

24,049 views | Oct 26, 2019, 12:40am

How Data Science Could Become A Mainstream Career

**Gregory Ferenstein** Contributor[Sharing Economy](#)*I write about technology, politics and personal health.*

The screenshot shows the AutoML Natural Language interface. The top navigation bar includes a hamburger menu, the text 'AutoML Natural Language BETA', the user 'kaggle3', and buttons for 'ADD ITEMS', 'LABEL STATS', 'EXPORT DATA', and a trial ID 'collaboratory-trial-2430'. Below the navigation bar, a dropdown menu shows 'kaggle3_v20191019213230'. The main section is titled 'Test your model on new text' with a subtitle 'Test your model on text items that capture the diversity of your expected inputs. [Learn more](#)'. A text input field contains 'walking among the trees' with a character count '9976 characters remaining' and a 'PREDICT' button. To the right, a 'Predictions' table shows the following results:

Predictions	Score
exercise	0.649
leisure	0.148
nature	0.112
achievement	0.066
enjoy_the_moment	0.024
bonding	0.001
affection	0.000

Below the screenshot, the text 'Use your custom model' is visible, followed by the caption 'AutoML screenshot FERENSTEIN'.

One of the big questions about the future of work is how we will train people to take on some of the fastest growing technical jobs in the economy. After all, [occupations such as Data Science](#) are incredibly complex and can require years of formal statistical and programming training, which is a big change from the kinds of occupations likely to be automated (such as trucking or retail sales).

But I think that technology not only will expand technical jobs but also make them much more accessible for people without training. As an example, being a taxi driver used to be an incredibly knowledge intensive task with the driver test in London [famously](#)

requiring drivers to know extraordinary detail about the city's layout. But, thanks to GF apps, all drivers have to know is how to take directions from a voiceover and follow algorithmically routed directions.

I think data science is going to follow a similar path. For the past year or so, I [decided that I wanted to learn](#) the underlying programming of automation and machine learning from sites such as Udacity and DataCamp. Right now, if you want to be a machine learning engineer, you have to know a decent amount of python or other programming languages and have to know how to construct neural networks manually. Even though there are some programming packages which make it easier to make machine learning models, it's still pretty important to understand a lot of the underline computer science and that can take quite a bit of training, especially for folks who don't have any experience in computational thinking.

Today In: [Innovation](#)



But, now that machine learning and data science is becoming more commonplace, there's starting to be more user-friendly, no code tools that would allow someone to perform sophisticated analysis without knowing programming or math.

One such tool is Google's [AutoML](#). Without any coding or even knowledge of neural networks, a user can easily upload a data set and, with a few clicks, can start to perform pretty impressive prediction algorithms around sentiment analysis or text classification

To give an example, there is an [open data set](#) freely available on Kaggle, the data science learning website, of thousands of people describing something good that happened to them on a particular day. For instance, the description "I went to the gym this morning and did yoga." is categorized under "exercise." The dataset is a light-hearted, and relatively intuitive way to train machine learning models (in this case, it's a model that attempts to predict the different categories to identify how someone would have a good day, such as if they were relaxing or going out into nature).

To build a machine learning model that can identify tens of thousands of categories from scratch is not an easy undertaking for someone who doesn't know how to code. But,

when I use AutoML, it didn't take me more than a few clicks to run the algorithm and accurately predict new data that it had never seen before.

In the picture above, you can see the user interface for Google AutoML where I typed in a brand new phrase that is not in the original data set, "walking among the trees". AutoML accurately predicted that this was most likely "exercise" (as opposed to a category such as "achievement" or "bonding").

I don't yet think tools such as AutoML are sufficient to be a good data scientist. But, I do think my experience with the software is a good illustration of how, in the near future, even something as complex as machine learning could be quite accessible to the general population without years of training. and, that makes me optimistic.



Gregory Ferenstein

Gregory Ferenstein is the editor of the Ferenstein Wire, a syndicated publication on tech, health, and politics. In his spare time, he teaches Mathematics and enjoys C... **Read More**