

Using Your Data to Give Your Customers Superpowers

Kirsten Westeinde
Senior Software Engineer
@kmkwesteinde
May 29, 2018



- 1. Identify a problem**
- 2. Frame your problem as a machine learning model**
- 3. Train your model**
- 4. Productionize your model**

Voila



Source: Giphy

Step 1

Identify a problem

**“When you have a hammer
Everything looks like a nail.”**

- Abraham Maslow

Rant over



Source: Giphy

Definitions

- Two types of machine learning problems:

Supervised & Unsupervised

- Two types of supervised learning problems:

Classification & Regression

**If you can frame your problem as a
classification or regression problem
with consistent input features,
It is a very good candidate to solve with
machine learning**



Examples

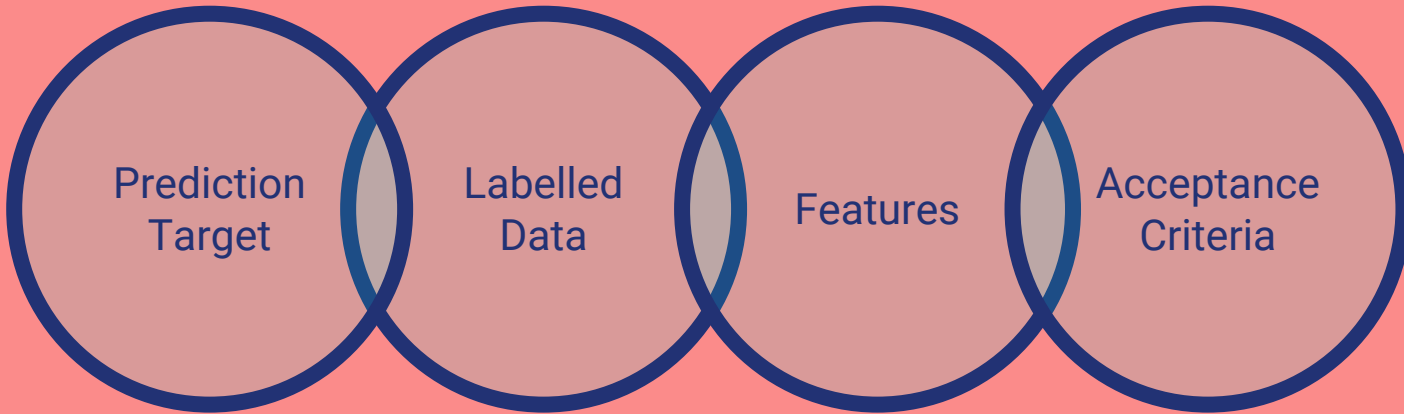


- Order fraud detection
- Deciding eligibility for Shopify Capital cash advances
- Classifying merchants by industry

Step 2

**Frame Your Problem as a
Machine Learning Model**

An ML algorithm needs:



A Prediction Target



Source: clipartpanda

Labelled Data



Source: buffalotechconsulting.com

Features



Source: komparu.com

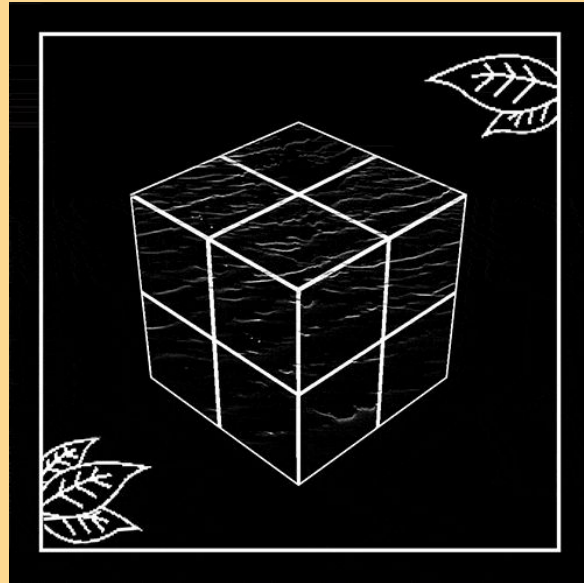
Acceptance Criteria



Source: [giphy.com](https://www.giphy.com)

