

# What is Machine Learning?

---



**David Chappell**

PRINCIPAL, CHAPPELL & ASSOCIATES

@DChappellAssoc [www.davidchappell.com](http://www.davidchappell.com)



# What machine learning does



**Finds patterns in data**

**Uses those patterns to predict the future**

**Examples:**

- Detecting credit card fraud
- Determining whether a customer is likely to switch to a competitor
- Deciding when to do preventive maintenance on a factory robot
- Lots more





What  
Does It Mean  
to Learn?

**How did you learn to read?**

**Learning requires:**

- Identifying patterns
- Recognizing those patterns when you see them again

**This is what machine learning does**

# Finding Patterns: A Simple Example

<i><b>Name</b></i>	<i><b>Amount</b></i>	<i><b>Fraudulent</b></i>
Smith	\$2,600.45	No
Potter	\$2,294.58	Yes
Peters	\$1,003.30	Yes
Adams	\$8,488.32	No

*What's the  
pattern for  
fraudulent  
transactions?*



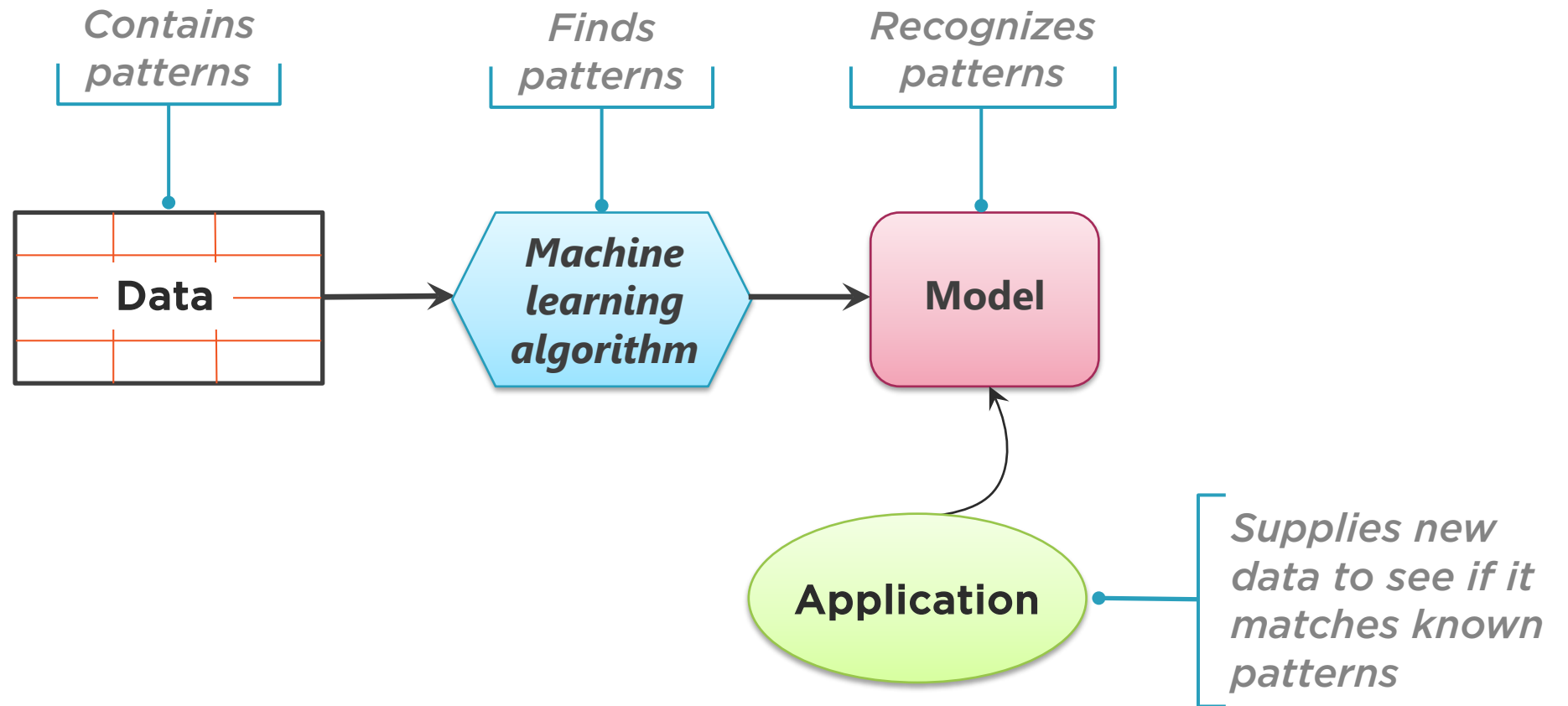
# Finding Patterns: Another Example

<i><b>Name</b></i>	<i><b>Amount</b></i>	<i><b>Where Issued</b></i>	<i><b>Where Used</b></i>	<i><b>Age</b></i>	<i><b>Fraudulent</b></i>
Smith	\$2,600.45	USA	USA	22	No
Potter	\$2,294.58	USA	RUS	29	Yes
Peters	\$1,003.30	USA	RUS	25	Yes
Adams	\$8,488.32	FRA	USA	64	No
Pali	\$200.12	AUS	JAP	58	No
Jones	\$3,250.11	USA	RUS	43	No
Hanford	\$8,156.20	USA	RUS	27	Yes
Marx	\$7,475.11	UK	GER	32	No
Norse	\$540.00	USA	RUS	27	No
Edson	\$7,475.11	USA	RUS	20	Yes

*What's the  
pattern for  
fraudulent  
transactions?*



# Machine Learning in a Nutshell





Why is machine learning so hot right now?

