its function, location, and units. In general, the process is very well known, the Physics is well understood.

What is unknown is its interaction with other components, and the effect of the environment upon it. And when I say physics I really mean Differential Equations that are being neglected or ignored.

hashtags:: AI SPE Petroleum Engineering Differential Equations SciML Physics Of Data



Alfonso R. Reyes • You
VP Artificial Intelligence Engineering - Energy Division
10mo • Edited •

The Physics of Data - Part II

Few days ago I published a post where I referred to the disappointing results we get in some #dataScience and #machineLearning projects due to poor quality of the data. Here, when I say "data" I am referring to real-world data; data that originates from sensors and instruments in the field.

#Data with sources in natural environments such as remote locations: the dessert, the jungle, the mountains, and the ocean are affected by innumerable factors. Picture those sensors for a moment, and the harsh conditions - external and internally - under which they perform their measurements. Second, think about the frequency at which you get that data. In the real world you don't get readings every second. Third, periodically you perform a **physical control test** to verify that your #sensor measurements are within range. You wish you could do more of these control tests but they are costly, they consume resources and time, and tend to affect your production cycle.

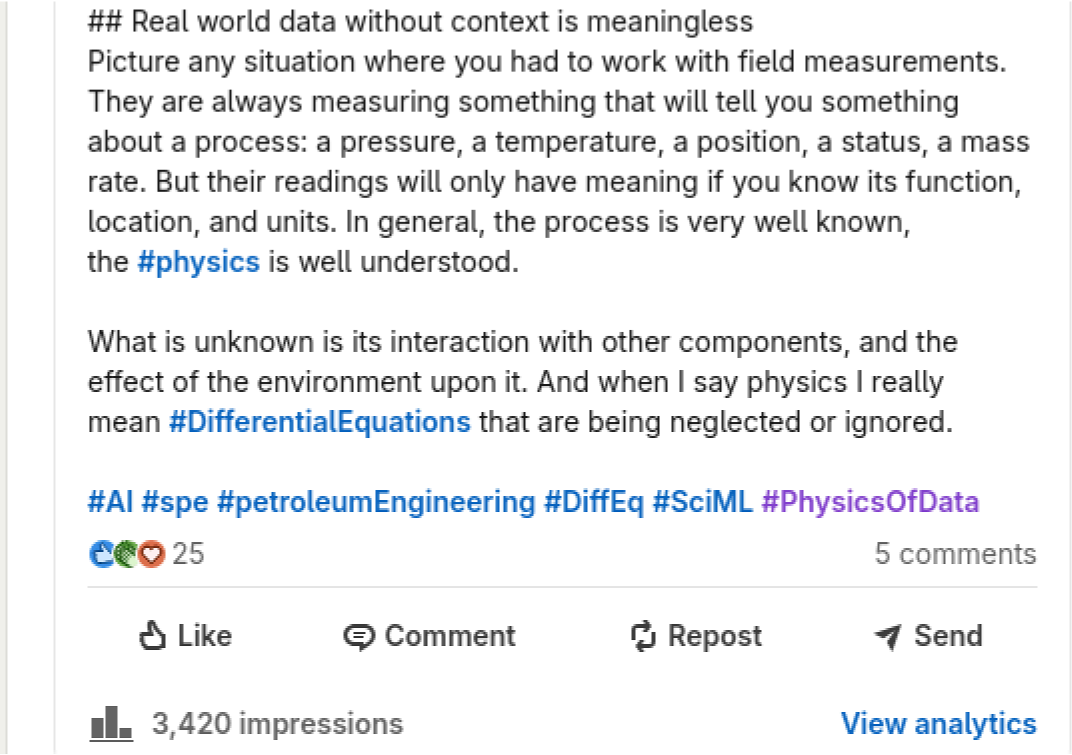
Big data is not only a misnomer; it's been misleading
Remember the first time you heard about "big data" and the prediction miracles you read about it. The term #BigData" originated from the immense accumulation of input coming from the internet browsers and your clicks. Billions of users, millions of computer devices, and quadrillions of interactions tend to produce huge amounts of data. But this is not the outside world; it is a universe of interconnected devices that never stops producing data. The world of *big data* is not the real world!

The data science and machine learning of *big data* environments are ideal for any kind of #algorithm, and it's been a mistake to extrapolate the ways and methods of that *ideal world* to the *real world* that produces resources, materials, goods, and energy.

It is no wonder that the world of big data achieves those high rates of #accuracy in the #predictions of their models. There are no forces of nature impinging on the data stream.

The real world is a universe of small and imperfect data
The world that produces resources, goods, and #energy is not necessarily abundant in "big data". There might be some pockets of it and represents a tiny fraction of the data that move the industries.

The real world is a world of small data.



Real world data without context is meaningless
Picture any situation where you had to work with field measurements. They are always measuring something that will tell you something about a process: a pressure, a temperature, a position, a status, a mass rate. But their readings will only have meaning if you know its function, location, and units. In general, the process is very well known, the #physics is well understood.

What is unknown is its interaction with other components, and the effect of the environment upon it. And when I say physics I really mean #DifferentialEquations that are being neglected or ignored.

#AI #spe #petroleumEngineering #DiffEq #SciML #PhysicsOfData