Artificial Intelligence and Machine Learning: Challenges for the audit professional

What do you need to know

PANAGIOTIS DROUKAS
ISACA Athens Chapter President

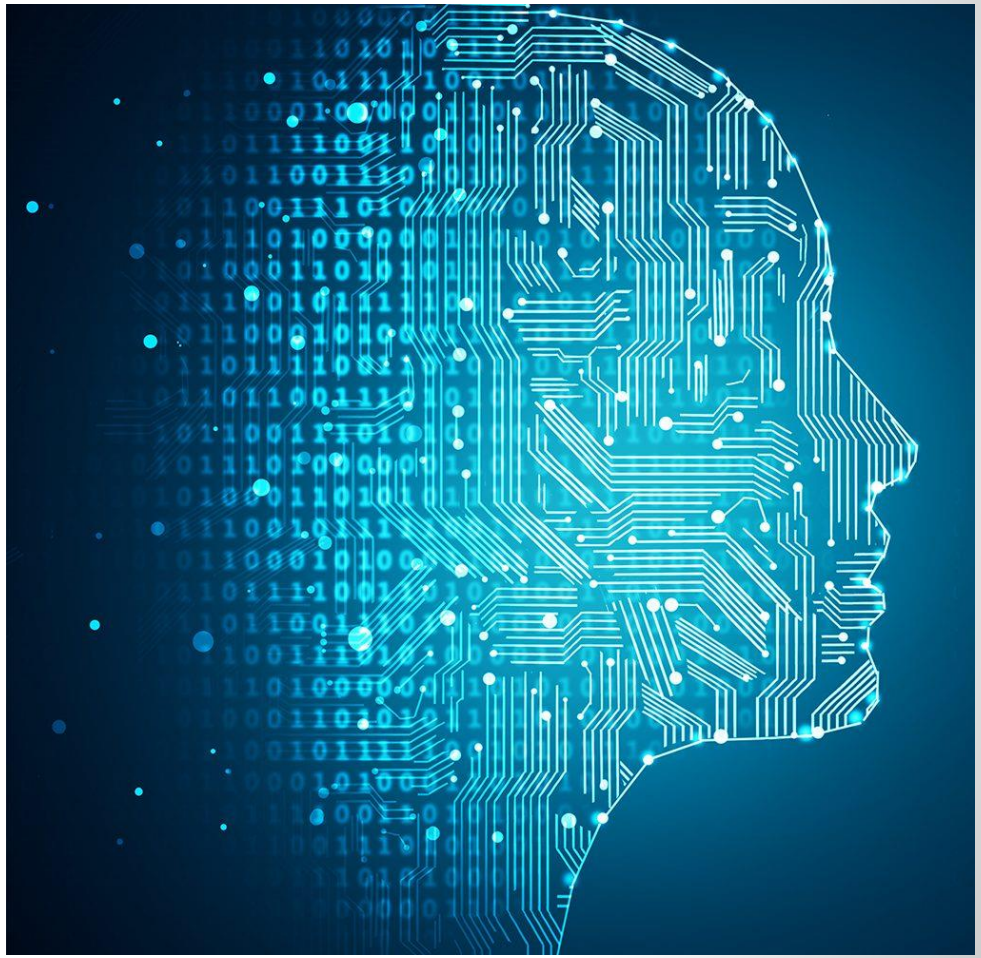
October 18th, 2018

ISACA®
Athens Chapter

ISACA Athens Chapter

Overview

- *What is machine learning?*
- *Why is it important?*
- *Non-technical introduction*
- *What are the types of machine learning?*
- *How does it pertain to auditors?*
- *Where can I learn more?*



ISACA®
Athens Chapter

Machine learning (ML) is a computer recognizing patterns without having to be explicitly programmed