**Doing machine learning well requires:**

- Lots of data
- Lots of compute power
- Effective machine learning algorithms

**All of those things are now more available than ever**

# Who's Interested in Machine Learning?



**Business Leaders**  
Want solutions to  
business problems



**Software Developers**  
Want to create better  
applications



**Data Scientists**  
Want powerful, easy-  
to-use tools







What is a data scientist?

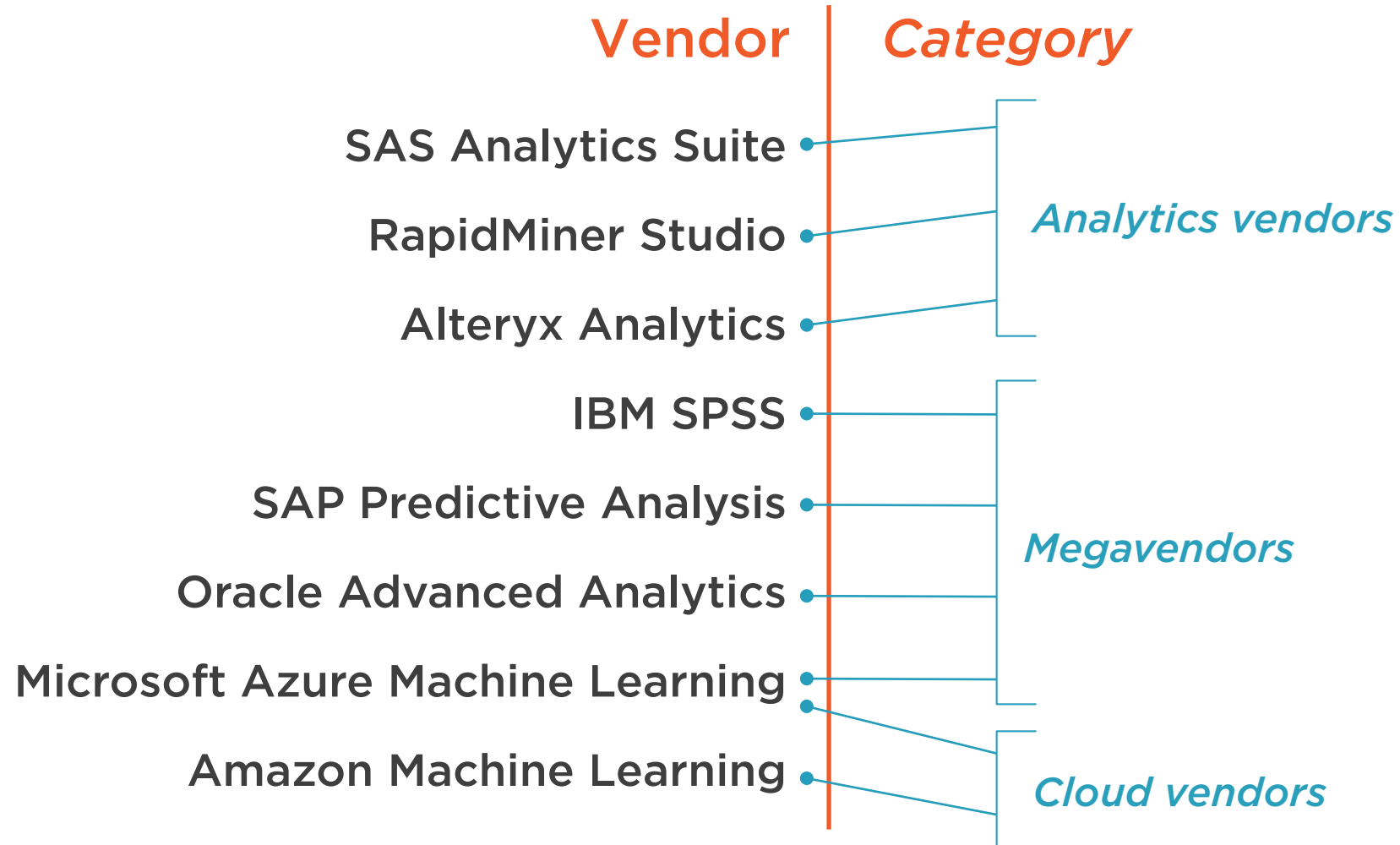
### Someone who knows about:

- Statistics
- Machine learning software
- Some problem domain (ideally)

### Key facts about data scientists:

- Good ones are scarce
- Good ones are expensive

# Example Machine Learning Offerings





## The Role of R

### **R is an open source programming language and environment**

- Supports machine learning, statistical computing, and more
- Has many available packages

### **R is very popular**

- Many commercial machine learning offerings support R

### **But it's not the only choice:**

- Python is also increasingly popular





## The Main Points

**Machine learning lets us find patterns in existing data, then create and use a model that recognizes those patterns in new data**

**Machine learning has gone mainstream**

- Big vendors think there's big money in this market

**Machine learning can probably help your organization**

