

A PRÁTICA DE UM CIENTISTA DE DADOS



Ao vivo em 2 horas
17 de abril 19:00



Definir lembrete

VIVIANE MARTINS

MÃO
massa



FABIO LIMA

sexta
17/04
19h

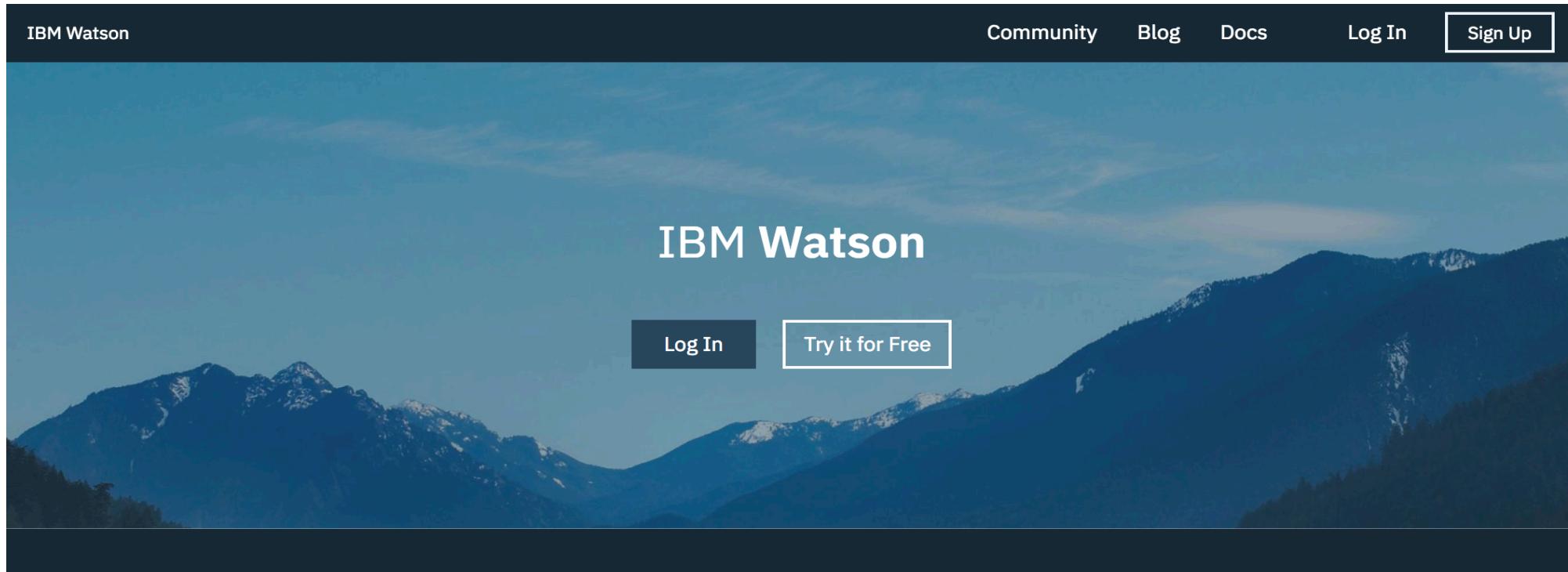


#VamosJuntos

double

Access the IBM Watson Studio

<https://dataplatform.cloud.ibm.com/>



Data Scientist

DATA SCIENCE SKILLSET

The diagram illustrates the Data Science Skillset as a Venn diagram with four overlapping circles:

- Hacking Skills** (Red circle)
- Substantive Expertise** (Light Blue circle)
- Machine Learning** (Orange circle)
- Math and Statistics Knowledge** (Yellow circle)

The central intersection of all four circles is labeled **Data Science**. To the right of the diagram is a vertical stack of icons representing the **Danger zone**, which occurs when Hacking skills are combined with Substantive scientific expertise without rigorous methods.

Data science, due to its interdisciplinary nature, requires an intersection of abilities: **Hacking skills, math and statistics knowledge**, and **substantive expertise** in a field of science.

Hacking skills are necessary for working with massive amounts of electronic data that must be acquired, cleaned and manipulated.

