EXPECTATIONS FROM DATA SCIENCE

Hello all,
Today, I would like to make you all commit something to your memory at an early stage – something which took me a long time to realize.

DATA SCIENCE IS NEVER DEFINITE. IT IS EXPERIMENTAL.

i.e You never have just **one** "right" answer to your data problem. For example, in Software Development (coding jobs), if someone says to write a code to perform XYZ task – there might be a "definite" code to do it! In Data Science – you don't!

Basically, any data problem which you're trying to solve is like a NUT and SCREW DRIVER problem. You want to find a **right screw driver** to open up the nut in front of you. And you never get it right just by looking at it. You TRY various screw drivers (models/methods) and then try to come up with an approximate one which opens it up! Basically JUGAAD.

This is known as **Hypothesis driven testing**. Basically, you **HYPOTHESIZE** that my problem "might" be solvable via METHOD XYZ. So, you try with that method (**EXPERIMENTATION**), then **EVALUATE** the results and then you **REJECT**/**ACCEPT** the hypothesis!

And there's a catch! There exist (almost) infinite methods for each specific case. How? There's a decision making involved in each step of Data Science. For example, in **imputing missing values** steps - one might choose to fill values by mean, while someone else by choose to fill it by -1. In outlier analysis, one might choose to delete outlier variables while one might keep them as it is. (And none of the decision is wrong!). So, in a multistep process you can clearly see that there's a decision involved.

By the end of a Data Science project which say has 10 steps and 2 decisions at each step, you'll automatically have 2^10 decisions to make! Basically, you have **2^10 experimentations** to make (in practice, you don't exhaustively experiment each one of them, you smartly select a "vital few"). And that is the beauty of Data Science. No one ever solves a question correctly. Each person tries to solve a problem by a different methodology. And there is not a SINGLE right decision. Each one can come up with different set of decisions yet correct answer.

CONCLUSION

This is the reason that throughout your career you will never come across a **COMPLETE DATA SCIENCE COURSE** which teaches you all the cases completely. They teach you the basic standard methods that is used in almost all cases. However, in real projects each case is different.

And when you solve many datasets, you'll come across many such different cases – you'll then search over the net on how people have previously solved them and then apply to your problem! **Therefore, a Data Science course can just teach you the basics**. And these courses will teach you just 40% of actual Data Science. The remaining 60% comes by practice, observations and reading solutions of other people.

Cheers!

Rishabh Agrawal