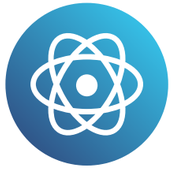
Supervised machine learning

DATA SCIENCE FOR BUSINESS

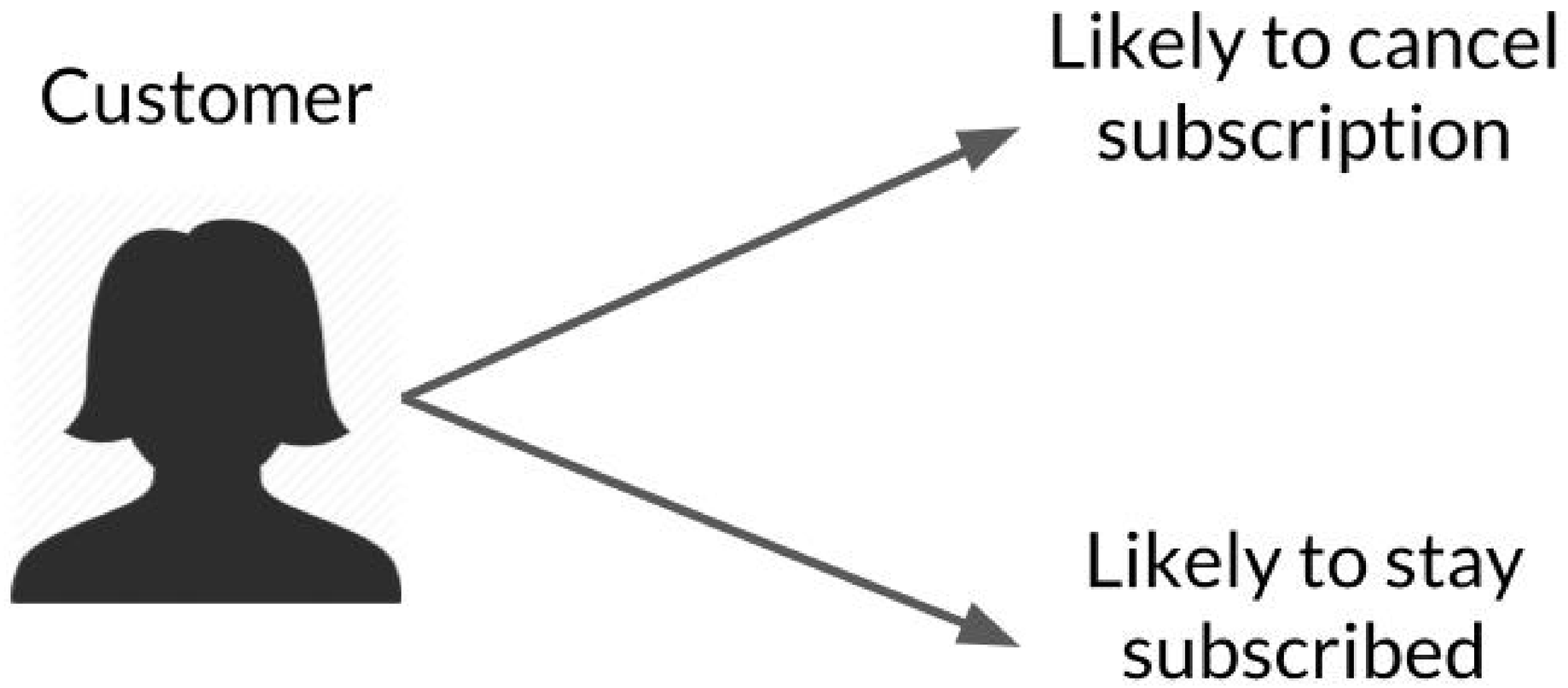


Ramnath Vaidyanathan
VP of Product Research, DataCamp

What is supervised machine learning?

- **Machine learning**: Predictions from data
- ***Supervised machine* learning**: Predictions from data with *labels* and *features*
  - Recommendation systems
  - Email subject optimization
  - Churn prediction

