An introduction to Machine Learning

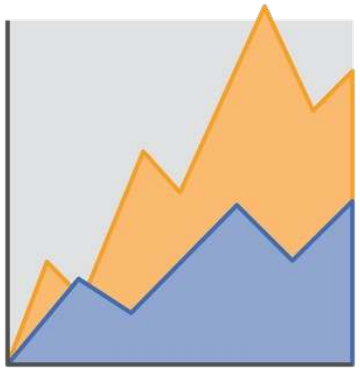
Julien Simon, Al Evangelist, EMEA
@julsimon

What to expect

- Why Machine Learning?
- Use cases
- The basics of Machine Learning
- Demos
- Q&A

Why Machine Learning?

Three types of data-driven development

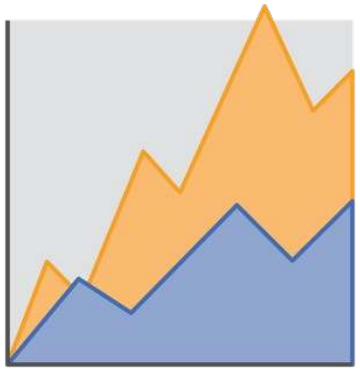


Retrospective

analysis and reporting

Amazon Redshift
Amazon RDS
Amazon S3
Amazon EMR

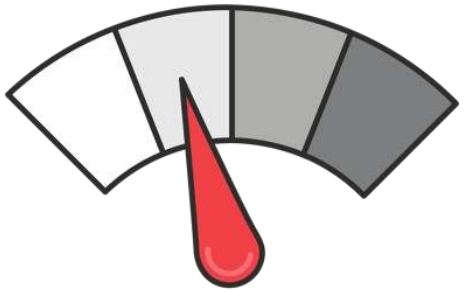
Three types of data-driven development



Retrospective

analysis and reporting

Amazon Redshift
Amazon RDS
Amazon S3
Amazon EMR

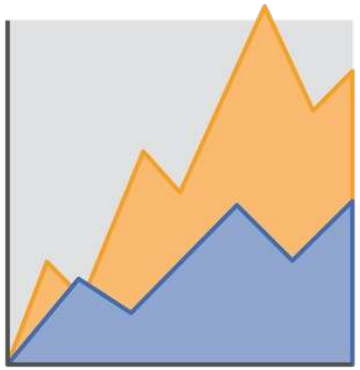


Here and now

real-time processing and dashboards

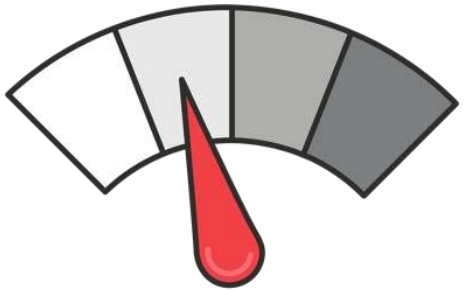
Amazon Kinesis
Amazon EC2
AWS Lambda

Three types of data-driven development



Retrospective
analysis and reporting

Amazon Redshift
Amazon RDS
Amazon S3
Amazon EMR



Here and now
real-time processing and dashboards

Amazon Kinesis
Amazon EC2
AWS Lambda

Predictions
to enable smart applications

Machine learning and smart applications

Machine learning is the technology that automatically finds patterns in your data and uses them to make predictions for new data points as they become available

Machine learning and smart applications

Machine learning is the technology that automatically finds patterns in your data and uses them to make predictions for new data points as they become available

Your data + machine learning = smart applications

Building smart applications – a *counter pattern*

```
Dear customer,

This awesome quadcopter is on sale
for just $49.99!
```

Smart applications by *counter example*

```
SELECT  c.ID
FROM    customers c
        LEFT JOIN orders o
            ON c.ID = o.customer
GROUP   BY c.ID
HAVING  o.date > GETDATE() – 30
```

We can start by sending the offer to all customers who placed an order in the last 30 days