Math and statistics knowledge allows a data scientist to choose appropriate methods and tools in order to extract insights from data

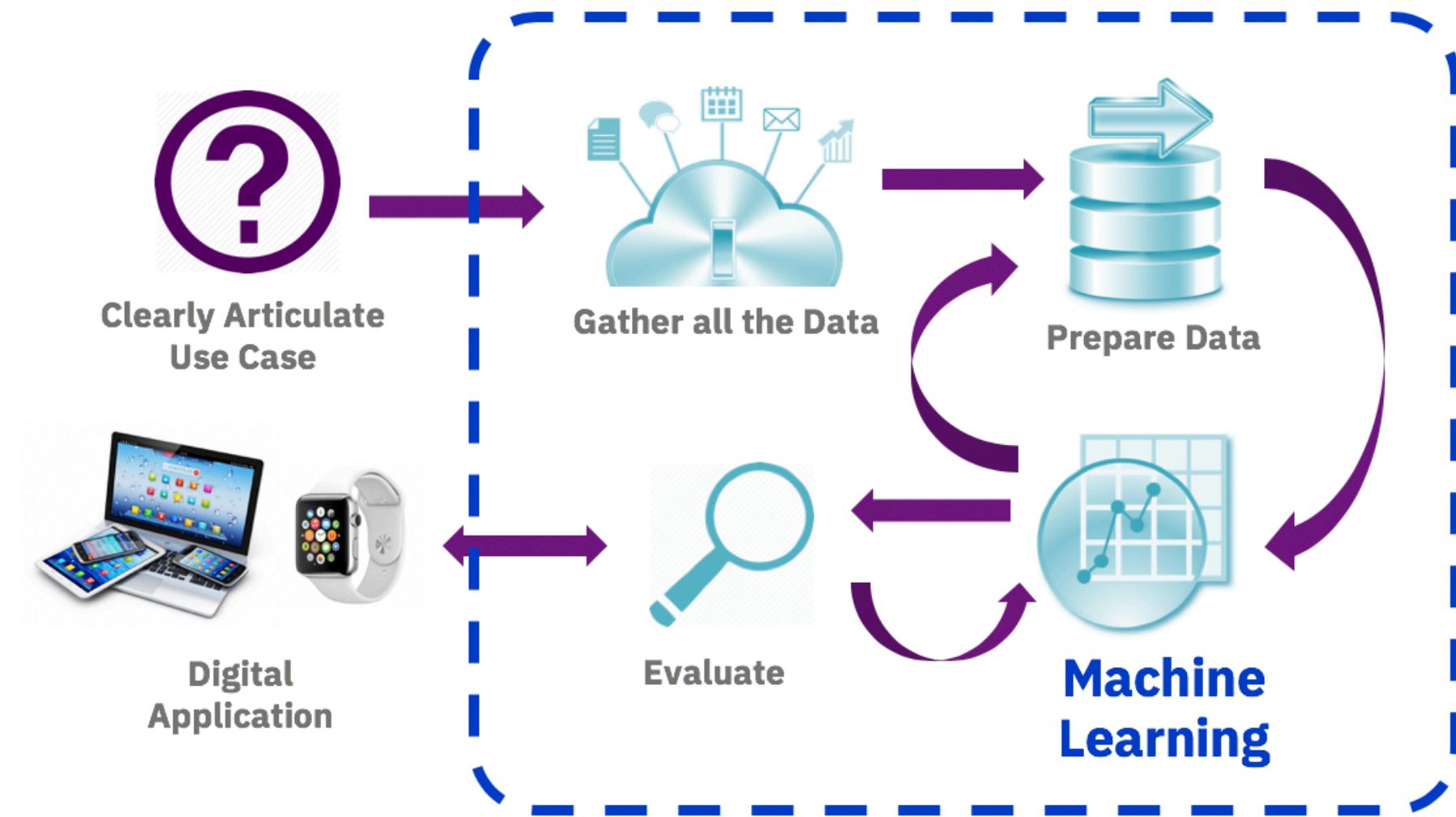
Substantive expertise in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.

Traditional research lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.