Case study: churn prediction



Case study: churn prediction

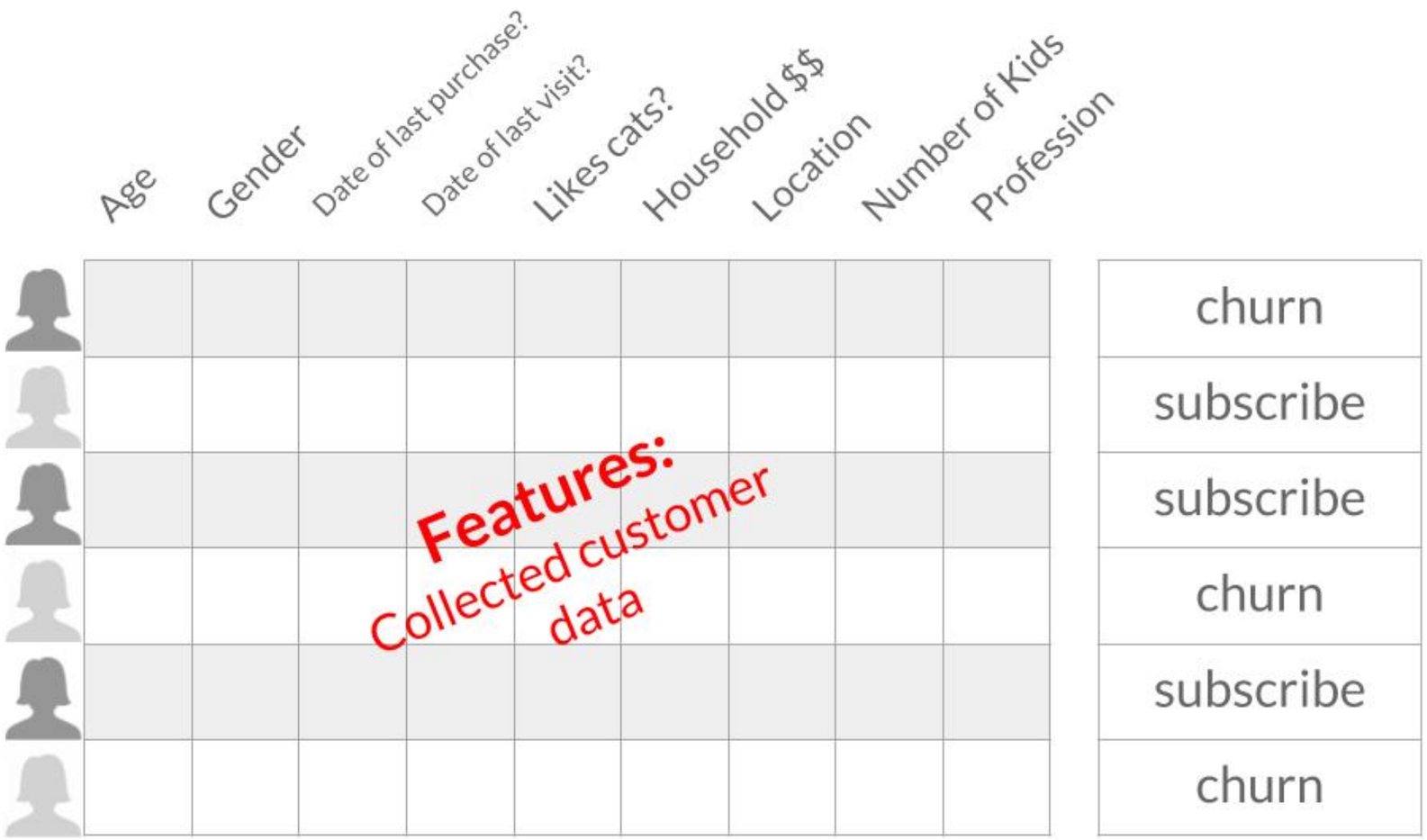
Training Data:
Customers

Case study: churn prediction

Labels
Customer outcomes

churn
subscribe
subscribe
churn
subscribe
churn

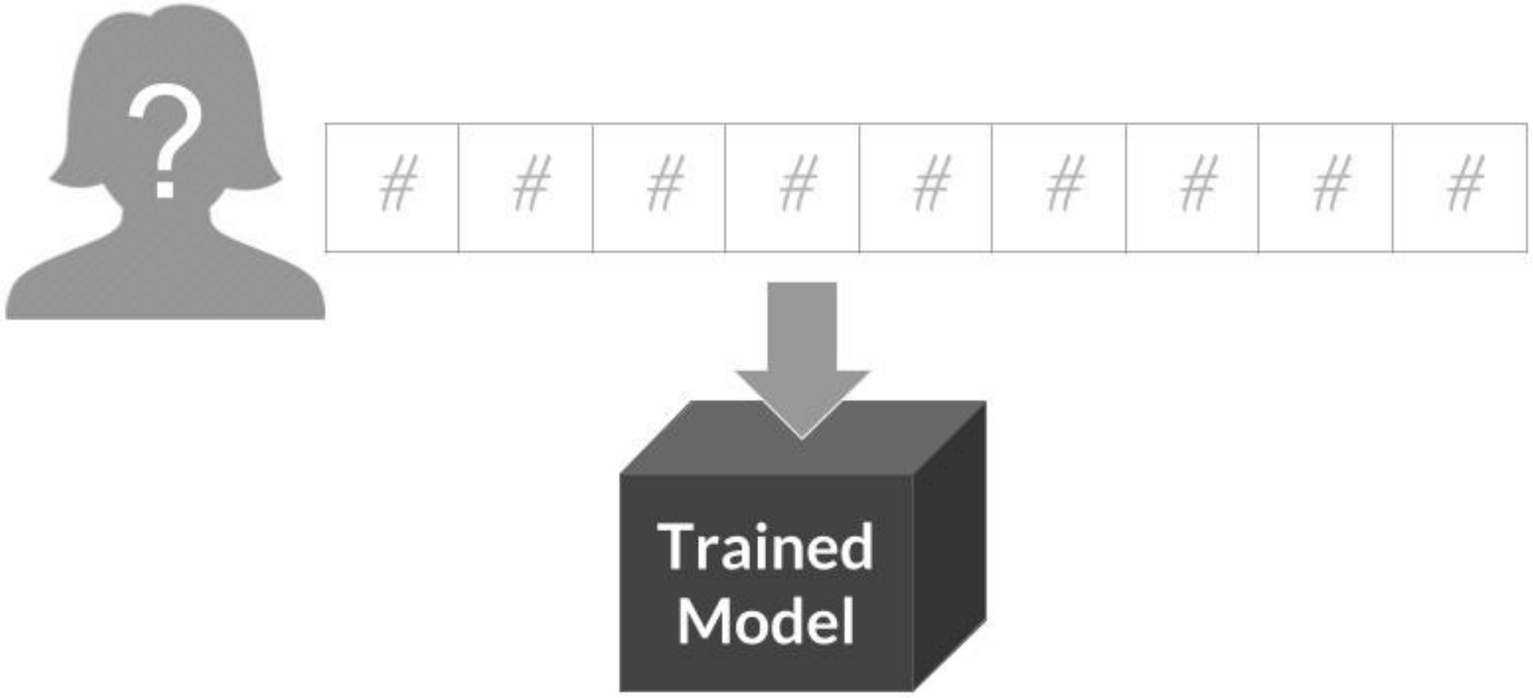
Case study: churn prediction



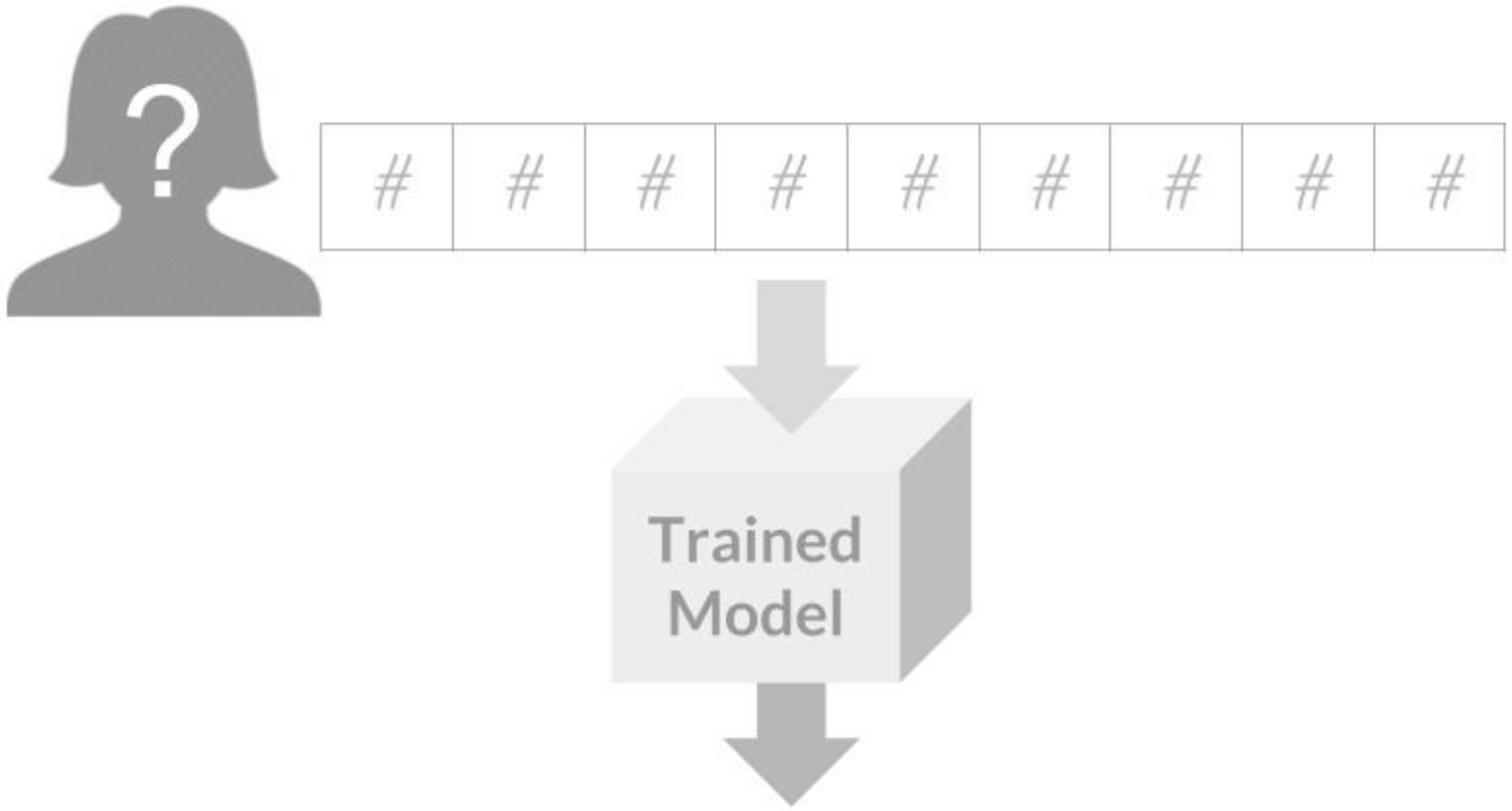
Case study: churn prediction

#	#	#	#	#	#	#	#	#

Case study: churn prediction



Case study: churn prediction



Recap

- Make a prediction based on data
- Data has *features* and *labels*