Smart applications by *counter example*

```
SELECT  c.ID
FROM    customers c
        LEFT JOIN orders o
            ON c.ID = o.customer
GROUP   BY c.ID
HAVING  o.category = ‘toys’
   AND  o.date > GETDATE() – 30
```

… let’s narrow it down to just customers who bought toys

Smart applications by *counter example*

```
SELECT  c.ID
FROM    customers c
        LEFT JOIN orders o
            ON c.ID = o.customer
        LEFT JOIN products p
            ON p.ID = o.product
GROUP   BY c.ID
HAVING  o.category = 'toys'
   AND ((p.description LIKE '%helicopter%'
         AND o.date > GETDATE() - 60)
     OR (COUNT(*) > 2
         AND  SUM(o.price) > 200
         AND  o.date > GETDATE() – 30)
       )
```

… and expand the query to customers who purchased other toy helicopters recently

Smart applications by *counter example*

```
SELECT  c.ID
FROM    customers c
        LEFT JOIN orders o
            ON c.ID = o.customer
        LEFT JOIN products p
            ON p.ID = o.product
GROUP   BY c.ID
HAVING  o.category = 'toys'
   AND ((p.description LIKE '%copter%'
         AND o.date > GETDATE() - 60)
     OR (COUNT(*) > 2
         AND  SUM(o.price) > 200
         AND  o.date > GETDATE() – 30)
       )
```

… but what about quadcopters?

Smart applications by *counter example*

```
SELECT  c.ID
FROM    customers c
        LEFT JOIN orders o
            ON c.ID = o.customer
        LEFT JOIN products p
            ON p.ID = o.product
GROUP   BY c.ID
HAVING  o.category = 'toys'
   AND ((p.description LIKE '%copter%'
         AND o.date > GETDATE() - 120)
     OR (COUNT(*) > 2
         AND  SUM(o.price) > 200
         AND  o.date > GETDATE() – 30)
       )
```

… maybe we should
go back further in time

Smart applications by *counter example*

```
SELECT  c.ID
FROM    customers c
        LEFT JOIN orders o
            ON c.ID = o.customer
        LEFT JOIN products p
            ON p.ID = o.product
GROUP   BY c.ID
HAVING  o.category = 'toys'
   AND ((p.description LIKE '%copter%'
         AND o.date > GETDATE() - 120)
     OR (COUNT(*) > 2
         AND  SUM(o.price) > 200
         AND  o.date > GETDATE() – 40)
       )
```

… tweak the query more

Smart applications by *counter example*

```
SELECT  c.ID
FROM    customers c
        LEFT JOIN orders o
            ON c.ID = o.customer
        LEFT JOIN products p
            ON p.ID = o.product
GROUP   BY c.ID
HAVING  o.category = ‘toys’
   AND ((p.description LIKE ‘%copter%’
         AND o.date > GETDATE() - 120)
     OR (COUNT(*) > 2
         AND  SUM(o.price) > 150
         AND  o.date > GETDATE() – 40)
       )
```

