# Lifetime Lessons: 20 Things Every Data Scientist Must Know Today



## Introduction

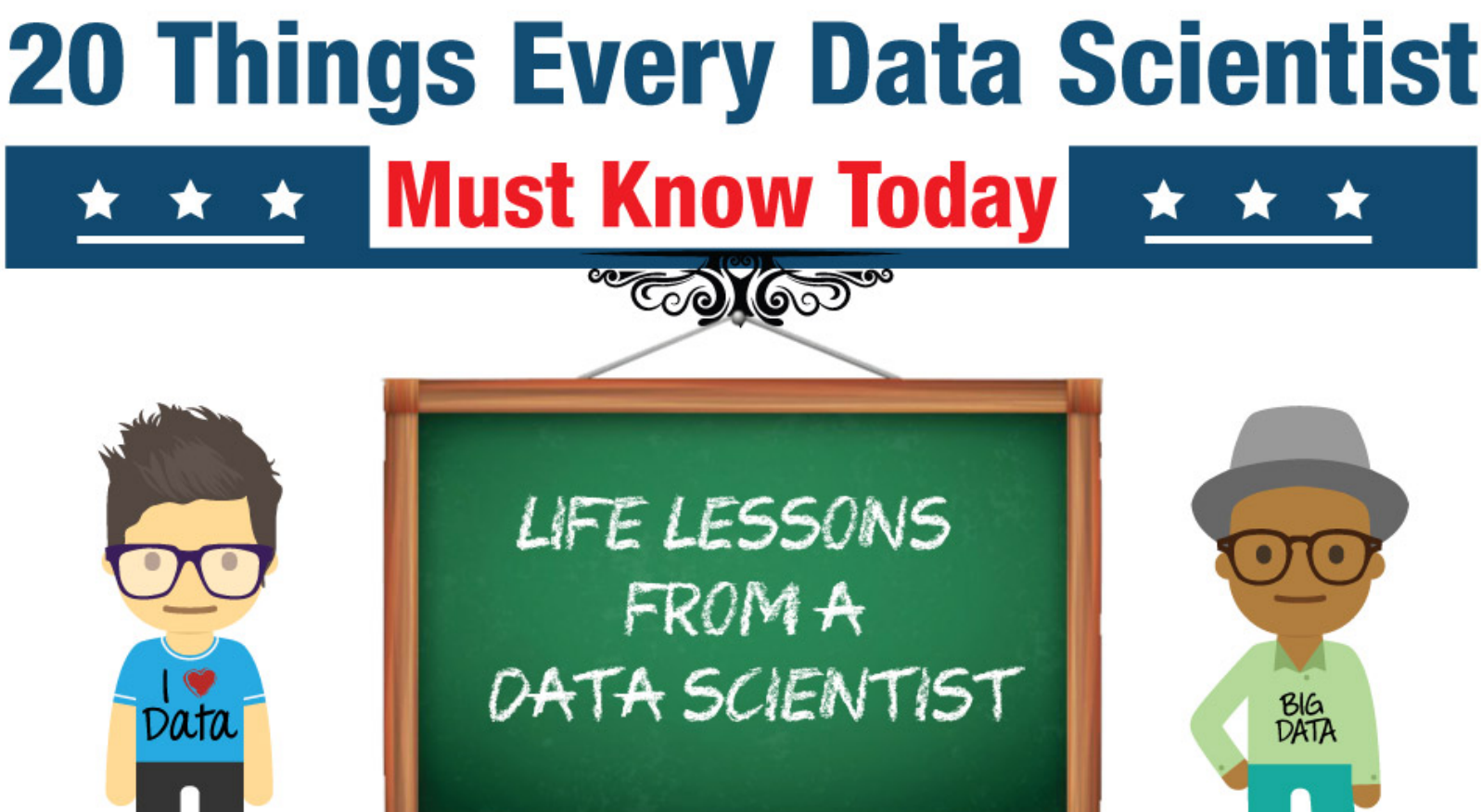
I've spent close to a decade in data science & analytics now. Over this period, I have learnt new ways of working on data sets and creating interesting stories. However, before I could succeed, I failed numerous times. Success doesn't come easy!

How did I succeed? The answer is simple. Every time I failed, I said to myself, 'Let's take one more step'. And I managed to travel a long distance. I learnt statistics, data mining, SAS, R, Python, Machine Learning on the way.