  - Label: what we want to predict
  - Features: data that might predict the label
- Trained model can make predictions

Model evaluation

Split historical data into training and testing sets

Model evaluation

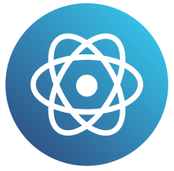
	Prediction	Reality
Churn	0%	3%
Remain	100%	97%

Let's practice!

DATA SCIENCE FOR BUSINESS

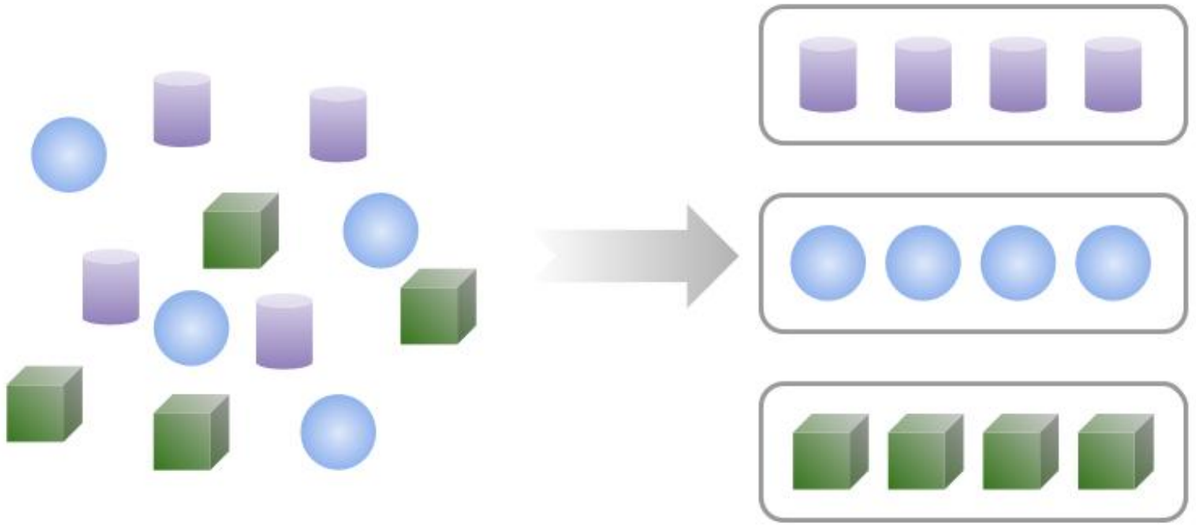
Clustering

DATA SCIENCE FOR BUSINESS



Ramnath Vaidyanathan
VP of Product Research, DataCamp

What is clustering?

