**Sub optimal Solution vs The Best Solution for Your ML Project – Which one is Better for Your Business?**

**WHY? Is this not obvious? Not Really…**

**The Big Bang**

In a real-life Data Science or Machine Learning Project, the Data Scientists, ML Engineers, and Developers work on a project with each other in a team. This layout of work is a lot different from what we learn or do in our degree programs or university research projects.

Generally, it is observed in organizations working on data intensive problems, or machine learning projects than the problem statement decided, and the probable solution discussed in the initial meetings of the project are different from what the finally presented product or data solution at the time of delivery or deadline of the project. This difference is not very big, as in an apple was asked but a banana was delivered, but it is surely similar to the difference between a red apple and a green apple, or an apple and an apple juice. The final presented solution is ‘kind of’ what everyone agreed, but not really the one what everyone agreed. This difference in the decided objective of a project and the final product means a lot in the corporate or business world.

This unexpected result is never intended but often witnessed, and this is the ‘Big Bang’ in Data Science world. But why or how does this happen?

Now, what happens here is that when you ask a good problem solver to solve a problem, we often go on for looking the best solution , we think about all the possible edge cases, and all the possible issues. We always go for solving the problem in one big iteration, which is kind of working in the dark, before getting feedback from our supervisors, as we want to come up with the solution, which is robust and with all the cases, as this is also a reflexion of what kind of work we do as an employee. This delayed big feedback loop is often too long to be useful and the reason of surprises at the end. The spot we end up reaching through this process is not the one we were supposed to be or what we decided at the beginning. Though, this is definitely a solution, but not the one we were looking for. This practice of solving a data intensive problem in one long iteration, influenced by many factors is a root cause of facing the ‘Big Bang’ in your Data project.

Can we solve it?

Of course, we can solve it. This article discusses about the process we data scientists, or ML engineers should follow while working on the data intensive projects, solving a data problem which would help us avoid the Big Bang in our project.

Let’s Dive In

**The Sub Optimal but Optimal Solution**

As discussed, the delayed feedback cycle makes us work on the solutions in the dark with no idea where we are heading, is it the right direction. And often, these feedbacks are old and un-useful as not given on time it was really needed. This results in us developing a solution which is not what we wanted at the beginning of the project and are surprised and disappointed at the time of final submission.

The surprises experienced at the end causes various problems for us as well as the business. Because a lot of money, time, and manpower is invested in the project only to get the most optimal but un-wanted solution at the end. This is much more crucial than it sounds. So, how to deal with this?

* One Solution

One solution to solve the problem of delayed feedback is to make everyone work more, for longer hours so to get the feedbacks more frequent. Which means, everyone is working double the work hours. This is not the optimal or sustainable solution for the problem and would create many other issues instead of solving the problem of delayed feedback loop.

* Other Solution Or opt for sub optimal solutions and keep the feedback loop smaller and frequent.

Another solution to solve the problem of ‘delayed feedback’ is to opt for finding and trying the sub-optimal solution to solve the problem statement and submit it for the feedback. This would allow us to reduce the time of the feedback loop (which is the first-hand problem we see here), and also allow us to keep checking the path we are walking for solving the problem.

This procedure can allow us to know what is working and what isn’t very frequently as we climb up with different solutions to find the best solution, instead of finding the best solution first, which is actually best in our views, but no guarantee if it is the best as per the business perspective, as we aren’t getting the feedback as frequently as we need to.

This brings us to the mental model we want to discuss in this article, which is ‘Wabby Sabby’.

**Wabby Sabby – The idea to deal with the ‘Big Bang’ in your ML Project**

‘Wabby Sabby’ is a concept from the Japanese philosophy, in contrast with the Greek philosophy which is about the perfection, symmetry and balance. The Japanese concept, however, is based on the idea that ‘everything is the world is imperfect, impermanent, and incomplete and therefore, there is beauty in creation’. This is the mental model, we can use as a problem solver to solve both the problems we have here, the one about the delayed feedback loop and the other one about the surprise at the submission table of our data project.

What this really mean is our perspective as a problem solver is that we can use a non-perfect solution to solve the problem. This would make us reach the result of the solution for the problem relatively faster and allow us to ask for feedback relatively faster to further improve our solution and eventually reach the best and expected solution to the problem.

By the feedback here, I mean not only the business feedback but also the feedback from the domain of our problem, feedback from the environment we are working in, feedback from the data about how well does it fit with the solution, what exceptions are found and what part of the data are not turning up with the solution as expected. The ‘Wabby Sabby’ thinking also gives us the time to sharpen our solutions. We are free to try the sub-optimal solutions, and this allows us to explore the different directions and dimensions of the data and our solution finding new faces of our problem statement. This way allows us to get more information about the problem and the data much faster, helping us to find the actual best solution for our problem, the one which also aligns with our business perspectives, environment, and the data.