I confess that, in last 10 years, the methods of predictive modeling have become faster. Data is becoming larger than ever. We faced constraints when faced with Big Data. But, people came out with several big data technologies.

It's overwhelming to see how the things have changed. But, there would still be many who are lagging to catch up with success in data science industry.

Hence, I decided to share 20 things which experience has taught me in the last 10 years. Hope you find them useful. The idea is to help people, who don't have a mentor to provide them this advice all the time. So go ahead and read!



# Technical Lessons

# Non-Technical Lessons

## 1

## Master 6 Algorithms

Machine Learning has many algorithms. If not all, you must master Random Forest, Logistic Regression, Neural Networks, FTRL, XGBoost, & SVM. If you practice them well, you can surely provide useful solution to data problems.

## 2

## Structured Thinking

Doing data science without structured thinking is like exploring a country without its map. It is a gateway to your success in data science. Companies can't teach you this skill. You need to masterit on your own.

## 3

## Feature Engineering

Young folks in data science rely too much on ML algorithms. Don't be under an impression that algorithms can take away your pain. They certainly can't extract all features for you. So, take a step back. And, learn the art of creating new variables in the data.

## 4

## Develop Business Understanding

You can't come up with a data based solutionwithout understanding the industry. Each industry has a different trend, behavior and driving factors. Make sure you understand them before working on data.

## 5

## Ensemble Modeling & Boosting

Learn the art of combining models and

## 6

## Stakeholder Management

70% (upto) of data science projects fail

algorithms to boost model accuracy.
This will allow you to create models faster.
Make sure that you must understand the mathematics behind it.

because of various reasons. The most common being, lack of consensus between analytics team and stakeholders. Hence, make sure you keep your stakeholders on the same page while executing your project.

## 7

## Data Cleaning

Learn to deal with missing data, bad data & outlier values. You wouldn't realize when these unwanted parts would interfere in your model accuracy. I've seen people taking data cleaning for granted. But, performing this step exclusively can make a big difference.

## 8

## Story Telling

I understand building predictive models is not an easy job. But, presenting it to stakeholders is even harder. I've had tough time presenting these models to non-technical people. But, I learnt a trick. The trick is, don't show just numbers. Rather, show charts, patterns and create a story out of them.

## 9

## Cross Validation

90% of the time your model will suffer from overfitting. You might feel contended after achieving low error values. But, that's an illusion. Your model wouldn't do well on unseen data. Master this concept as quickly as you can.

## 10

## Networking

You can't become successful alone. In order to succeed you'll will need support of your connections to meet exciting opportunities. In data science industry, networking rocks. Thus, whenever you get time, try to meet, talk and socialize with data science professionals.

## 11

## Time Series Data

Companies usually work with Time series data. You must understand time series concept well. You can't create models on time series data with your commonly used approach. Instead, you must learn to work on techniques like ARIMA models.

## 12

## Implementation Matters

You've built a predictive model with highest accuracy. Great! But, what use it would be, if it can't be implemented in real life.
I've faced this situation. And, I wouldn't want you to face it too. Hence, focus

on models which are practically sound.

## 13

## Learn Python

I started with SAS, R and now Python. Python is easiest to learn. It is best suited for machine learning. It takes less time in code execution. And you'll find a huge universe of python codes available for practice at Github. If you are a R user, sooner or later you'll make this move.

## 14

## Work to Learn Not to Earn

If you chose data science because it's a lucrative industry, soon you'll find this job monotonous. I've seen people failing because they got bored of their work. Hence, learn to love data, numbers, challenges to become successful in data science.

## 15

## Deep Learning

Deep Learning is our future. It emanates from machine learning research and influences products like Cortana, Siri, Google DeepMind, Antiviruses and many more. If you wish to go beyond machine learning, deep learning has to be your next destination.

## 16

## Read Books

This is the most important lesson I've learnt in life. Invest your time in productive things. If you don't find any, pick a book and read. This will not only improve your grammar, but introduce you to a world of knowledge and possibilities. I started with 'Atlas Shrugged'. Read it 3 times.