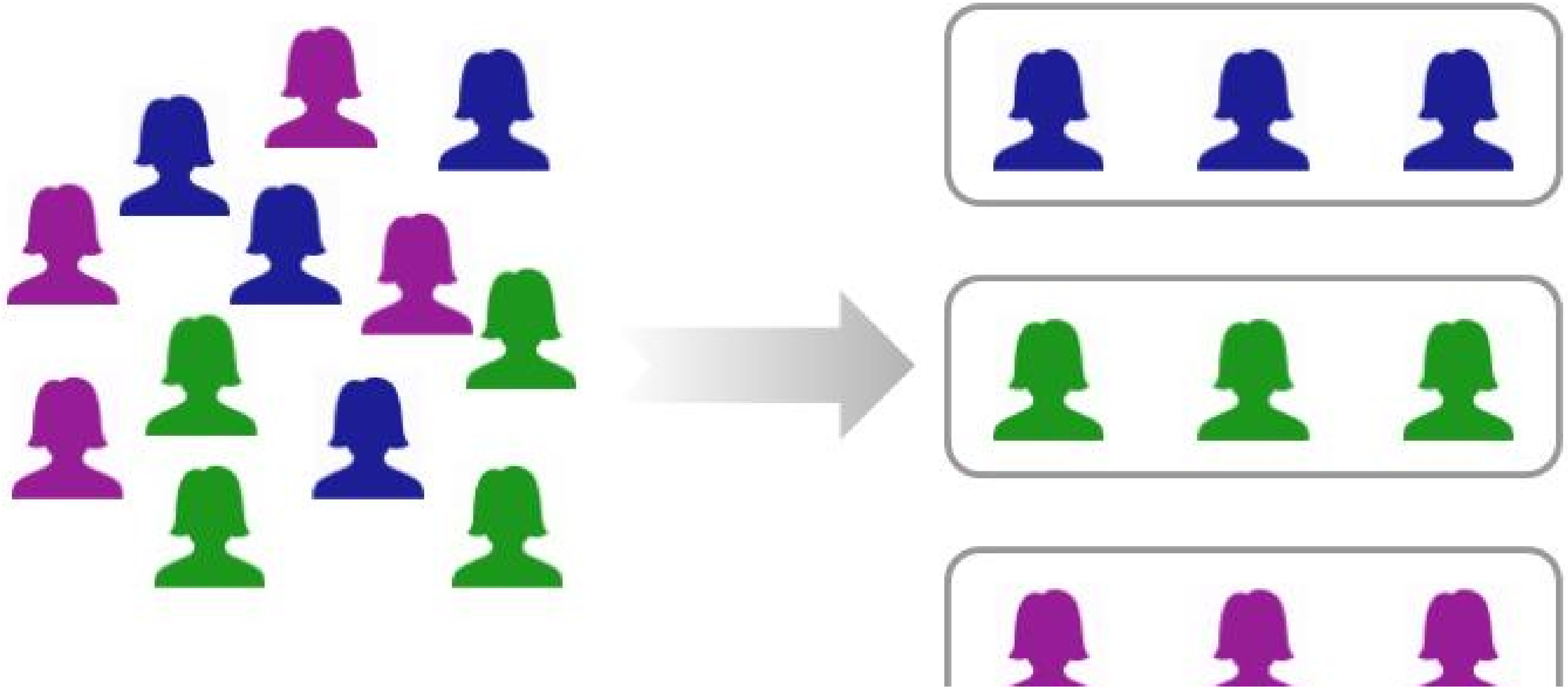


- Divide data into categories
- Use cases
  - Customer segmentation
  - Image segmentation
  - Anomaly detection