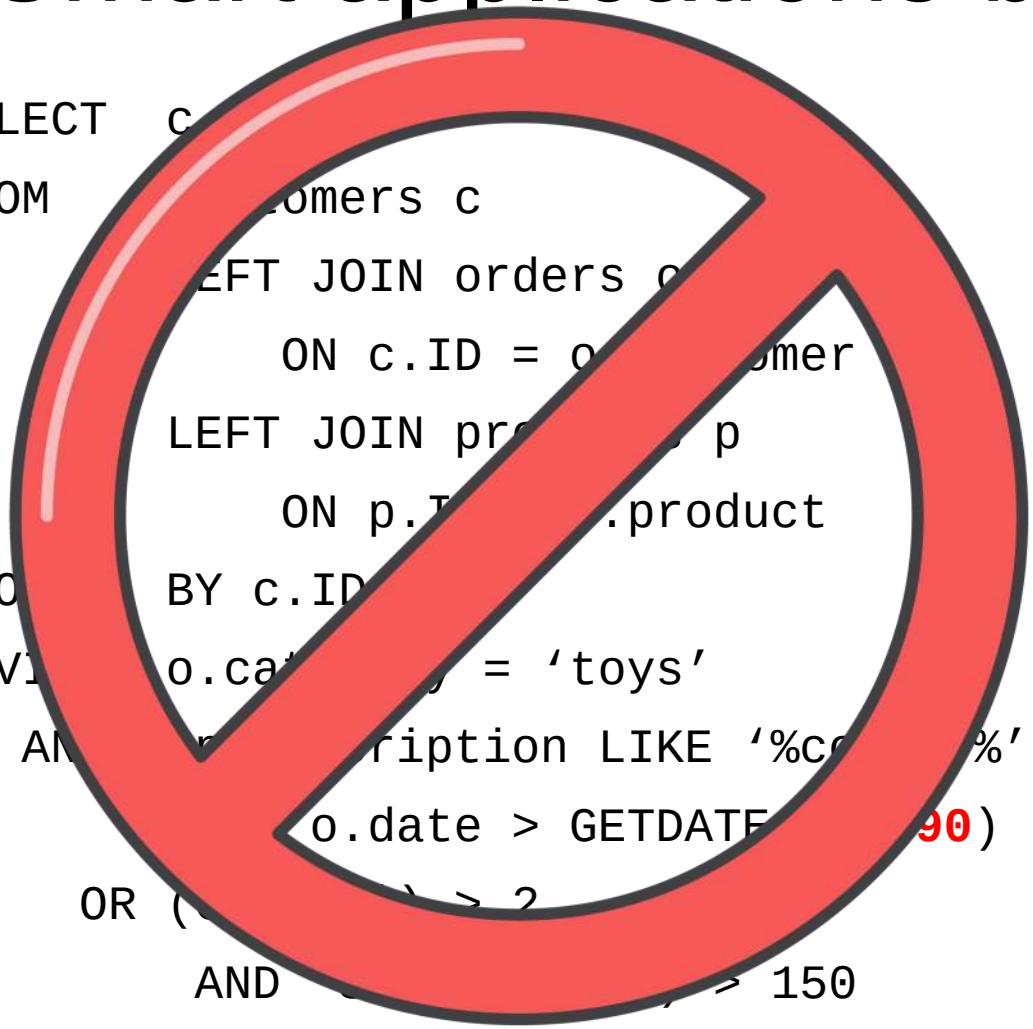
… again

Smart applications by *counter example*

```
SELECT  c.ID
FROM    customers c
        LEFT JOIN orders o
            ON c.ID = o.customer
        LEFT JOIN products p
            ON p.ID = o.product
GROUP   BY c.ID
HAVING  o.category = ‘toys’
   AND ((p.description LIKE ‘%copter%’
         AND o.date > GETDATE() - 90)
     OR (COUNT(*) > 2
         AND  SUM(o.price) > 150
         AND  o.date > GETDATE() – 40)
       )
```

… and again

Smart applications by *counter example*

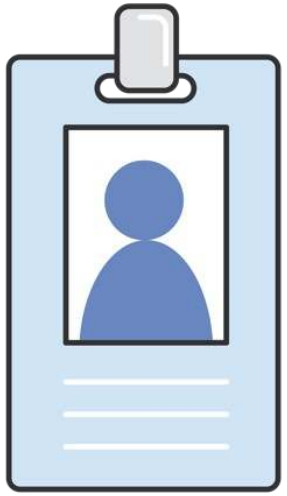


```
SELECT  c
FROM        omers c
        LEFT JOIN orders o
            ON c.ID = o      omer
        LEFT JOIN pr       p
            ON p.I       .product
GRO     BY c.ID
HAVI    o.ca       = 'toys'
    AN      ription LIKE '%co      %'
            o.date > GETDATE      90)
    OR (        ) > 2
        AND                  > 150
        AND  o.date > GETDATE() - 40)
    )
```

Use machine learning to learn your business rules from data!