## 17

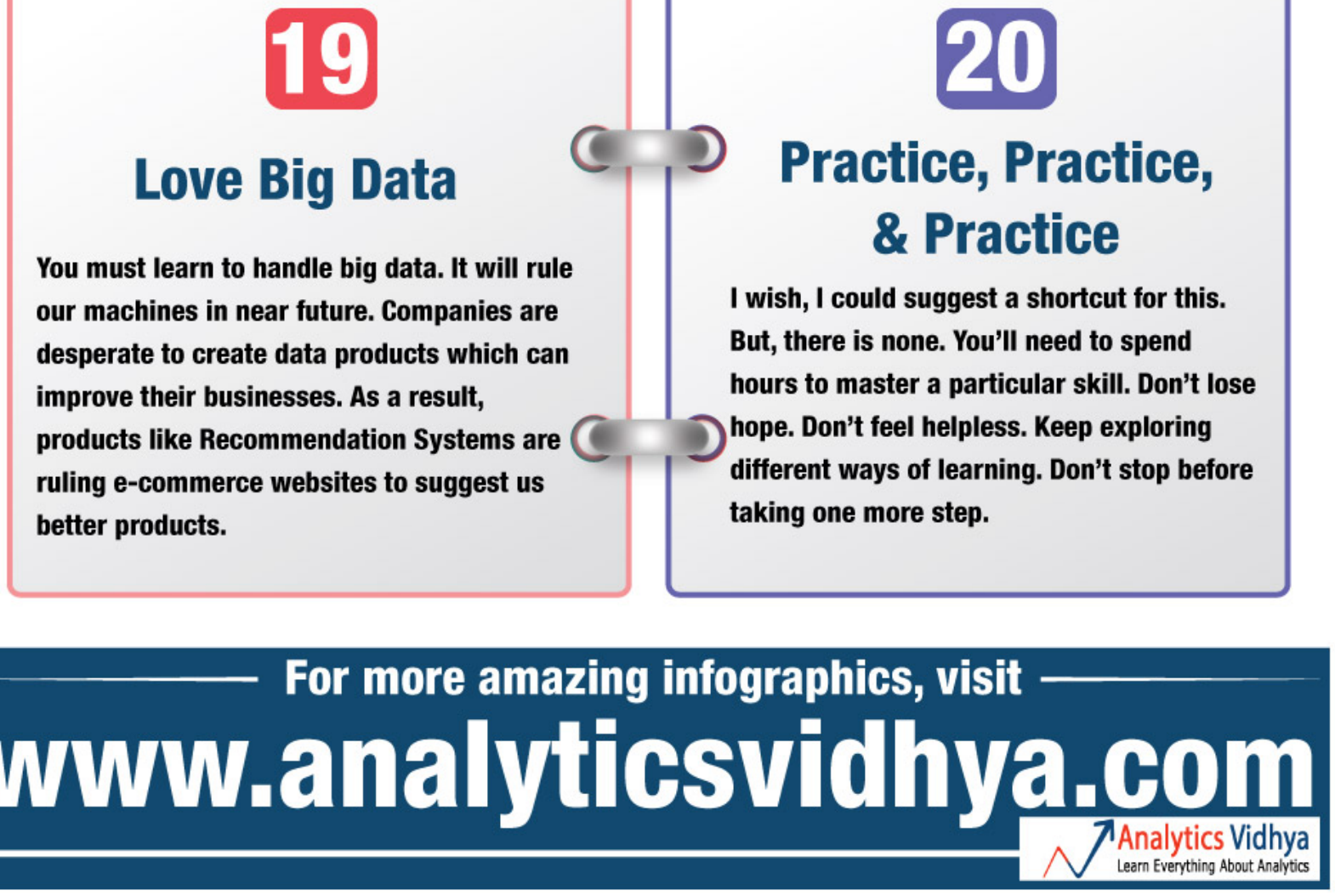
## Cloud Computing

Cloud is our future. We are deluged by data. Investing in machines upgrade will soon become costly for companies (if it isn't already) as well as individuals. Don't work hard, instead work smart. Invest in cloud computing. It's economical, scalable and faster. You can start with AWS free tier!

## 18

## Competitions

You must participate in data science competitions. To become successful, you need to step out of your comfort zone. There is a lot which you will learn, explore and practice. Participating in competition will help you gain confidence, learn new techniques & provide much needed exposure. Bonus: You will network too!




## Some Useful Resources

Learn Python

Learn Ensemble Modeling

Learn Boosting Algorithms

Learn Machine Learning Algorithms

Learn k- fold Cross Validation

Learn Feature Engineering

Resources on Neural Networks and Deep Learning

Master Structured Thinking Skill

**If you like what you just read & want to continue your analytics learning, subscribe to our emails, follow us on twitter or like our facebook page.**