Supervised Machine Learning

Unsupervised Machine Learning

Case study: customer segmentation



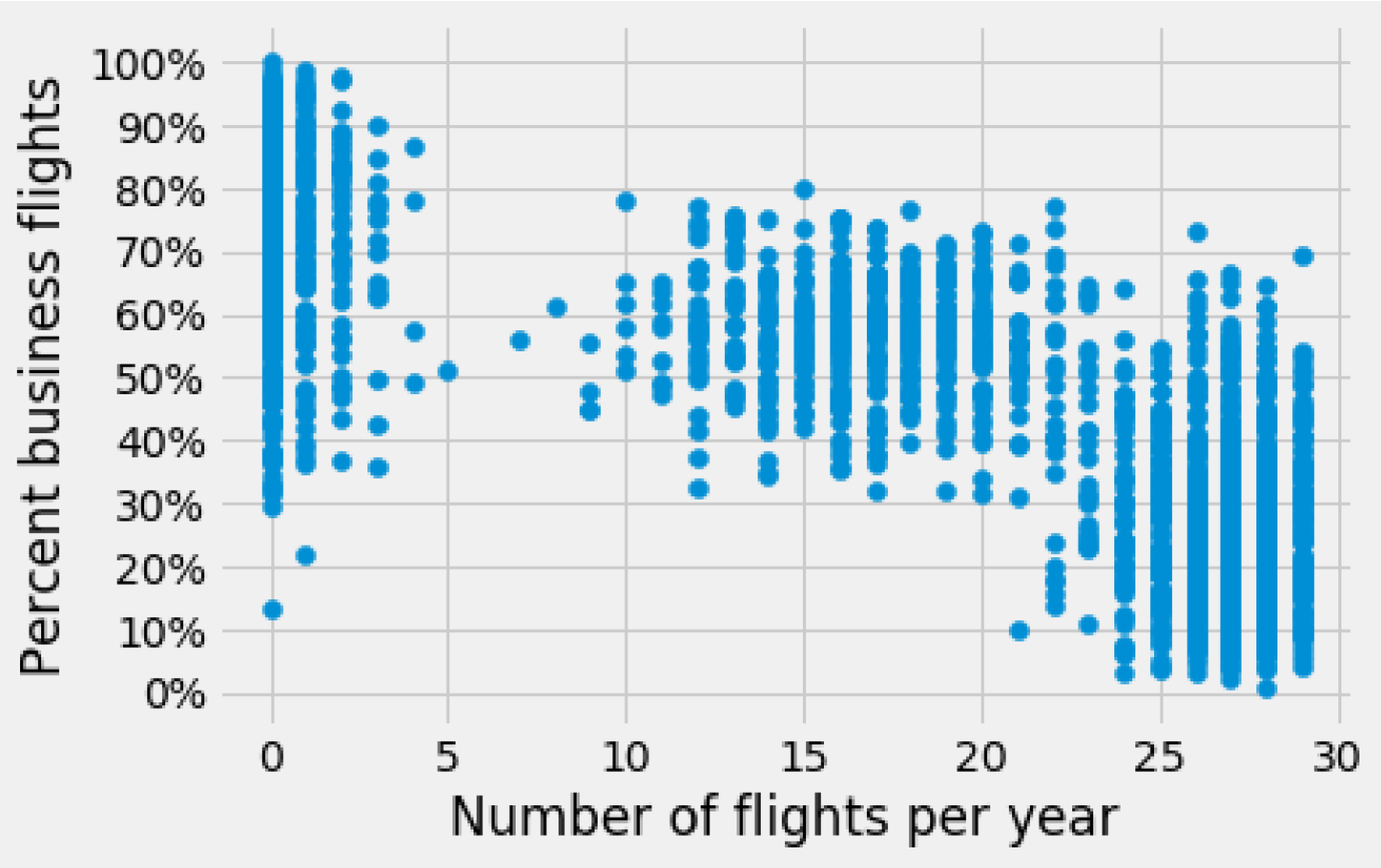
Case study: customer segmentation

Define features

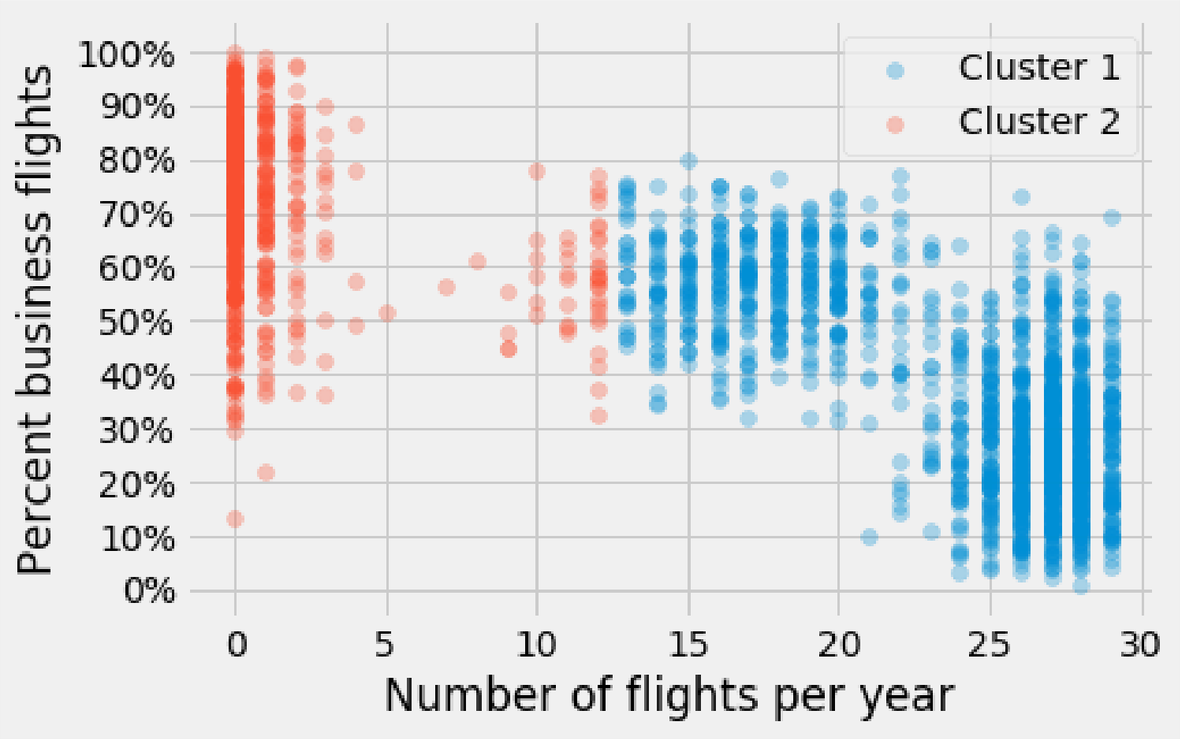
- Number of flights in the past year
- Percent international
- Advanced planning
- Percent business class