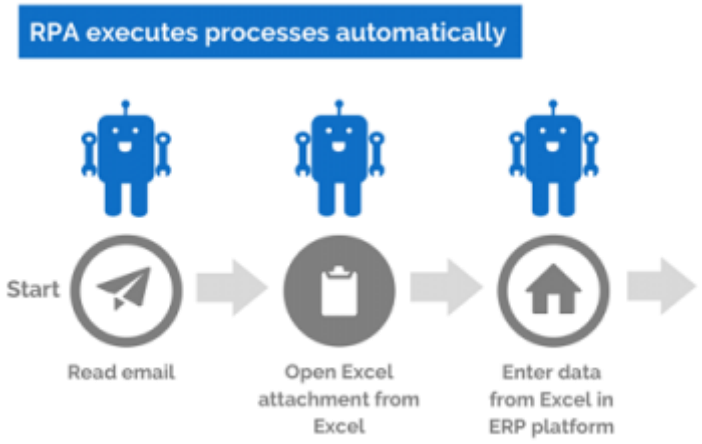
Why is machine learning important?

- *Disrupting businesses*
- *Revolutionize existing business models*
- *Driving economic growth*

NETFLIX

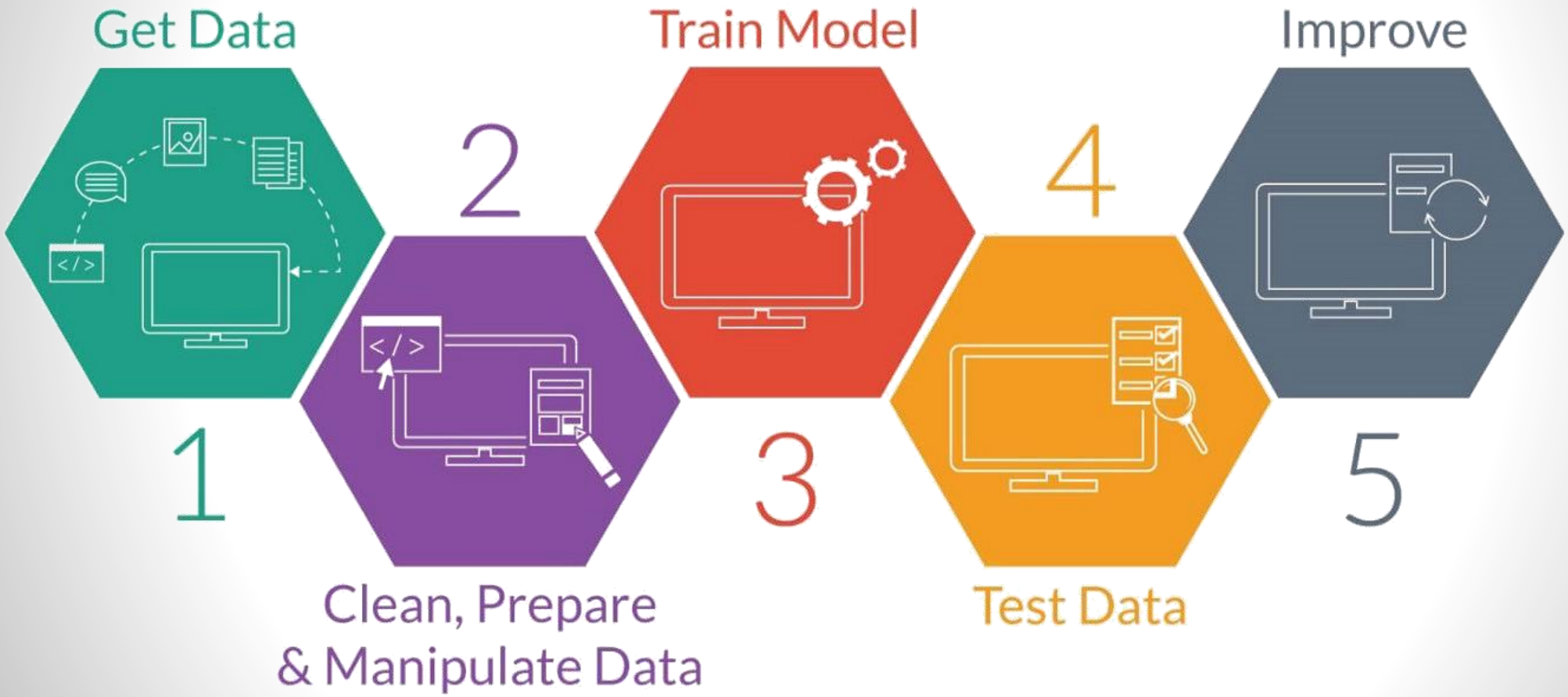
Spotify®

amazon.com®



WORLD ECONOMIC FORUM

A non-technical introduction



ISACA Athens Chapter

1. Get Data

Business Understanding

- *Why are we doing this?*
