

# Applied Machine Learning in Python

Mathijs de Jong

Turing Students Rotterdam

Dutch Data Science Week, 4 June 2019

# Table of Contents

1. Artificial Intelligence & Machine Learning — Tomorrow's Technologies
2. Python — The Language of Tomorrow
3. Kaggle — The AirBnB for Data Scientists

# Table of Contents

1. Artificial Intelligence & Machine Learning — Tomorrow's Technologies
2. Python — The Language of Tomorrow
3. Kaggle — The AirBnB for Data Scientists

## Definition: Artificial intelligence

Artificial intelligence (AI) is typically defined as the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and problem solving.

## Definition: Artificial intelligence

Artificial intelligence (AI) is typically defined as the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and problem solving.

- Examples of technologies that enable AI to solve business problems are
  - robotics and autonomous vehicles,
  - computer vision,
  - language,
  - virtual agents, and
  - machine learning.

## Definition: Machine Learning

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. (...) Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.

## Definition: Machine Learning

Machine learning (ML) is the scientific study of **algorithms and statistical models** that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. (...) Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.

## Definition: Machine Learning

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively **without using explicit instructions**, relying on patterns and inference instead. (...) Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.



## Definition: Machine Learning

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, **relying on patterns and inference instead**. (...) Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.

## Definition: Machine Learning

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. (...) Machine learning algorithms build a **mathematical model based on sample data, known as "training data"**, in order to make predictions or decisions without being explicitly programmed to perform the task.

## Definition: Machine Learning

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. (...) Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.

# Types of Analytics

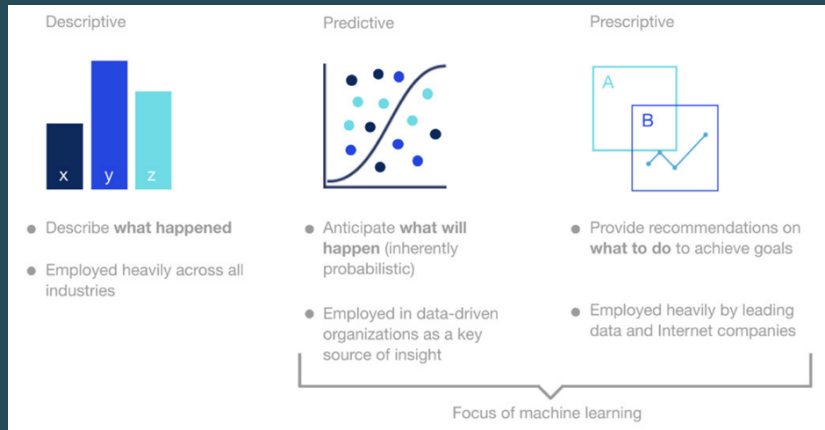
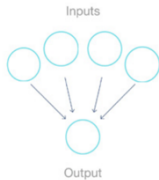


Figure: Types of analytics in order of complexity

# Major Types of Machine Learning

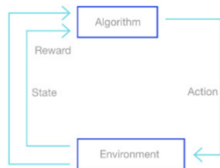
Supervised learning



Unsupervised learning

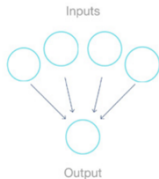


Reinforcement learning



# Major Types of Machine Learning

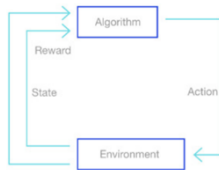
## Supervised learning



## Unsupervised learning



## Reinforcement learning



### What it is

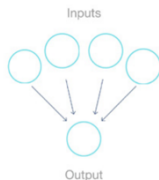
An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output (eg, how the inputs “time of year” and “interest rates” predict housing prices)

An algorithm explores input data without being given an explicit output variable (eg, explores customer demographic data to identify patterns)

An algorithm learns to perform a task simply by trying to maximize rewards it receives for its actions (eg, maximizes points it receives for increasing returns of an investment portfolio)

# Major Types of Machine Learning

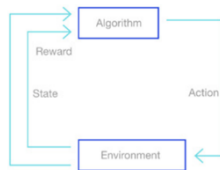
## Supervised learning



## Unsupervised learning



## Reinforcement learning



You know how to classify the input data and the type of behavior you want to predict, but you need the algorithm to calculate it for you on new data

You do not know how to classify the data, and you want the algorithm to find patterns and classify the data for you

You don't have a lot of training data; you cannot clearly define the ideal end state; or the only way to learn about the environment is to interact with it

# Major Types of Machine Learning



1. A human labels every element of the input data (eg, in the case of predicting housing prices, labels the input data as “time of year,” “interest rates,” etc) and defines the output variable (eg, housing prices)
2. The algorithm is trained on the data to find the connection between the input variables and the output
3. Once training is complete—typically when the algorithm is sufficiently accurate—the algorithm is applied to new data

1. The algorithm receives unlabeled data (eg, a set of data describing customer journeys on a website)
2. It infers a structure from the data
3. The algorithm identifies groups of data that exhibit similar behavior (eg, forms clusters of customers that exhibit similar buying behaviors)

1. The algorithm takes an action on the environment (eg, makes a trade in a financial portfolio)
2. It receives a reward if the action brings the machine a step closer to maximizing the total rewards available (eg, the highest total return on the portfolio)
3. The algorithm optimizes for the best series of actions by correcting itself over time



# Characteristics of the Machine Learning Approach

# Characteristics of the Machine Learning Approach

- Data driven

# Characteristics of the Machine Learning Approach

- Data driven
  - 1 Start simple

# Characteristics of the Machine Learning Approach

- Data driven
  - 1 Start simple
  - 2 Test the model

# Characteristics of the Machine Learning Approach

- Data driven

- 1 Start simple
- 2 Test the model
- 3 If (Performance is good enough):  
    DONE  
    else:  
        Go back to step 1.