Use cases

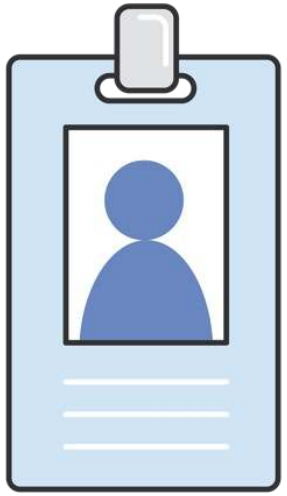
Smart applications by example

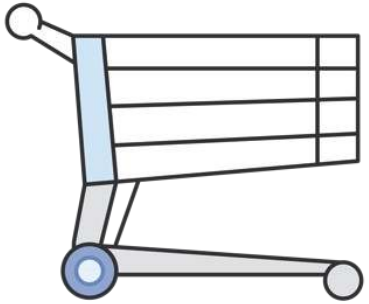


Based on what you know about the **user**:

Will they use your product?

Smart applications by example





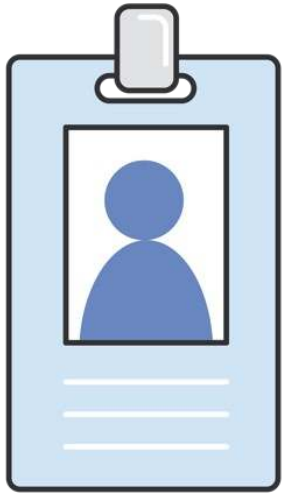
Based on what you know about the **user**:

Will they use your product?

Based on what you know about an **order**:

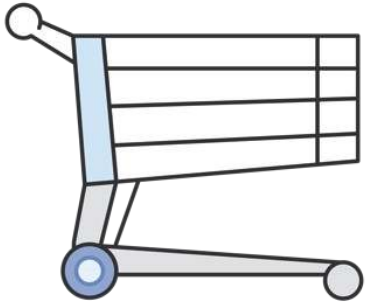
Is this order fraudulent?

Smart applications by example



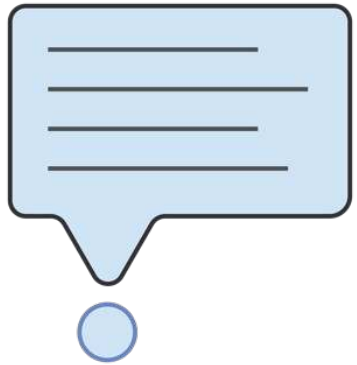
Based on what you know about the **user**:

Will they use your product?



Based on what you know about an **order**:

Is this order fraudulent?



Based on what you know about a **news article**:

What other articles are interesting?

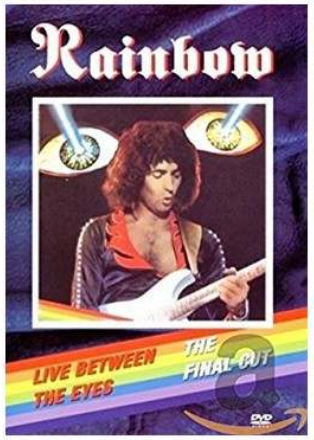
And a few more examples…

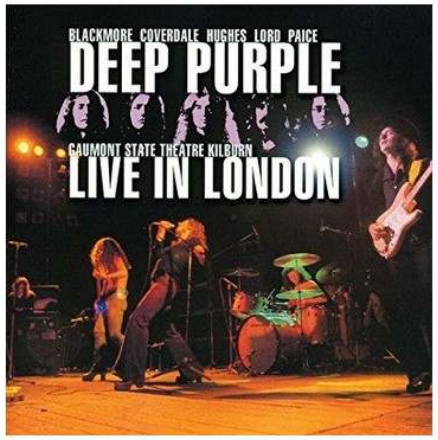
Fraud detection	Detecting fraudulent transactions, filtering spam emails, flagging suspicious reviews, …
Personalization	Recommending content, predictive content loading, improving user experience, …
Targeted marketing	Matching customers and offers, choosing marketing campaigns, cross-selling and up-selling, …
Content classification	Categorizing documents, matching hiring managers and resumes, …
Churn prediction	Finding customers who are likely to stop using the service, upgrade targeting, …
Customer support	Predictive routing of customer emails, social media listening, …
Predictive	Detect early signs of equipment failure