- *Framing of the problem we want to solve*
- ***Why** is this needed and **what** is the desired outcome*

Data Understanding

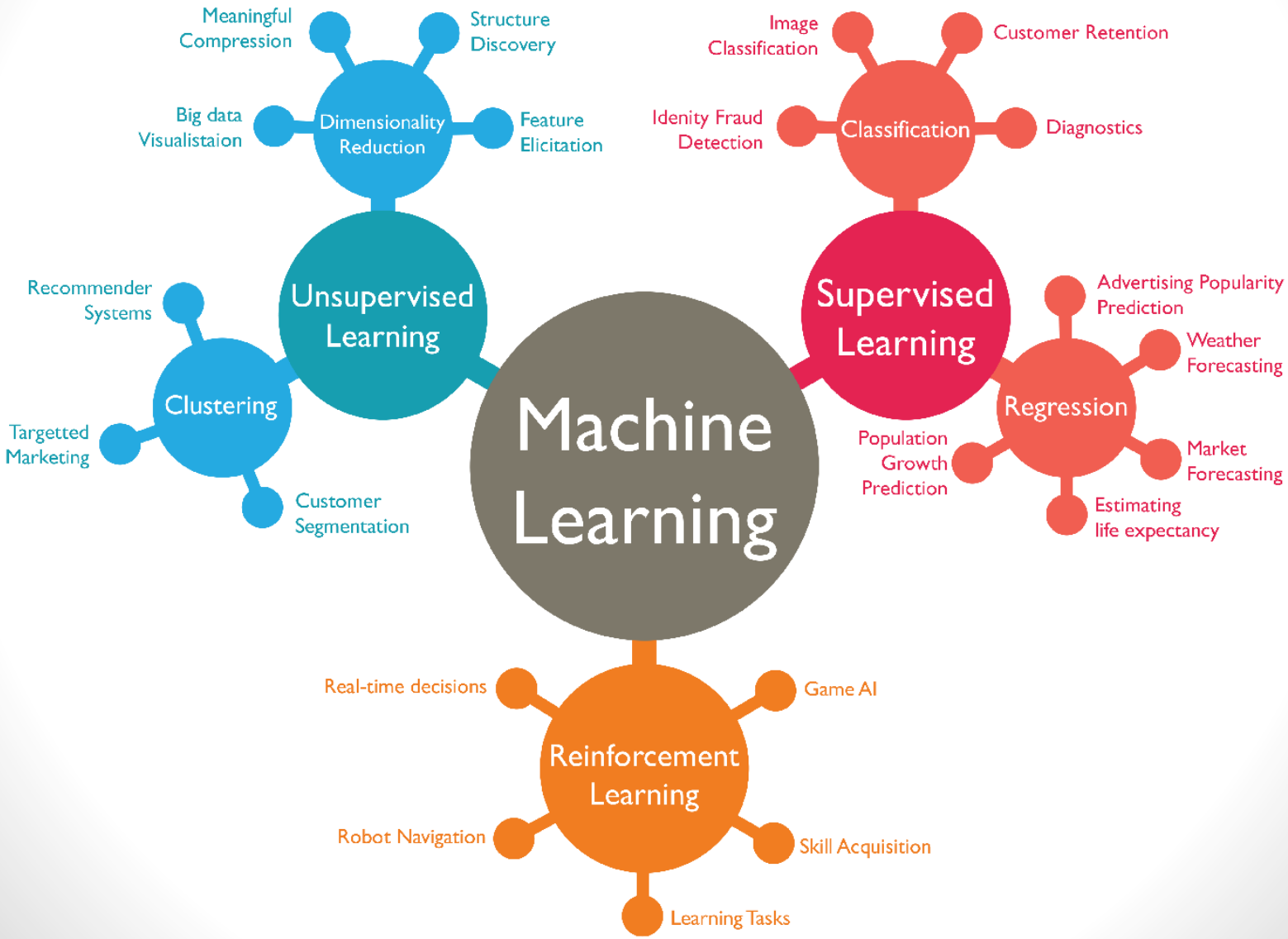
- *Where the data is coming from?*
- *In what format the data is stored (e.g. database, text file, webpage, etc.)?*
- *What is the scape of the data (e.g. cm or in)?*
- *Are transformations required (e.g. from in to cm)?*

