MIT | EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

# Big Data and Social Analytics certificate course

**MODULE 5 UNIT 1**
**Video 4 Transcript**

# MIT BDA Module 5 Unit 1 Video 4 Transcript

**Speaker key**

YM: Yves-Alexandre de Montjoye

HY: Hapyak

YM: So, in the previous video we learned how to extract behavioral indicators using Bandicoot and then used these indicators to predict individual characteristics such as age and gender. In this video we will now see how to use Bandicoot's behavioral indicators for personalized marketing.

00:00:29

Doing truly data-driven marketing is hard. In fact, a recent IBM study showed that 80% of marketing decisions are based on someone's gut feeling. What I will show you here is the result of an experiment we ran at large scale in partnership with a mobile phone operator. In this experiment we showed how you can use data to improve click-through rates of a marketing campaign.

This took place in an Asian country where the internet penetration rate is low. The goal of the campaign was to improve the penetration rate by sending mobile text messages to customers, offering them a half price internet data plan for a month. Traditionally, the way this works is the marketing team organizes a meeting. They look at the product they would want to promote, in this case a data plan, and then pick the group of customer they think will be interested in it.

In this case they picked people who were sending and receiving at least four texts a month and whose phone could access the Internet. They also picked customer who were spending already more on their phone than other users. Historically, this approach of handpicking features used to work. However, when we have billions of mobile phone records and hundreds of features, having a human decide which one to pick and how to weight them, just doesn't make sense anymore.

HY: When working with billions of data records, why is it not feasible to rely on a human to select which features are most important and weight them accordingly?

Thank you for your reflection, please continue watching the video to hear about a better solution.

00:01:49

YM: So what we proposed instead is to learn from the data. There's already a small number of subscriber in that region who have a data plan so why don't we let an algorithm figure out what distinguishes them from the rest of the population? In the previous video we learned about machine learning algorithm and how that works. In this video we will use another algorithm, bagging classification trees, and train it to recognize people in the country who already have a data plan.

Once this is done we can use our trained algorithm to find in the customer base, people who don't have a data plan but who behave very much like people who have one. So, as often, the best way to compare the two approaches is to try them out. We just ran a large-scale experiment. We had a database of subscribers. The marketer teams handpicked in the database 50,000 people who they think are likely to be interested in a data plan. Then out of the subscribers they did not pick we selected the one to who our algorithm gave the highest score. We picked 200,000 of them.

We now have two groups. We sent texts to all of them and measured the response rate for each group; the fraction of people who subscribed to our offer for a half price internet data plan. Out of the people the marketers picked 0.5% signed up for our data plan. Out of the people picked by our algorithm 6.5% of them signed up. That's a 13 times increase in sign-up rates.

00:03:20

We then looked at retention rate; the fraction of people who renewed their plan the second month at full price. While only 37% of the people in the marketer's group who had signed up for the plan renewed them the second month, a striking 98% of the people in our group did.

The best approach for doing personalized marketing with big data is really to tell an algorithm what to look for but then to let it figure out what the right combination of variable is. We found this to outperform the traditional approach.

HY: Once you have told your algorithm what to look for, the best approach is to explicitly define the right combination of variables to analyze.

True

Incorrect. The best approach is to tell your algorithm what to look for and then leave it to figure out the right combination of variables.

**False**

Correct, well done. The best approach is to tell your algorithm what to look for and then leave it to figure out the right combination of variables.

YM: So to gain some insights into what our algorithm does we looked at the top ranked features, the features that are the most useful to predict who uses a data plan in the country. Here we can see that the marketer's intuition wasn't wrong. The third and the second most useful features are related to how a user texts and how much he spends; his average monthly spending on texts and the average number of texts he sends every month.

It's not about the intuition. It's really about finding the right variable, weights, and thresholds and this is what our algorithm does. The most useful feature is different though and might be something that the marketing team had not thought about. In economic term there might be a network effect going on with data plan. What network effect means is basically that the value of a product to you is a function of the number of people who are also using the same product.

We did find that the most useful feature for our algorithm was actually the amount your close contacts were spending on data. And again it doesn't mean that the marketer might not have thought about it, but really what it means is that algorithm are much better than humans at picking the right combination of variables.

00:05:12

So, to summarize, in this video we looked at how we can combine Bandicoot indicators with machine learning algorithm to do personalized marketing. What we showed is that by our combining the strength of the human with the strengths of the machine, we can outperform the traditional marketing approach in a large-scale experiment.

**Tel:** +1 224 249 3522 | **Email:** info@getsmarter.ac | **Website:** getsmarter.mit.edu

**Learning Outcomes:**

# 1. Introduction