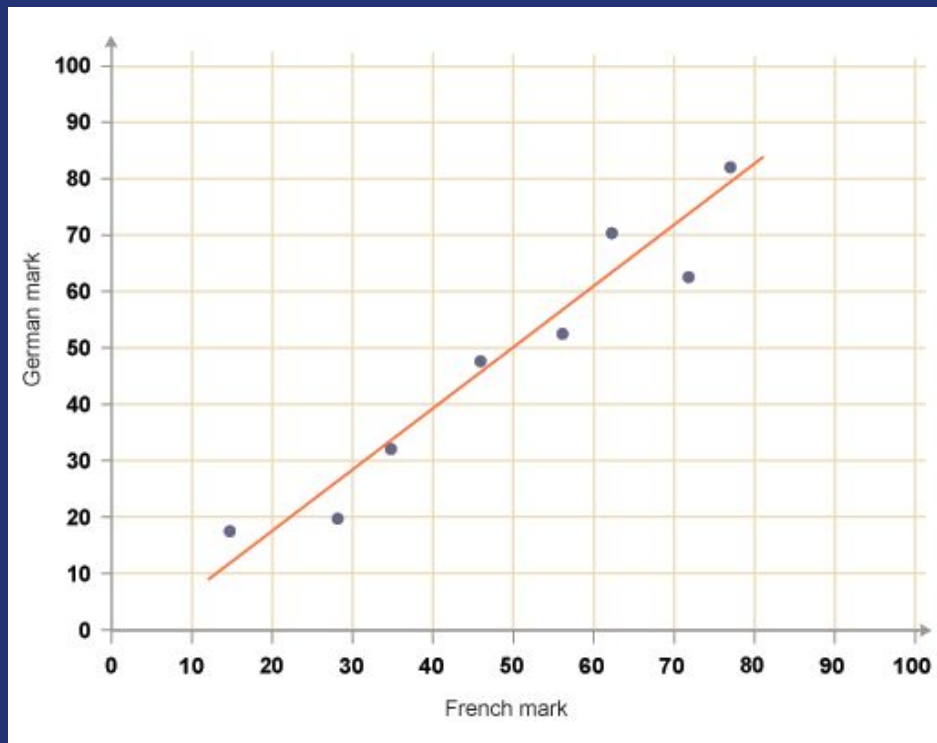
Step 3

Train Your Model



A very very oversimplified example

$$y = mx + b$$



Source: tes.com

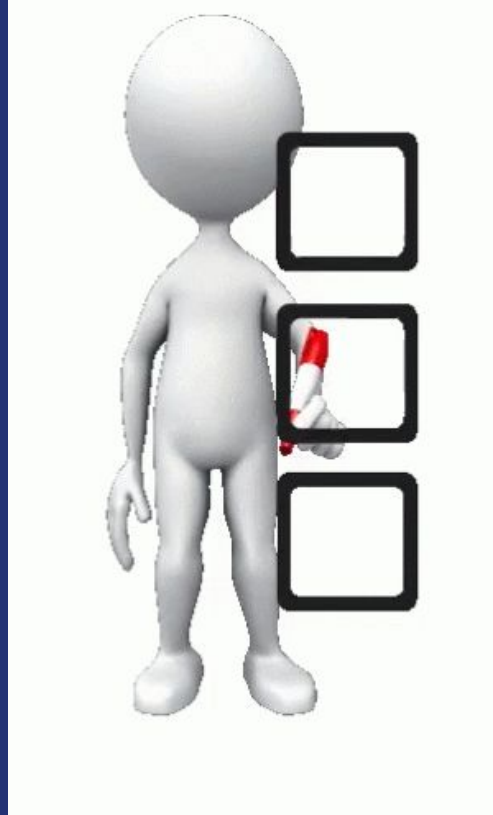
Further Reading:

- Models: Linear Regression, Neural Networks, Random forest, Support Vector Machines (SVM's), decision trees
- Libraries/Frameworks: Tensor Flow, PMML
- Detailed courses: Google Developer's "Machine Learning Crash Course", Coursera's "Machine Learning" course

Step 4

Productionize your model

Verification



Source: tenor.com

Deployment



A word cloud on a light purple background featuring various terms related to software deployment. The words are arranged in a cluster, with 'Verification' and 'Testing' being the largest and most central. Other prominent words include 'Reusability', 'Updates', 'Test', and 'Visibility'. Smaller words like 'Release', 'Real-time', 'A/B', 'Performance', 'set', 'Monitoring', and 'Reconciliation' are also present, scattered around the larger terms.

Reusability
Updates Test
Release Real-time
A/B
Verification
Testing
Performance set Monitoring
Reconciliation Visibility

The Future



Source giphy.com

Requirements for ML Products



| Required | Not Required |
|---|---|
| <ul style="list-style-type: none">• Understanding problem domain• Holistic knowledge of ML process• Software engineering/CS core concepts• Curiosity to experiment | <ul style="list-style-type: none">• Degree in Mathematics/Statistics• Formal data science training |

Thanks!