2. Clean, prepare and manipulate data

Data preparation

- *Currently, close to 90% of what data scientists do*
- *"Munging"*
- ***Josh Wills** - "I am a data janitor. That is the sexiest job of the 21st century. It's very flattering but also a little baffling"*

3. Train model



ISACA Athens Chapter

4-5. Evaluation & Improvement

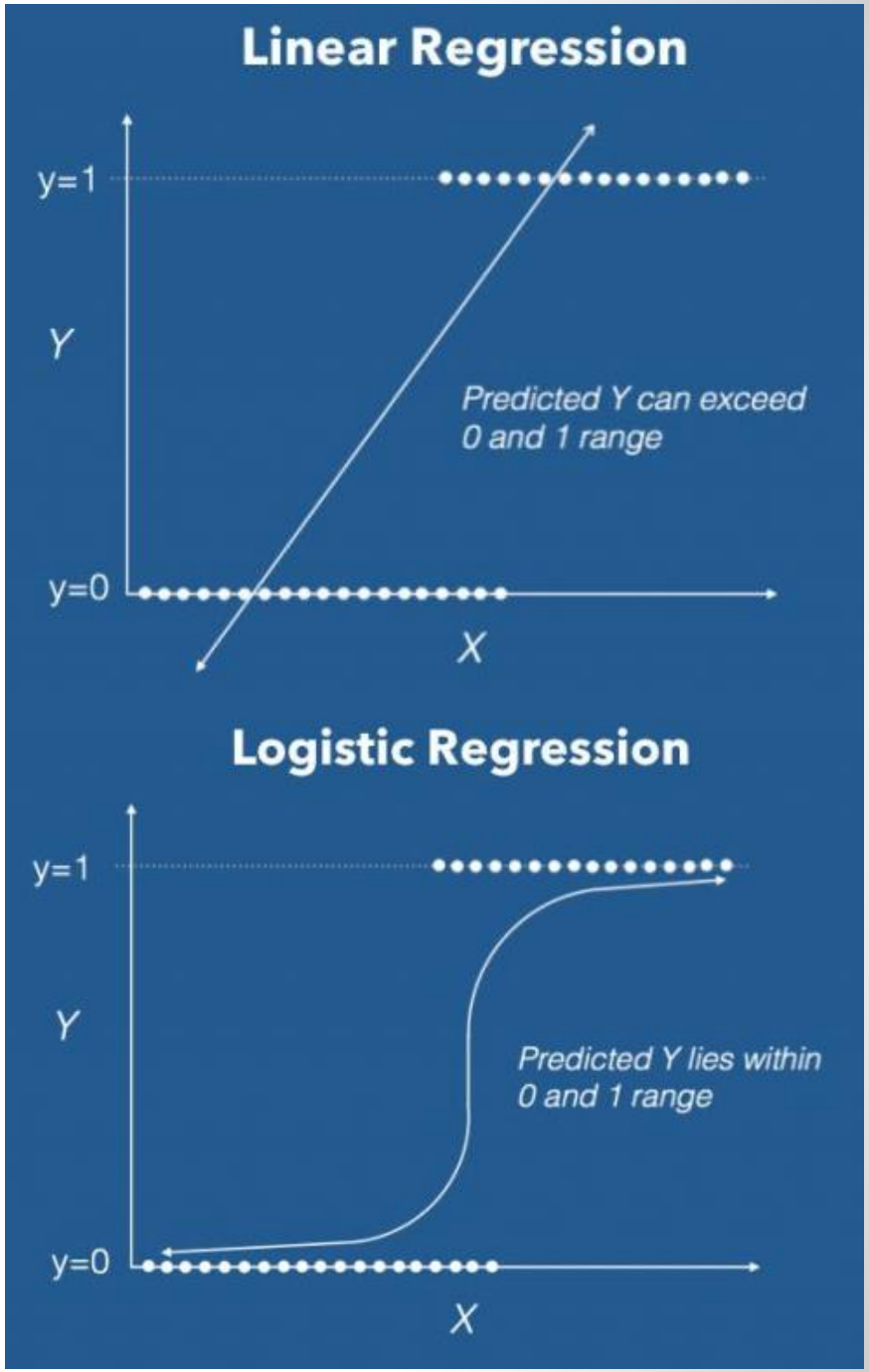
- *Accuracy*
- *Precision*
- *Does the selected model solve the problem?*
- *What adjustments are needed?*
- *How can we improve and automate the data gathering process?*

