

# MIT has Produce Data Science Machine that is Replacing Human Intuition for Big-Data Analysis

By **Muktasid Adib** -

October 17, 2015



Researchers from the Computer Science and Artificial Intelligence Laboratory (CSAIL) at Massachusetts Institute of Technology (M.I.T) have successfully created a machine that is looking out to replace the need for human intuition out of big data analysis by letting the ‘super-computers’ choose the feature set used to identify predictive patterns in the data.

Even though it is a given that super-computers are awesome in crunching numbers and figures in comparison to human, but the question always remained – can artificial intelligence choose the most optimal path to begin in the first place? That was always the key, because computers will have to go through each and every set of possibilities, even if it is a ridiculous one – while humans are tremendously efficient in this regard as we can filter out decisions that do not make sense even

though it is still a theoretical possibility. Well that is what researchers from MIT have proved as their machine beat the human counterparts in a data test.

The prototype of the software, named “Data Science Machine” participated in three data science competitions. The results – it successfully beat 615 out of 906 human teams.

Data Science Machine’s predictions in two out of the three competitions were 94 percent and 96 percent – as accurate as the winning submissions made by the human teams. In the third competition, the result was a modest 87 percent. However the winning play was due to the fact that the artificial intelligent machine took somewhere between 2 to 12 hours to produce each of the results. And the human counterparts – yes you guessed it right – they labored tirelessly for months to come to the same outcome.

“We view the Data Science Machine as a natural complement to human intelligence,” commented Max Kanter, who is a key figure behind this achievement, because his MIT master’s thesis in computer science is the basis of the Data Science Machine. “There’s so much data out there to be analyzed. And right now it’s just sitting there not doing anything. So maybe we can come up with a solution that will at least get us started on it, at least get us moving.”

Kanter and his thesis advisor, Kalyan Veeramachaneni, a research scientist at MIT’s Computer Science and Artificial Intelligence Laboratory (CSAIL), describe the functional aspects of the Data Science Machine in the paper that will be presented next week at the IEEE International Conference on Data Science and Advanced Analytics.

Veeramachaneni also co-leads the Anyscale Learning for All group at CSAIL, which basically applies machine-learning techniques to attempt to solve realistic problems in big-data analysis. These include predicting which students are at most risk of dropping out of online-courses in the coming semesters or determining the power-generation capacity of wind-farm sites.

“What we observed from our experience solving a number of data science problems for industry is that one of the very critical steps is called feature engineering. The first thing you have to do is identify what variables to extract from the database or compose, and for that, you have to come up with a lot of ideas,” commented Veeramachaneni.

For predicting the rate of dropouts for instance, two of the most crucial indicators were how long before a deadline a student begins working on an assignment and how much time does each student spend on the course website in relation to his/her course mates.

Needless to state that such an artificial intelligent machine capable of analyzing massive volumes of big data in unbelievably short periods of time will have numerous practical applications. Many corporations, institutions, and governments already deal with a large amount of data and sometimes even avoid collecting particular pieces of data which are relatively essential due to network, storage and sensor constraints.

However with such a breakthrough by artificial intelligence in big data analysis will certainly augment the job of a big data scientist by adding another layer of automation. Even though big data scientists would have to code the analysis portions for the engines manually to let the computerized brains loose, it would still be a massive improvement in the existing field of big data analysis.