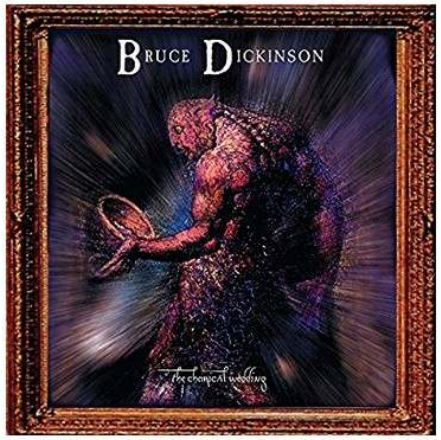
Product recommendation

amazon.com

A découvrir Voir plus







Inspiré par votre historique de navigation Voir plus

Customer support

Not all customer interactions can be solved in a self-service mode. Therefore, Amazon operates large customer support centers where Customer Service Representatives (CSR) handle customer requests.

Machine learning models are used to optimize the human interactions of these requests.

For example, they are used to route the customer call to the best CSR before the customer has even started to speak! They are also used again during the call.

Market surveillance

https://aws.amazon.com/solutions/case-studies/finra/

FINRA, the primary regulatory agency for broker-dealers in the US, uses AWS extensively in their IT operations, including Market Surveillance and Member Regulation.

For market surveillance, each night FINRA loads approximately 35 billion rows of data into Amazon S3 and Amazon EMR (up to 10,000 nodes) to monitor trading activity on exchanges and market centers in the US.

FINRA estimates it will save up to $20 million annually by using AWS instead of a physical data center infrastructure.

Cost optimization & predictive maintenance

An introduction to Machine Learning

What Machine Learning is all about

The purpose of Machine Learning is to find the best possible function F (aka model) that will accurately predict a result Y for a given input X

$$F(X)=Y$$

F = algorithm + data set + learned parameters

Types of learning

- Supervised Learning: data set is labeled

  Here are the historical home prices in New York based on zip code, number of rooms, etc. Now what's the price of this house?

- Unsupervised Learning: data set is unlabeled

  Here are 100,000 customer profiles. Cluster them in 100 groups and tell me what group this new customer should belong to.

- Reinforcement Learning: learning is goal-oriented

  Here is a graph of 1,000 nodes. Find the shortest path between node A and node B.

Common problems and algorithms

- Supervised Learning
  - Regression: predicting a value
    - *“What’s the probability of this motor failing in the next 24 hours?”*
    - Linear Regression
  - Classification
    - *“Is this tumor benign or malignant?”*
    - Logistic Regression, K-nearest neighbors, Support Vector Machines
- Unsupervised Learning
  - Clustering: grouping samples
    - “*What group of customers does this customer most look like?”*
    - K-means

