35 DATA SCIENTIST QUALIFICATIONS AND SKILLS NEEDED TO SUCCEED

INTELLSPOT.COM

1

R
R is a programming language – an excellent tool for unlocking the patterns in large datasets

2

Python
Python is an absolute hit – a general purpose language, that is broadly used by the data science community

3

Java

Java
This popular general purpose language allows integrating data analysis directly into the codebase

4

SQL

SQL
SQL is efficient at querying and manipulating relational databases

5

sas

SAS
It provides a great range of statistical functions with a user-friendly GUI that helps you learn quickly

6

MATLAB
It is a numerical computing language that is a fast, stable, and works with solid algorithms for complex math.
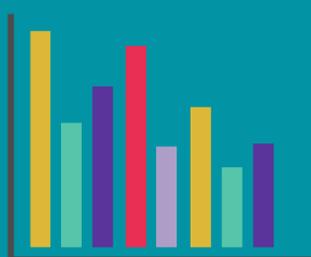
7

Scala
Scala is becoming a preferred weapon for those doing machine learning at high-volume data sets and creating high-level algorithms

8

Julia
It is a high-level dynamic language for programming that aims to meet the needs of high-performance numerical analysis

9
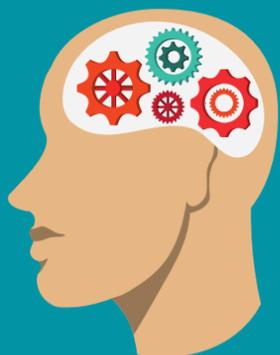
Discrete vs. Continuous Data
Many data mining decisions depend on whether the basic data are discrete or continuous
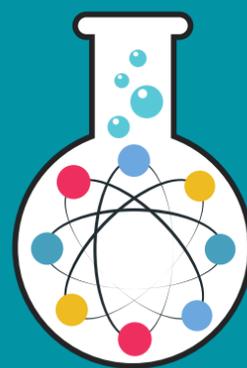
10

Binomial Distribution
The binomial concept has a core role for defining the probability of success or failure in many data science events

11

Regression
The regression models are the oldest and widely used supervised machine learning algorithms for predictive analysis

12

Hypothesis Testing
Hypothesis-driven thinking in data science is something you should be familiar with

13

Bayesian Thinking
Bayesian modeling is an extremely powerful suite of tools for modeling any random variable

14

Machine Learning
Machine Learning is a constantly growing area that is used when credit scoring, placing ads, stock trading and for many other purposes

15

Markov Chains
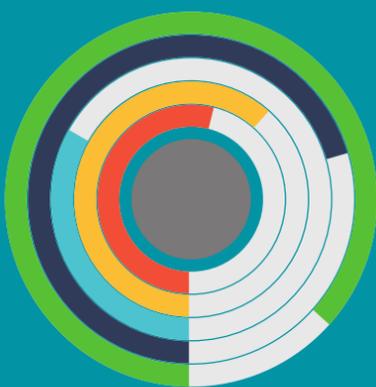Markov chains are simple methods to model random processes in a statistical way

16

KNIME
KNIME is a software company that offers an open source analytics platform for data reporting, data mining, and predictive analysis

17

Weka
Weka is machine learning software provided by The University of Waikato. Weka combines machine learning algorithms for data mining tasks

18

RapidMiner
RapidMiner is a software platform for data scientists to help you build predictive models faster. The tool unites data prep, machine learning, and predictive model deployment.
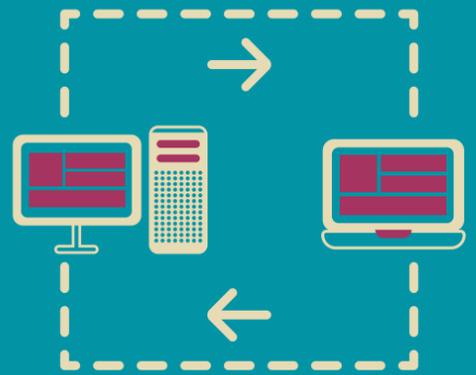
19

Apache Hadoop
Apache Hadoop software library (one of the best big data tools) is a framework, written in Java, for processing large and complex datasets
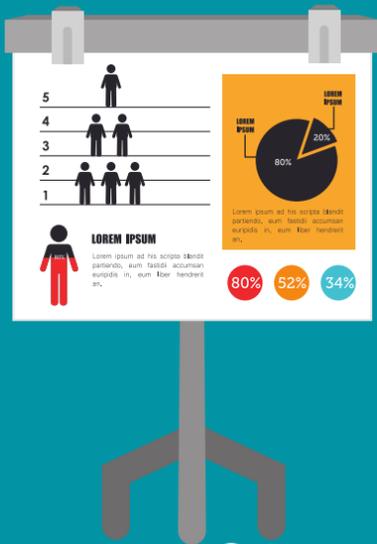
20

Apache Spark
Apache Spark is a unified analytics engine for large-scale data processing, a cluster-computing framework for data analysis

21

DataMelt
DataMelt is free graphing software that is used for numeric computation, statistics, symbolic calculations, data analysis, and data visualization

22

Apache Storm
Apache Storm is a free and open source distributed platform for real-time analytics

23

TensorFlow
TensorFlow is an open source machine learning framework for everyone – from students and researchers to data science professionals and innovators

24

Tableau
Tableau is business intelligence software that helps people see and understand their data. It allows you to connect and visualize your data in minutes, combine multiple views of data to get richer insight, and share dashboards on the web

25

Google chart
Google chart is a powerful, simple to use, and free solution for visualization of big data. It is totally free and has a great support from Google.

26

D3 Data-Driven Documents

D3.js
D3.js (stands for Data-Driven Document) is a JavaScript library for manipulating documents based on data

27

JupyteR
JupyteR is an open-source project allows you to analyze, visualize and real-time collaborate on software development across a variety of programming languages.

28

Orange
Orange is an open source data visualization and machine learning solution for everyone – from newbies to experts.

29

Communication Skills
Communication skills can help you build trust and understanding, which is incredibly important for those being stewards of the data
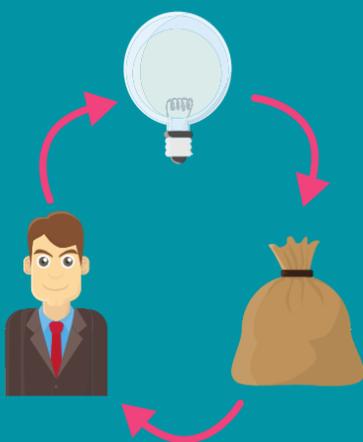
30

Data-driven Decision Making
Today, the whole management world talks about how to create a successful data-driven decision-making process and models in business to improve results

31

Domain Knowledge
You must know the business relevance of the algorithms and statistical models at hand

32

Teamwork
Every data scientist should be a team player. They are deeply involved in a company at different levels

33

Intellectual Curiosity
You need a passion for work and passion for finding patterns and answers to business problems

34

Good Data Intuition
Once you've exposed yourself to enough data work, you'll develop your data intuition to know the right questions and find the right answers

35

Project Management Skills
Most data analytics work is project-based and have to be handled effectively. That is why project management is a key skill in data analytics