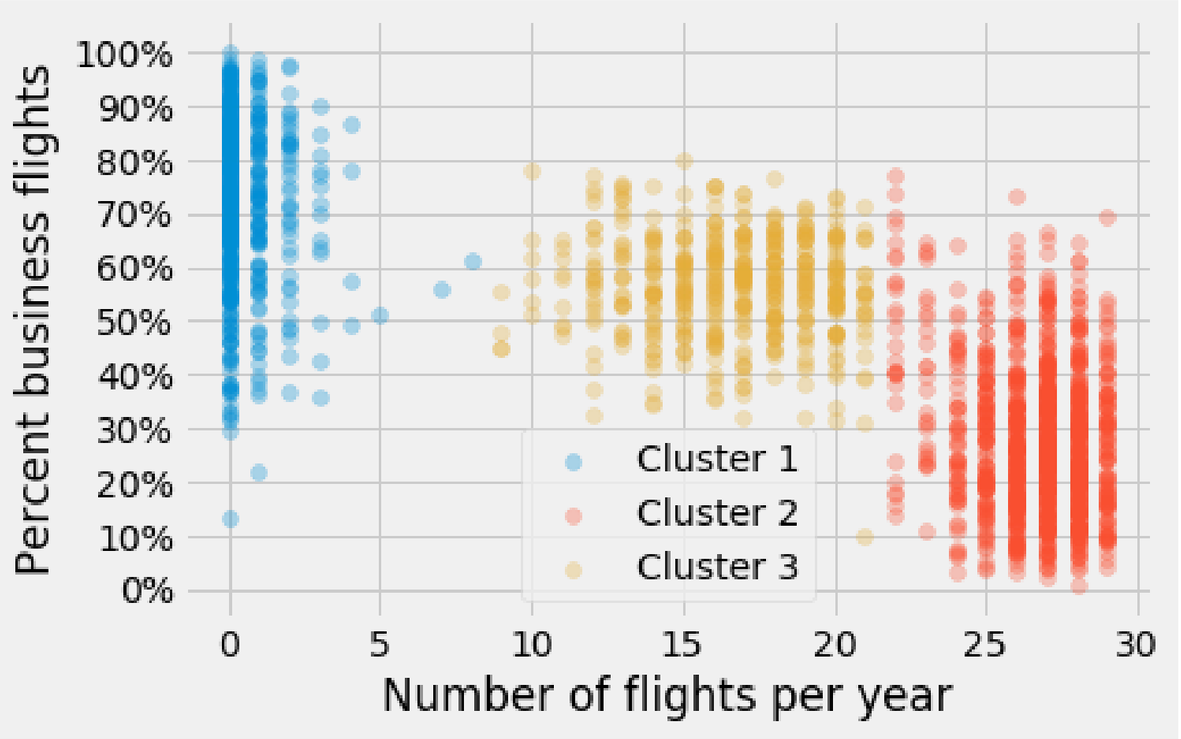
Case study: customer segmentation

- Define number of clusters



Case study: customer segmentation





Clustering review

Definition

- Divide unlabeled dataset into different categories

Steps

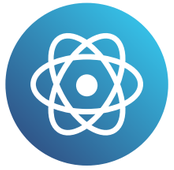
- Select features
- Select number of clusters
- Use clusters to solve business problems

Let's practice!

DATA SCIENCE FOR BUSINESS

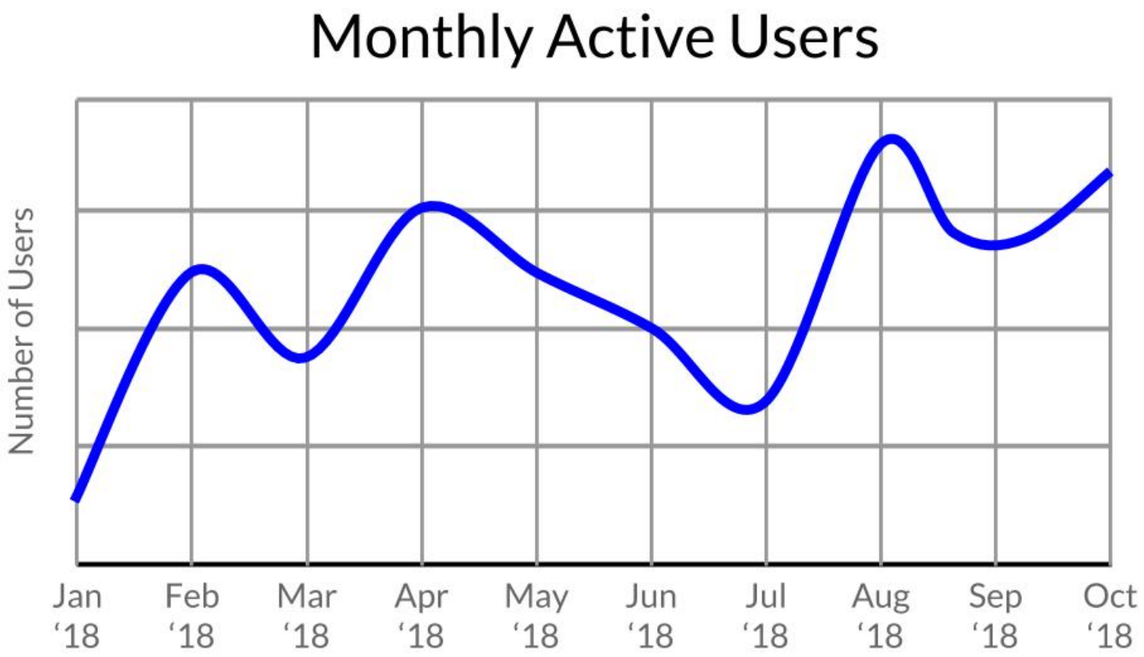
Special topics in Machine Learning

DATA SCIENCE FOR BUSINESS



Ramnath Vaidyanathan
VP of Product Research, DataCamp

Time series forecasting



- Time is a feature
- Accounts for weekly, monthly, or yearly trends

Seasonality

- **Weekly**: Lower television viewership on Fridays
- **Monthly**: Higher spending at end of pay periods
- **Yearly**: Less ice cream in the winter

Natural Language Processing

- Dataset is text
  - Customer reviews
  - Tweets
  - Medical records
  - Email subjects
- Possible uses
  - Classifying sentiment
  - Clustering medical records

Sentiment Analysis is an NLP methodology for quantifying how positive or negative the emotion expressed by a segment of text is. It is often used for automatically categorizing customer feedback messages or product reviews.