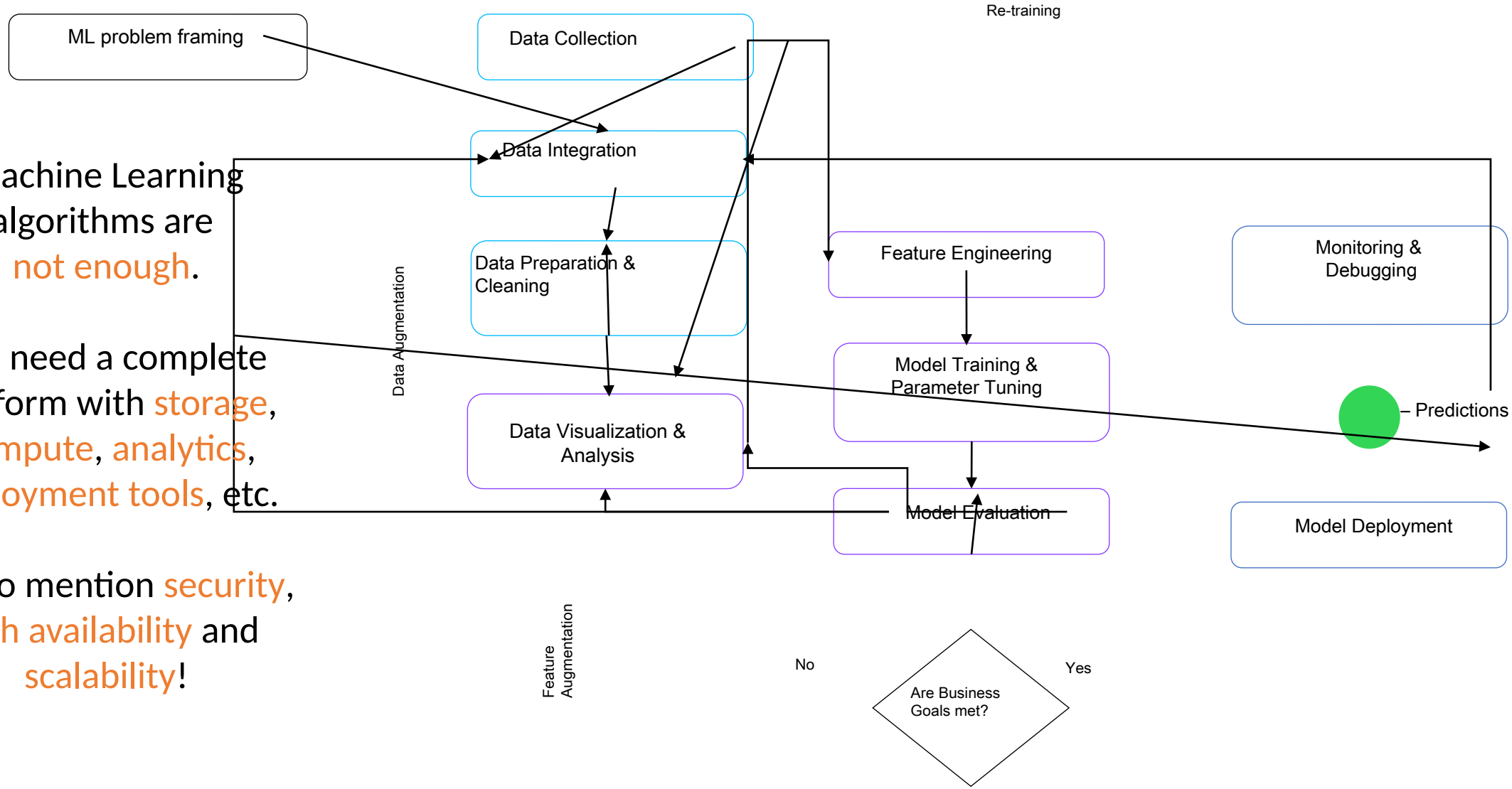
The Machine Learning process

Business Problem

Machine Learning algorithms are not enough.

You need a complete platform with storage, compute, analytics, deployment tools, etc.

Not to mention security, high availability and scalability!



Demo #1

Linear Regression… in Excel?

Demo #2

Linear Regression in Python with scikit-learn

Demo #3

Logistic Regression in Python with scikit-learn

Jeff Bezos' letter to Amazon shareholders

"We are solving problems with machine learning and artificial intelligence that were in the realm of science fiction for the last several decades. Natural language understanding, machine vision problems, it really is an amazing renaissance."

https://www.geekwire.com/2017/jeff-bezos-explains-amazons-artificial-intelligence-machine-learning-strategy/

Resources

https://aws.amazon.com/machine-learning

https://aws.amazon.com/blogs/ai

https://scikit-learn.org

https://machinelearningmastery.com/

https://www.coursera.org/learn/machine-learning

Machine Learning training @ Predicsis http://bit.ly/2yZjZHJ

https://medium.com/@julsimon

Thank you!

Julien Simon, AI Evangelist, EMEA
@julsimon