

Let's Bake a Data Science Cake

Attributed to Aritra Chattaraj, Engineering, Extentia Information Technology

By TimesTech - April 11, 2019

94 0



Cake – the four letter word responsible for a myriad of emotions from joy and all the way to euphoria. With its vast ubiquity across countries and cultures, it makes us wonder what all cake might have been responsible for.

- Was it responsible for making birthday parties a (~38) billion dollar industry? Absolutely.
- Was it responsible for gainful employment across cities? Probably.
- Was it responsible for the French Revolution? Can't say.

What we can say, however, is that cake is exceptional at making our lives better. You know what else has been claiming to make our lives better and more convenient? Artificial Intelligence (AI), and the buck doesn't stop there.

AI is an exciting field and an even more elusive one once we get to it. To make it implementable, we have a group of statisticians with expertise in the domains of business, economics, programming, and machine learning, driving a brand new 'data revolution' and an even newer profession – **data science**. Although we are far away from the Skynet era of technological advancement, data science is the one discipline dedicated to making data-driven decisions in order to drive business growth and social welfare. While we cannot comment on the exact future of AI, we can conclusively discuss what data science means. So let's get baking!

Cake – we all love it but what would we do to bake it to perfection? Well, I suppose the answer varies from person to person but taking this variation in our stride, let us change the question – how would a data scientist bake a cake?



For a **data scientist**, a more important question would be, "Does anybody even want to eat cake right now?" and that is where our journey begins with the first question or problem statement.

A baker would also ask questions like that before they start their bakery business but to what extent and to what precision? A data scientist would start exploring this question by

understanding the market demand, the consumer preference, the time of the day, whether or not their consumers are suffering from Celac. For instance, our data scientist finds out that nobody with a fondness for apple pie would like to eat cake at 3 a.m. Inference? 3 a.m. is probably a terrible time to bake a cake.

Next, a baker would already know their select ingredients for baking a cake. A data scientist, in contrast, will employ **descriptive analytics** – a number of statistical measures to derive insights from past instances – to understand which ingredients have been not only been historically significant but

also most preferred. For instance, flour is a must for baking but gram flour is probably not the best ingredient for cake since, historically, people have been using gram flour to bakedhoklas.

With the ingredients all in place, the data scientist would now employ **predictive analytics** – statistical modeling or **machine learning** techniques in order to build a model that can be used to predict some value (numerical such as prices or even categorical such as true or false) – to predict the number of cakes that need to be baked for their whole family as well as the size for each cake. This means, hypothetically speaking, we either have four small cakes (1lb each) or two big cakes (1.5lb each). However, it is the end of the month and our data scientist is low on cash which makes it imperative for him to optimize on his costs. The question that is posed then is, "Is it cheaper to make four small cakes or two big cakes while also accounting for the eating habits of all family members?"

Using **predictive analytics**, our data scientist predicts the two-big-cakes option to be the most optimized of all solutions before moving forward to make it. Congratulations! Our cakes have finally been baked. The process, however, is far from over.

In a few hours, we realize that only one of the two cakes have been consumed. Baffled, our data scientist sits down to understand why – employing a new kind of technique, **diagnostic analytics**, which, as the name suggests – questions the reason for why things did not go the way they were supposed to. Using advanced analytics, they find that instead of chocolate syrup, they had put balsamic vinegar in the uneaten cake mix. After having lamented at this unfortunate situation, our

data scientist starts thinking – have we exhausted everything in our data scientist's toolbox? Well, not quite.

Prescriptive analytics is a technique to not only diagnose a problem but also to suggest solutions for resolving the same. Applying this, our data scientist analyzes the big, brown cake to understand what remedies can be applied to fix this mistake. Unfortunately, as it turns out, the only 'prescription' is to dump the cake and start anew. This leads us to an interesting point – a faulty analysis can lead to disastrous results and cost us a lot of money.

Would our neighborhood baker go through the same processes? Probably not in the same way our data scientist did. What was really sad, though, was that they never got a chance to eat the cake. As the saying goes, – you can't have the cake and eat it too.

Now that we've explained the basics of **data science** with a tasty example, let's turn our attention to its application in an organization.

For more information please visit at : www.extentia.com