Word counts

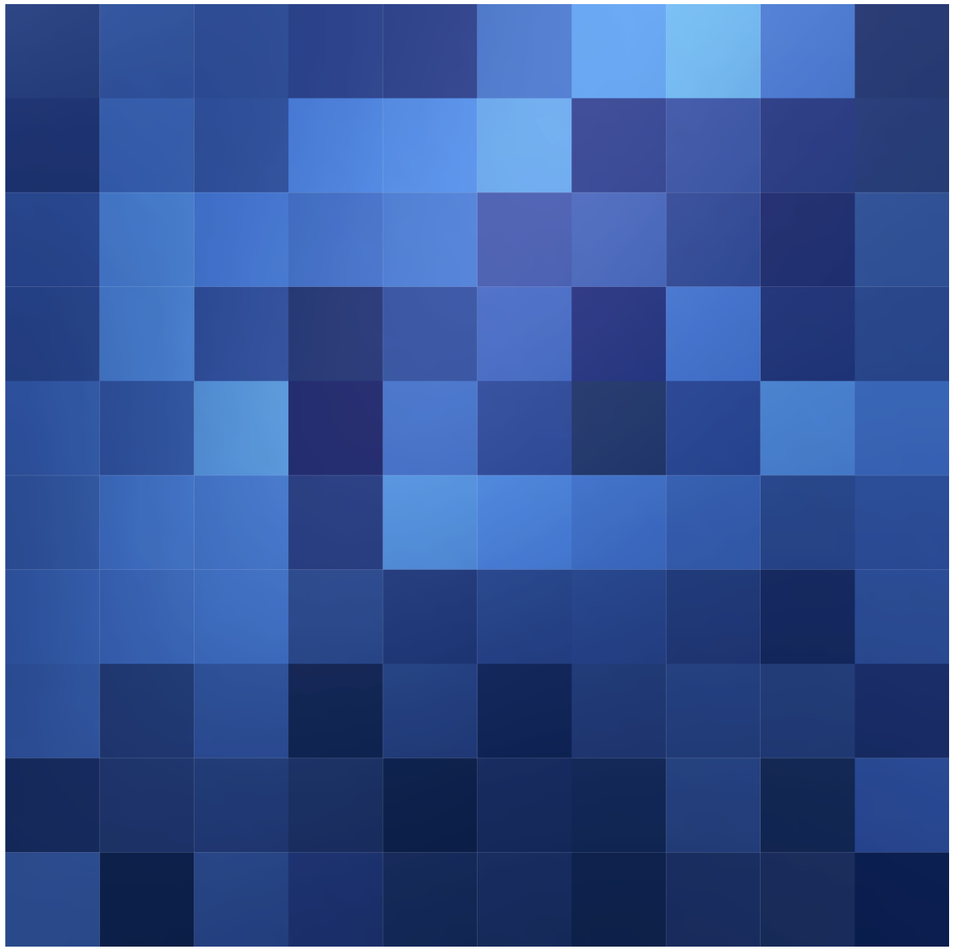
Sentence	Texans	Giants	football	great
The Texans are a great football team.	1	0	1	1
The Giants are a great football team.	0	1	1	1

Problems with word counts: negation

Sentence	Texans	Giants	football	great	not
The Giants are a great football team.	0	1	1	1	0
The Giants are not a great football team.	0	1	1	1	1

Word counts and synonyms

- Word counts don't help us consider synonyms
- Example: "blue"
  - "sky-blue"
  - "aqua"
  - "cerulean"
- Want to group as a single feature



Word embeddings

- Create features that group similar words
- Features have a mathematical meaning:

```
king - man + woman = queen
```

Review

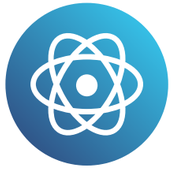
- **Time series forecasting**
  - Time is a feature
  - Seasonality
- **Natural Language Processing (NLP)**
  - Text as input data
  - Word counts
  - Word embeddings

Let's practice!

