

The main tools of a data scientist

[Technology](#), [Computer](#)



It is quite tempting to assume you are a computer savvy simply because you have used it over a long period. This should not be the case because there are some skills you must get absolutely correct in order to be considered a computer-literate. However, becoming a computer expert is not easy. It usually requires determination and high levels of focus. If you are a newbie and wish to become a data scientist, familiarizing yourself with python, a programming language, would be a great place to start.

With that being said, let look at key data science tools that you should master.

Rapidminer

This is a data science tool developed by Rapidminer company which allocates an integrated environment for machine learning, data preparation, predictive analysis and text mining. Its key area of application includes research, training, education, rapid prototyping, application development and modern commerce among other fields. Additionally, Rapidminer elaborates every step of the data mining, which allows optimization of advanced queuing mechanism. The incorporation of Rapidminer Radoop software eliminates the complexity of data preparation and the computer learning procedures.

DataRobot

This is a platform developed solely to provide machine learning for data scientists within a short period. The technology is developed to cater for the shortage of data scientists by altering the speed and economics of predictive analytics of a machine. This is made possible as the Datarobot is able to

search through millions of possible combinations of pre-processing steps, algorithms, features, transformations and tuning parameters to deliver the best models for any dataset and the predicted target.

The invention of DataRobot cloud has made it extremely easy to create a world-class prediction model. Additionally, the partnership of the RobotCloud with Web Services (AWS) has enabled the software to be more robust, flexible and highly secured. This has resulted to higher demand by potential customers.

Alteryx

This is an American computer software company located in Irvine, California. Its products are used for data science and analytics both by business analysts and data scientists. This has brought light in these industries as the Alteryx is able to break data barriers which is a big game changer!

Qubole

This is a well-detailed data platform which is self-managed, optimized and is capable of learning from your usage allowing the data scientists to focus on the productivity of the business rather than the platform. The ability to control itself means Qubole has unbeatable agility and flexibility.

Paxata

This is an automatic, flexible and intelligent self-service data preparation software which transforms unprocessed data into readymade and reliable information instantly. The platform is capable of extracting information from any source, many it be a cloud or any information environment. The results are highly accurate and can be trusted. Paxata has partnered with Amazon

and cloudera and can be easily connected to BI tools, Microsoft excel, Salesforce Wave, Qlik and Tableau.

Trifacta

The main product under this category is Wrangler. Its main purpose is to help data scientists clean and come up with a wide range of data more quickly, which have a remarkable accuracy. By importing your information in the application, it will automatically structure and organize your data. Additionally, you can make suggestions to the software such as transformations and aggregations to suit your specifications. The agile Trifacta data wrangling engine makes it possible sharing the process of structuring, exploring and analysis of datasets for faster and more accurate results.

Lumen Data

This is the leading provider of information management solutions for enterprises. Having strong knowledge in implementing data persistence, predicting systems, data strategy, data lakes, data governance as well as quality implies that Lumen Data has successfully delivered a good plan, implementation process, integration, maintenance, and training services to a good number of newbies.

Feature Labs

By using FeatureLab, one is able to utilize artificial intelligence and machine learning. This is important as one is able to deploy new products and (or) services, identify critical points, and understand what the data analyzed says

about the future of their investments as FeatureLab is a predictive platform made to analyze data automatically while following the right strategy.

Bottom Line

The knowledge concerning data science in the current world cannot be ignored. This branch of knowledge has greatly developed across major industries including finance, travel, government and energy. More importantly, universities have begun acknowledging its role of data science tools and technology. Additionally, having this knowledge will not only improve your skills and knowledge but will give you an upper hand when job opportunities strike.