Applied Machine Learning in Python

Mathijs de Jong

Turing Students Rotterdam

Dutch Data Science Week, 4 June 2019

TS Turing Students Rotterdam

Table of Contents

Turing Students Rotterdam

Table of Contents

TS **Turing**
Students
Rotterdam

Artificial Intelligence

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.

McKinsey Analytics - An executives guide to AI

Turing Students Rotterdam

Artificial Intelligence

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.

McKinsey Analytics - An executives guide to AI

TS | **Turing Students**
Rotterdam

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.

Wikipedia - Machine Learning

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.

Wikipedia - Machine Learning

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.

Wikipedia - Machine Learning

Turing Students
Rotterdam

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.

Wikipedia - Machine Learning

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.

Wikipedia - 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.

Wikipedia - Machine Learning
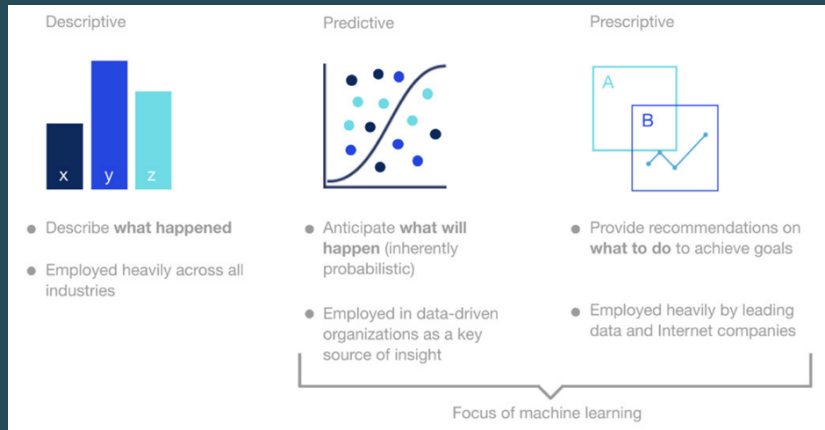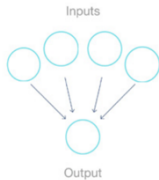
TS Turing Students
Rotterdam

Figure: Types of analytics in order of complexity

Major Types of Machine Learning

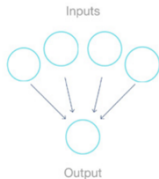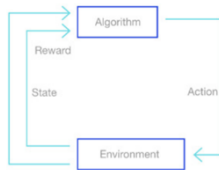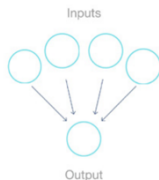| Supervised learning	Unsupervised learning	Reinforcement learning
What it is	An algorithm uses training data and feedback from humans to learn the relationship of given inputs to a given output (eg, how the inputs "time of year" and "interest rates" predict housing prices)	An algorithm explores input 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)
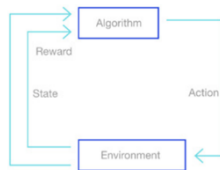
Major Types of Machine Learning

How it works

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

- Data driven

Characteristics of the Machine Learning Approach

- Data driven
  1. Start simple

TS Turing
Students
Rotterdam

Characteristics of the Machine Learning Approach

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

TS Turing
Students
Rotterdam

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?

TS Turing Students Rotterdam

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

TS Turing Students Rotterdam

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

TS | Turing
Students
Rotterdam

- 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?

TS Turing Students
Rotterdam

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?

Turing Students Rotterdam

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

About Kaggle

- Community of data scientists and machine learners

Getting Started — What is Kaggle, Why I Participate, What is the Impact?

Mathijs de Jong (TSR)   Applied Machine Learning in Python   DDSW 2019   14 / 16

Turing
Students
Rotterdam

About Kaggle

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

Getting Started — What is Kaggle, Why I Participate, What is the Impact?

Mathijs de Jong (TSR)　　　Applied Machine Learning in Python　　　DDSW 2019　　14 / 16

Turing
Students
Rotterdam

About Kaggle

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

Getting Started — What is Kaggle, Why I Participate, What is the Impact?

Mathijs de Jong (TSR)   Applied Machine Learning in Python   DDSW 2019   14 / 16

About Kaggle

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

Getting Started — What is Kaggle, Why I Participate, What is the Impact?

Mathijs de Jong (TSR)   Applied Machine Learning in Python   DDSW 2019   14 / 16

TS Turing Students Rotterdam

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

Turing Students Rotterdam

Getting Started — What is Kaggle, Why I Participate, What is the Impact?

Mathijs de Jong (TSR)   Applied Machine Learning in Python   DDSW 2019   14 / 16

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

Getting Started — What is Kaggle, Why I Participate, What is the Impact?

Mathijs de Jong (TSR)   Applied Machine Learning in Python   DDSW 2019   14 / 16

TS Turing Students
Rotterdam

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

Getting Started — What is Kaggle, Why I Participate, What is the Impact?

Mathijs de Jong (TSR)   Applied Machine Learning in Python   DDSW 2019   14 / 16

TS Turing Students Rotterdam

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.

TS Turing Students
Rotterdam

- 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

TS Turing Students
Rotterdam

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

TS **Turing Students** Rotterdam

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

TS | **Turing Students** Rotterdam

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

TS **Turing Students** Rotterdam

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

TS | **Turing Students** Rotterdam