**Benefits of Wabby Sabby**

As discussed, the ‘Wabby Sabby’ thinking allows us to explore different dimensions of our data and the problem. By this, what I really mean is that, whenever we are working with data, we are trying to extract some information from it. And this information is an aggregated format of our data only. Now, very often, the data has some data points which we call outliers, as they are quiet far from other data points when plotted on a graph. One practice of dealing with this is to get rid of all such noises in the data and get the best polished data we can to formulate the best optimal solution for our problem.

But I have experienced this in practice that these outliers plays a role, they also tells a part of the story possessed by the data. One dimension of the data lies behind these outliers, and we should explore that before getting rid of them as they might have an important insight for us.

The project workflow in organizations are very unique and diverse. One week we are working on extracting, cleaning, and storing the data, and the other week we have all the time to test the machine learning models on our data and analyse the results.

The concept of ‘Wabby Sabby’ thinking here is useful as it allows us to explore different dimensions of the data, using different models and checking the results, finding which dimension shares a valuable insight and which ones are not very useful for us.

The way of approaching the problem and finding the solution by intentionally picking up the non-optimal or should say, the sub-optimal solutions must be followed and encouraged among data professionals, as it allows us to reduce the uncertainty in our solutions significantly and also shortens the feedback loop period as we get much faster feedback loops which we can integrate into our solutions and move forward altogether towards the better solution eventually finding the best solution.

**CAUTION! While using Wabby Sabby**

Though ‘Wabby Sabby’ concept has a lot of plus points for optimizing the process of our data projects, there are also some cautions for using this concept in your project.

Zoom out and check + evaluate the imperfections. How much does it matter and how time does it take to make it better.

Throughout the project, there are various places where we need to ‘zoom out’ the situation at hand for a moment to scan the complete situation and check what imperfections do we have with us, how much these imperfections matter to us in formulating the solution and how much time do we need to improve these imperfections in order to get a better result.

Also, there are many places, or we come across many data projects or problem statements where we do not really need the perfect solution as the imperfection does not have that big of an impact on the result. In such cases, the speed at which we solve the problem really makes a big difference. Here, the ‘Wabby Sabby’ thinking, and zooming out technique helps us a lot to find the solution we need, and we want.

“Don’t forget your Assumptions”

As we realize that we make a lot of assumptions and decision along the way of exploring the data, and formulating the solutions, it is very crucial to keep a note of all the assumptions made which actually led us to the solution we have.

We can make a lot of shortcuts, and we need to be careful about the ‘Wabby Sabby’ decisions made by us on the go and shall they be passed on into environment.

Remember the assumptions made, the short cuts taken and the consequences of these decisions if not treated well before sending the product into production.

Communicate well what is the solution, what are the assumptions, and what need to be done to make the solution work.

Another important thing to keep in mind, and never forget is to ‘Communicate’. Communication is one of the most important parts of a project, and it is even more important if we are approaching our problem through a ‘Wabby Sabby’ thinking. Apart from remembering and recording your assumptions and decisions, you also need to communicate these to officials at meeting in order to bring all team members and managers on the same page. Share what you are presenting, what are the assumptions make, tell them that the solution is imperfect, and you and your team are going to improve it.

There can be times, where the businesspeople, or client or the user might find the solution best, but you know that it’s not. Share this with them along with the reason supporting your statement that the solution is not the best. Keep your reason ready to share, and also ensure that the reasoning is logical and understandable.

Hidden imperfections, mismatched thinking, and perspective towards solution among you and businesspeople can lead to unwanted results later. Share the consequences and other details about the presented solution. Discuss the impact of the issue, time required to fix them and probable change in the results.

Keep good explanations regarding your sub-optimal solutions about why the solution is not perfect and what is to be fixed. In the case, you do not find any logical reason for sharing that the presented sub-optimal solution is not best, then the sub-optimal solution is the solution we actually want.

Be vigilant about all this while using the ‘Wabby Sabby’ concept and see the much-optimized progress and success in your data projects. Do not forget to take care of the cautions mentioned above to avoid unwanted consequences of using the ‘Wabby Sabby’ thinking.

**Conclusion**

Concluding and summarizing the blog we can say that the data projects are a very big rock we need to break. We cannot break this rock in one big hammer punch, rather using several hammers at several different spots in a particular rhythm would definitely break the rock much faster and in the way we wanted. This is the controlled approach to solve problems as big as the rock. The concept we can use to solve massive data intensive problems is the famous Japanese philosophical concept, ‘Wabby Sabby’ which sees the world and life as an imperfect, incomplete, and impermanent, which is why the creation is beautiful. This concept allows us to approach the problem from different directions, exploring different dimensions and getting multiple solutions faster allowing us to analyse what is working and what’s not.