# Characteristics of the Machine Learning Approach

- Data driven
  - 1 Start simple
  - 2 Test the model
  - 3 If (Performance is good enough):  
    DONE  
    else:  
        Go back to step 1.
- When do machine learning methods excel?

# Characteristics of the Machine Learning Approach

- Data driven
  - 1 Start simple
  - 2 Test the model
  - 3 If (Performance is good enough):  
    DONE  
    else:  
        Go back to step 1.
- When do machine learning methods excel?
  - 1 Large data sets

# Characteristics of the Machine Learning Approach

- Data driven

- 1 Start simple
- 2 Test the model
- 3 If (Performance is good enough):  
    DONE  
    else:  
        Go back to step 1.

- When do machine learning methods excel?

- 1 Large data sets
- 2 Unstructured and complex problems



# Table of Contents

1. Artificial Intelligence & Machine Learning — Tomorrow's Technologies
2. Python — The Language of Tomorrow
3. Kaggle — The AirBnB for Data Scientists

# Python Programming Language

- Easy to use, flexible and accessible

---

The Economist — Python is becoming the world's most popular coding language

CBT Nuggets — Why Data Scientists Love Python

Quora — Why do we use Python for data analytics?

# Python Programming Language

- Easy to use, flexible and accessible
- Multi-purpose language

---

The Economist — Python is becoming the world's most popular coding language

CBT Nuggets — Why Data Scientists Love Python

Quora — Why do we use Python for data analytics?

# Python Programming Language

- Easy to use, flexible and accessible
- Multi-purpose language
- Scalability

---

The Economist — Python is becoming the world's most popular coding language

CBT Nuggets — Why Data Scientists Love Python

Quora — Why do we use Python for data analytics?

# Python Programming Language

- Easy to use, flexible and accessible
- Multi-purpose language
- Scalability
- Mass adoption
  - Choice of data science libraries
  - Great community support

---

The Economist — Python is becoming the world's most popular coding language

CBT Nuggets — Why Data Scientists Love Python

Quora — Why do we use Python for data analytics?

# Table of Contents

1. Artificial Intelligence & Machine Learning — Tomorrow's Technologies
2. Python — The Language of Tomorrow
3. Kaggle — The AirBnB for Data Scientists

# About Kaggle

- Community of data scientists and machine learners

# About Kaggle

- Community of data scientists and machine learners
- Platform for machine learning competitions



# About Kaggle

- Community of data scientists and machine learners
- Platform for machine learning competitions
  - Public data sets

# About Kaggle

- Community of data scientists and machine learners
- Platform for machine learning competitions
  - Public data sets
  - Real world data science problems

# About Kaggle

- Community of data scientists and machine learners
- Platform for **machine learning competitions**
  - **Public data sets**
  - Real world data science problems
  - Learn together in a community

# About Kaggle

- Community of data scientists and machine learners
- Platform for machine learning competitions
  - Public data sets
  - Real world data science problems
  - Learn together in a community
  - Place to build a portfolio and find your future job

# About Kaggle

- Community of data scientists and machine learners
- Platform for machine learning competitions
  - Public data sets
  - Real world data science problems
  - Learn together in a community
  - Place to build a portfolio and find your future job
- Data science education environment

# The Future Starts Now

- For the remainder of this session, you can work on Kaggle tutorials

- 1 Register an account at Kaggle

- 2 Start with one of the micro-courses on Kaggle:

Beginners: Python Micro-Course

Intermediates: Intro to Machine Learning Micro-Course

Others: Start with one of the other tutorials. Select the one you think is most relevant for you right now.

# Tomorrow and Afterwards

- Learn by doing

# Tomorrow and Afterwards

- Learn by doing
- Continue with the micro-courses on Kaggle



# Tomorrow and Afterwards

- Learn by doing
- Continue with the micro-courses on Kaggle
- Compete in Kaggle competitions

# Tomorrow and Afterwards

- Learn by doing
- Continue with the micro-courses on Kaggle
- Compete in Kaggle competitions
- Work on your own projects or join other projects

# Tomorrow and Afterwards

- Learn by doing
- Continue with the micro-courses on Kaggle
- Compete in Kaggle competitions
- Work on your own projects or join other projects
- Follow one of Turing Students Rotterdam's bootcamp courses

# Tomorrow and Afterwards

- Learn by doing
- Continue with the micro-courses on Kaggle
- Compete in Kaggle competitions
- Work on your own projects or join other projects
- Follow one of Turing Students Rotterdam's bootcamp courses
  - Python for Data Science

# Tomorrow and Afterwards

- Learn by doing
- Continue with the micro-courses on Kaggle
- Compete in Kaggle competitions
- Work on your own projects or join other projects
- Follow one of Turing Students Rotterdam's bootcamp courses
  - Python for Data Science
  - R for Data Science

# Tomorrow and Afterwards

- Learn by doing
- Continue with the micro-courses on Kaggle
- Compete in Kaggle competitions
- Work on your own projects or join other projects
- Follow one of Turing Students Rotterdam's bootcamp courses
  - Python for Data Science
  - R for Data Science
  - Data Analytics in R

# Tomorrow and Afterwards

- Learn by doing
- Continue with the micro-courses on Kaggle
- Compete in Kaggle competitions
- Work on your own projects or join other projects
- Follow one of Turing Students Rotterdam's bootcamp courses
  - Python for Data Science
  - R for Data Science
  - Data Analytics in R
  - Bletchley Machine Learning Bootcamp