DATA SCIENCE FOR BUSINESS

Deep Learning and Explainable AI

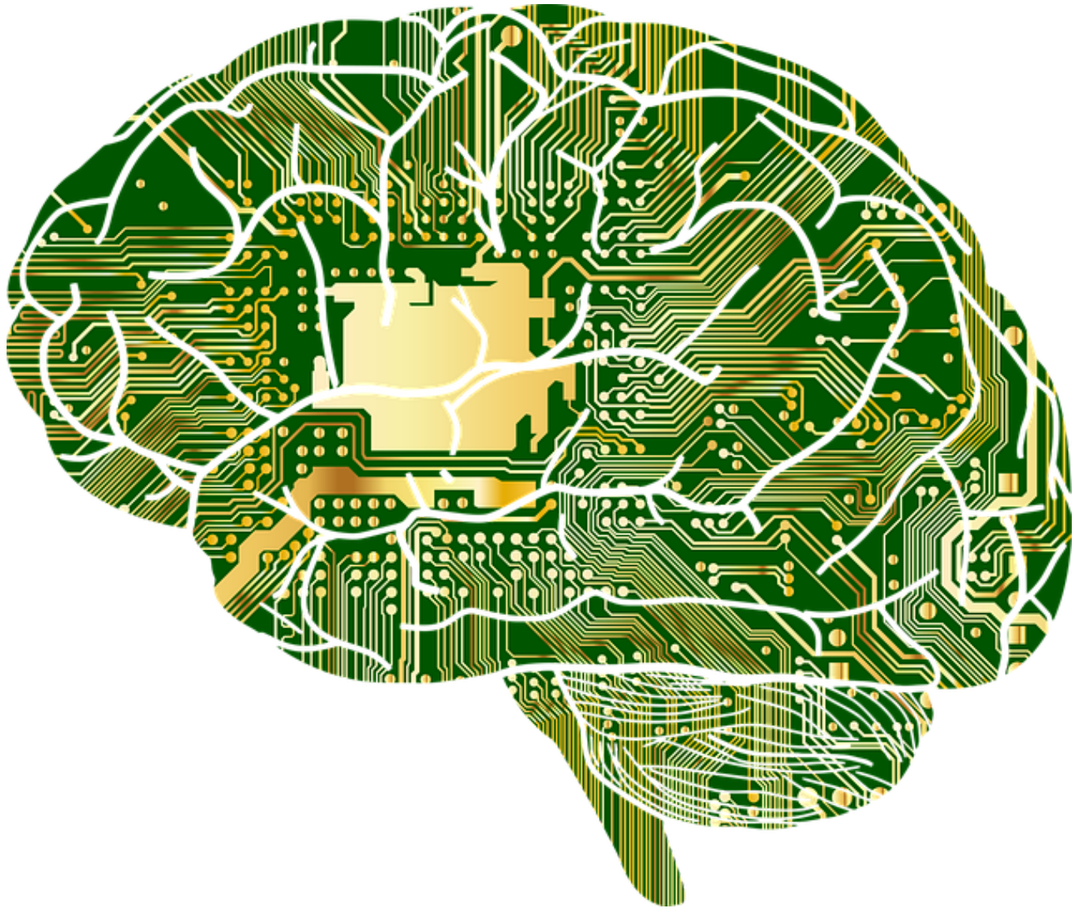
DATA SCIENCE FOR BUSINESS



Ramnath Vaidyanathan
VP of Product Research, DataCamp

What is Deep Learning?

- AKA: Neural Networks or Neural Nets
- Special area of Machine Learning
- Requires more data
- Best when inputs that are images or text



Explainable AI

Deep Learning	Explainable AI
Highly accurate predictions	Understandable by humans
Better for "What?"	Better for "Why?"

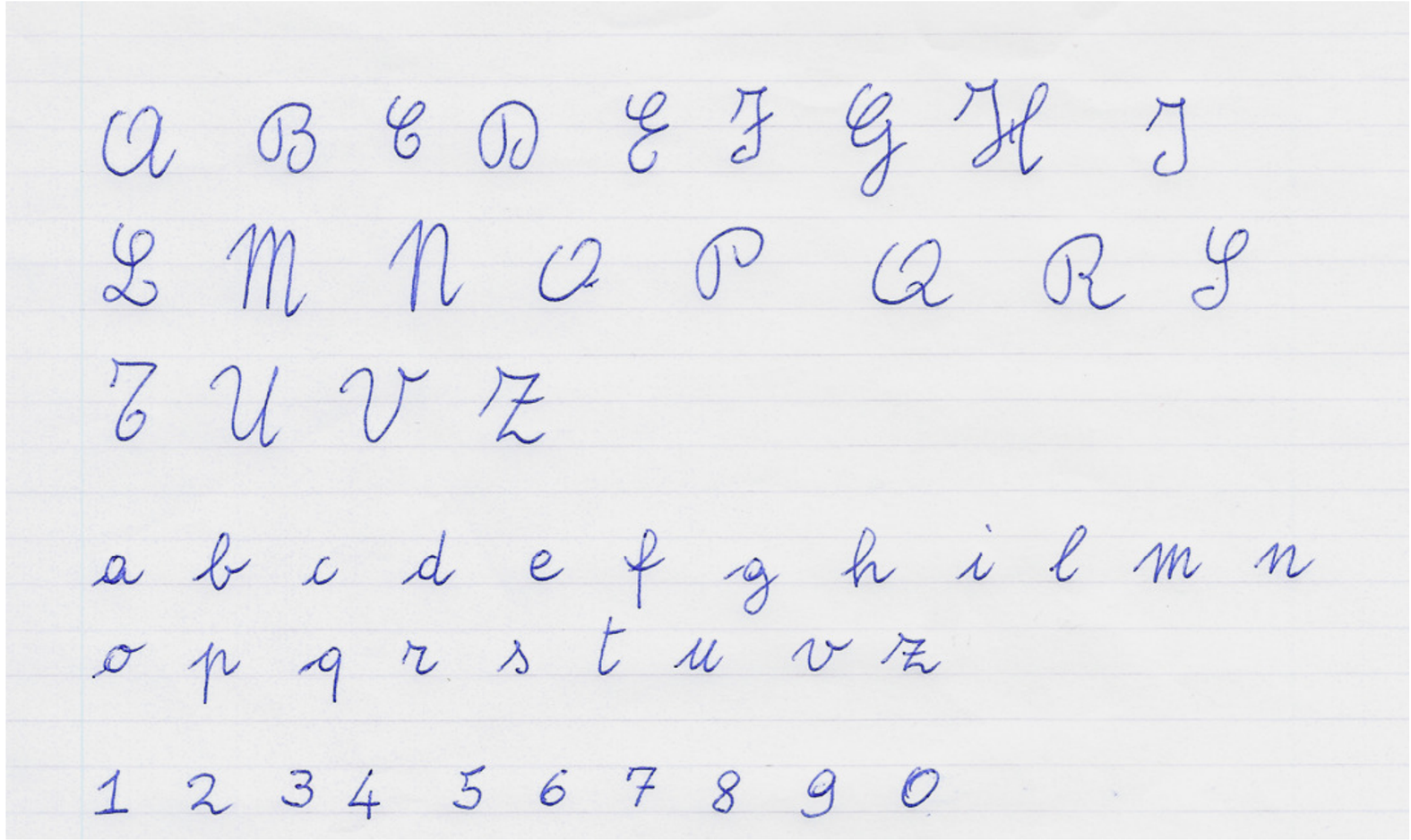
Problems with complex inputs like large quantities of text or images require a Deep Learning approach. Problems where it is helpful to know why the alogrithm chose a particular classification should be tackled with Explainable AI.

Case Study: Explainable AI

1. **Prediction**: What a customer is likely to do
2. **Explanation**: Why a customer is likely to do it

Case Study: Inexplicable AI

Prediction only: Which letter is this likely to be?



When to use Deep Learning

1. Is the training data complex?
2. Do we have a very large amount of training data?
3. Does the model need to be predictive or explanatory?

Deep Learning is great for predictive modeling, but can leave us perplexed if we care about why each prediction was made. Simpler models might have less predictive power but can be better when clarity is essential.

Let's practice!

DATA SCIENCE FOR BUSINESS