Case study: Classification

Business Case

- *In an airport, we have a machine that can automatically measure **height, weight** and **shoe size***

- *We need to figure out if the person passing through the machine is **male** or **female** in order to direct him / her to go to the correct human agent for spot inspection*



Train model

15 lines (9 sloc) | 470 Bytes

Raw | Blame | History

```python
from sklearn.linear_model import LogisticRegression

neigh = LogisticRegression()

#[height, weight, shoe_size]
X = [[181, 80, 44], [177, 70, 43], [160, 60, 38], [154, 54, 37], [166, 65, 40], [190, 90, 47], [175, 64, 39],
     [177, 70, 40], [159, 55, 37], [171, 75, 42], [181, 85, 43]]

Y = ['male', 'male', 'female', 'female', 'male', 'male', 'female', 'female', 'female', 'male', 'male']

neigh.fit(X, Y)

prediction = neigh.predict([[190, 70, 43]])

print prediction
```

Use cases in assurance and compliance

Unsupervised learning

- *Credit card fraudulent transactions*
- *Unsupervised journal entry anomaly detection*
- *Clustering of invoice / payments data for outliers*

Supervised learning

- *Credit default rates*
- *Expense report investigation*
- *Fraudulent claims*

ISACA Athens Chapter

Where can I learn more?

- A visual introduction to machine learning
  http://www.r2d3.us/visual-intro-to-machine-learning-part-1/

- *4-Steps to Get Started in Machine Learning*
  https://machinelearningmastery.com/4-steps-to-get-started-in-machine-learning/

- Weka 3: Data Mining Software in Java
  https://www.cs.waikato.ac.nz/ml/weka/

- Machine learning in Python
  http://scikit-learn.org/stable/

ISACA Athens Chapter

ISACA at a Glance: Global Reach, Local Impact*

SERVING MORE THAN

520,000

ENGAGED PROFESSIONALS

215+ CHAPTERS

WORLDWIDE

MEMBERS IN

190 COUNTRIES

11

INDUSTRY-LEADING CONFERENCES

1,000+

KNOWLEDGE ASSETS

Questions ?

Thank you!