Top of Form

What limitations do we think data science has? How far do we think we can apply it?

Data Science evolution has opened up a lot of new possibilities of Predictions, finding patterns in the real world by studying historical information and past performances. Now that Big data has become much more affordable to the companies and with the faster processing speed, future seems bright that the accuracy of these predictions will only increase. But, in reality, the results can be totally different. We can expect to see some surprising behavior and results because the models and algorithms may yield completely different results than the actual one. This can happen because the Data Scientists may not able to take all the unknown variables into consideration which can eventually lead to the failure of the model.

Let us look at a very famous example on how many pollsters who used advanced scientific algorithms have failed to Predict the correct outcome of election in 2016.  Just before the election, many of pollsters, with the help of the Big Data, predicted that Donald Trump will lose the election to Hillary Clinton and the chances of trump victory were predicted between 15% to 30%. Princeton Election Consortium even predicted that Hillary has a 99% chance to win in the presidential election. Even though some statisticians like Salil Mehta has warned about the unreliability of the polls. This has posed us questions on what the limits are for a data model and if Data Science has any limitations.  Listed below are a few of them.

1. **Data quality and understanding of the data:** Task selection and specification is not the only important place where human interaction is far more critical. Human creativity, knowledge and common sense adds value in understanding and selecting the right data to mine. Also, if there is lot of missing data, it may not be able to help the model to do its job accurately.
2. **Beware of biases:** We should remember the fact that data do not represent objective truth. Humans incorporate their biases, opinions, beliefs into the data collection systems and the meaning of data can change by their own beliefs. In the pollster’s example, many, who were supporters of Hillary, were not ready to question the results that favored their candidate.
3. **Models have only a half-life:** Data scientists spend a lot of time on building and nurturing their models with a hope that it’ll fit in the real world. But the fact is that even though these models work perfectly alright in the beginning when they were designed, they will lose their effectiveness overtime because the world will change over time. There may be many factors that may have changed during these times and all these variables should be captured and models should be refined to make it relevant.
4. **Probabilities, not guarantees:** We should remember that the predictive models that we use gives the probability of occurring of an event but not guarantee it. The technology always needs the experienced scientists to understand these gaps. Only by applying the experience, one can make informed decisions after considering these probabilities into consideration.
5. **irrational behavior** – most of the times, people may not do what they say, and this irrational behavior can result in the inconsistent results.

Data science involves the judicious integration of human knowledge and computer-based techniques to accomplish which neither of them could on their own. Data is the fuel, and algorithms borrowed from the tool kit of artificial intelligence, notably machine learning, are the engine.[3] But there’s still a lot of emphasis on the driver’s seat which humans should take and never take their eyes off the road in this journey.

Bottom of Form