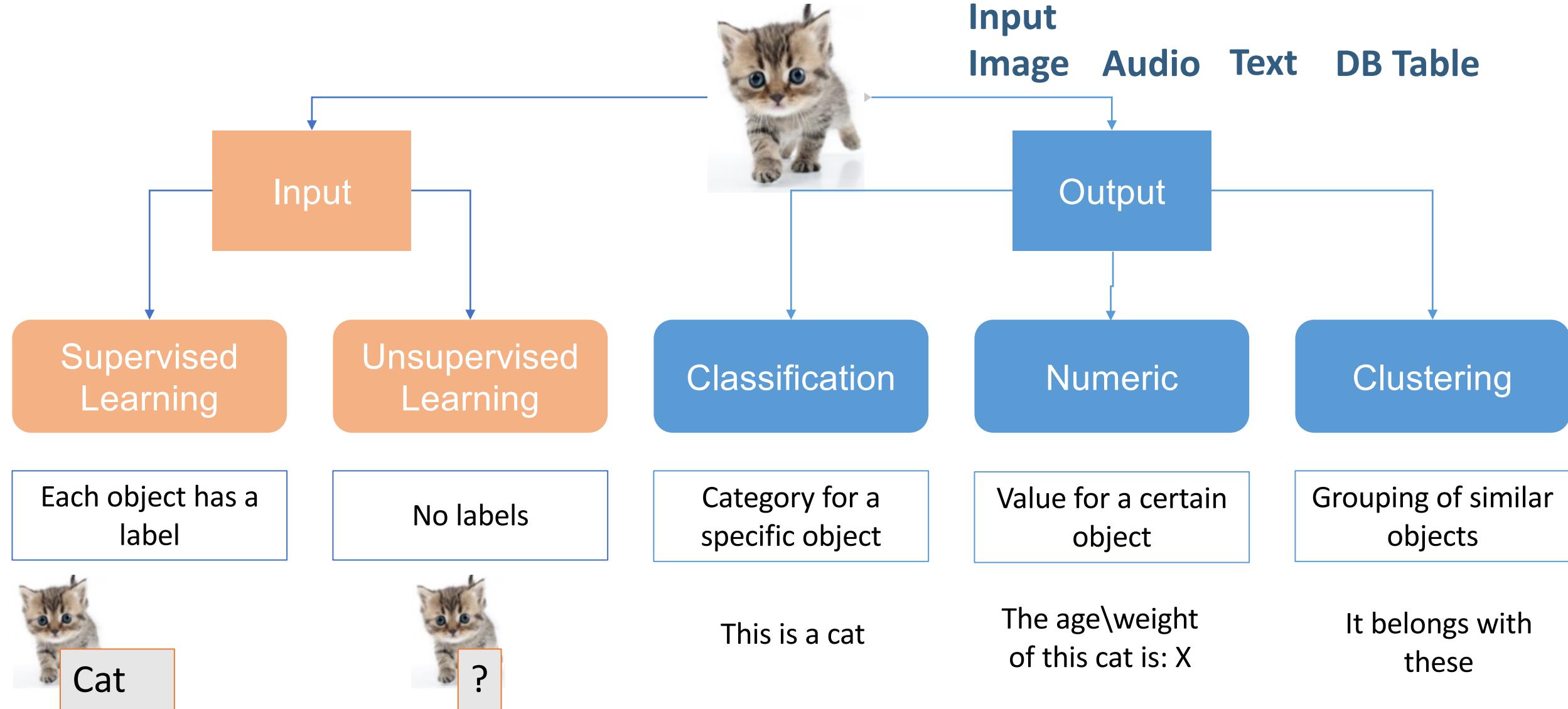
Machine learning stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.

Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Data Science - Fundamentals



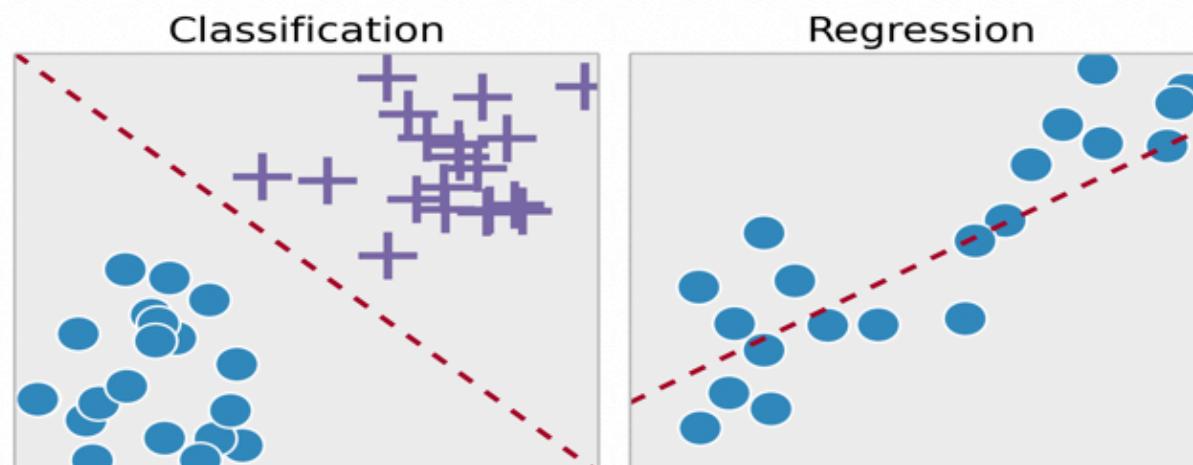
Machine Learning – Fundamentals



Supervised Learning

- Most Basic Type of Machine Learning
- Known ‘Target’ Values
- Either:

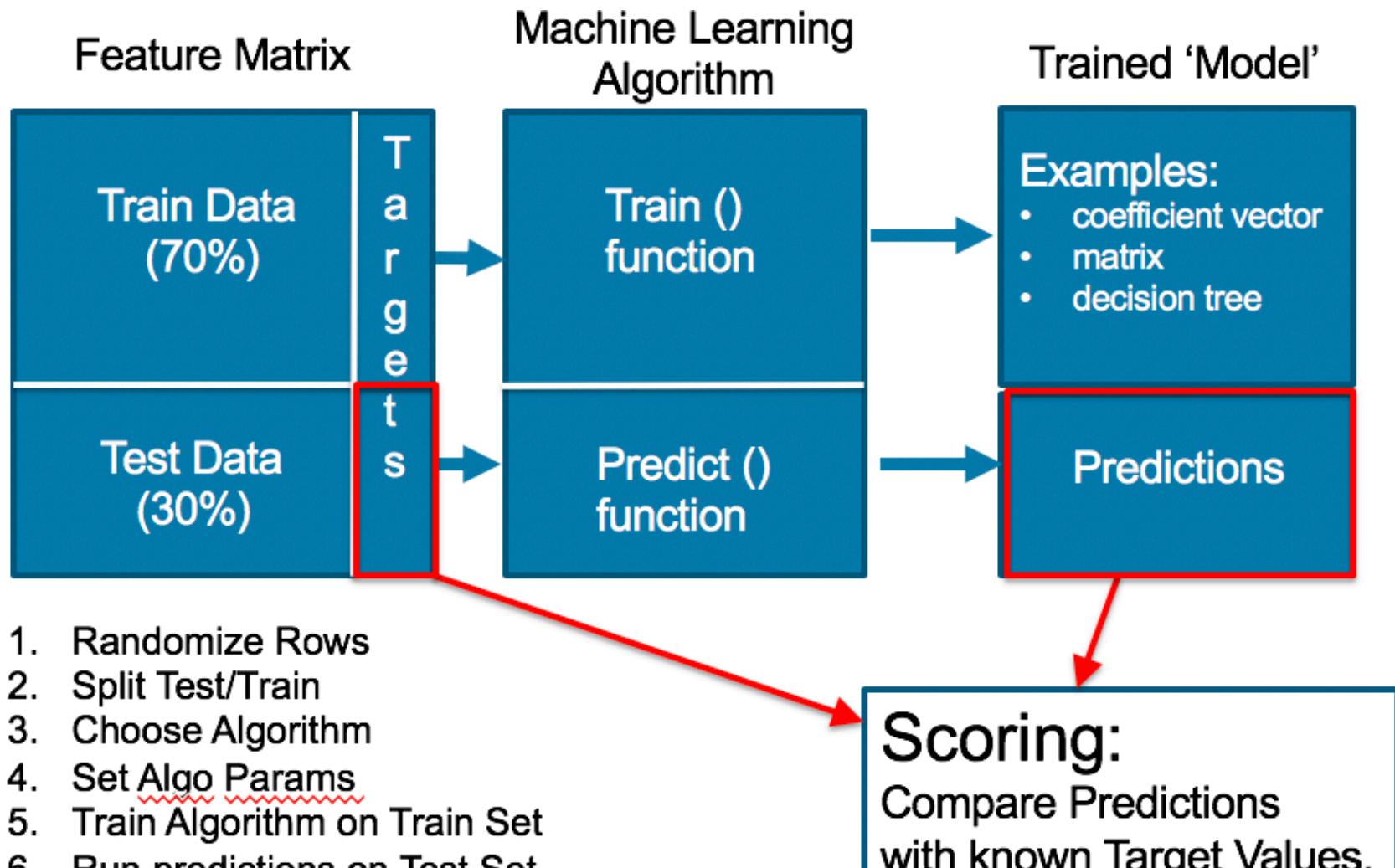
- Classification
- Regression



Source: <http://ipython-books.github.io/featured-04/>

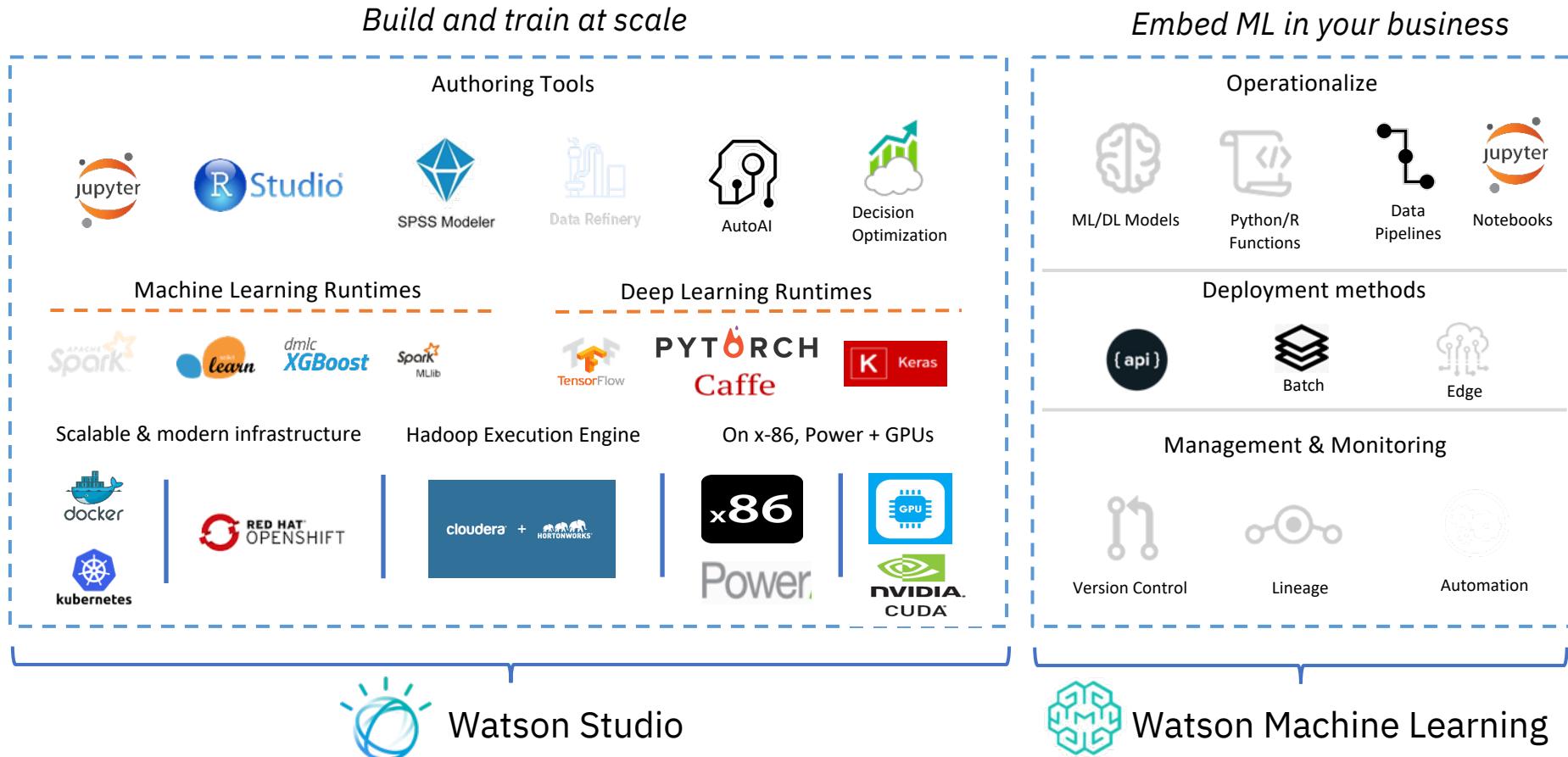
- Other types of ML include:
 - Unsupervised Learning, Recommender Systems, Anomaly Detection, Reinforcement Learning, Deep Learning, Cognitive Systems

Training and Scoring Process



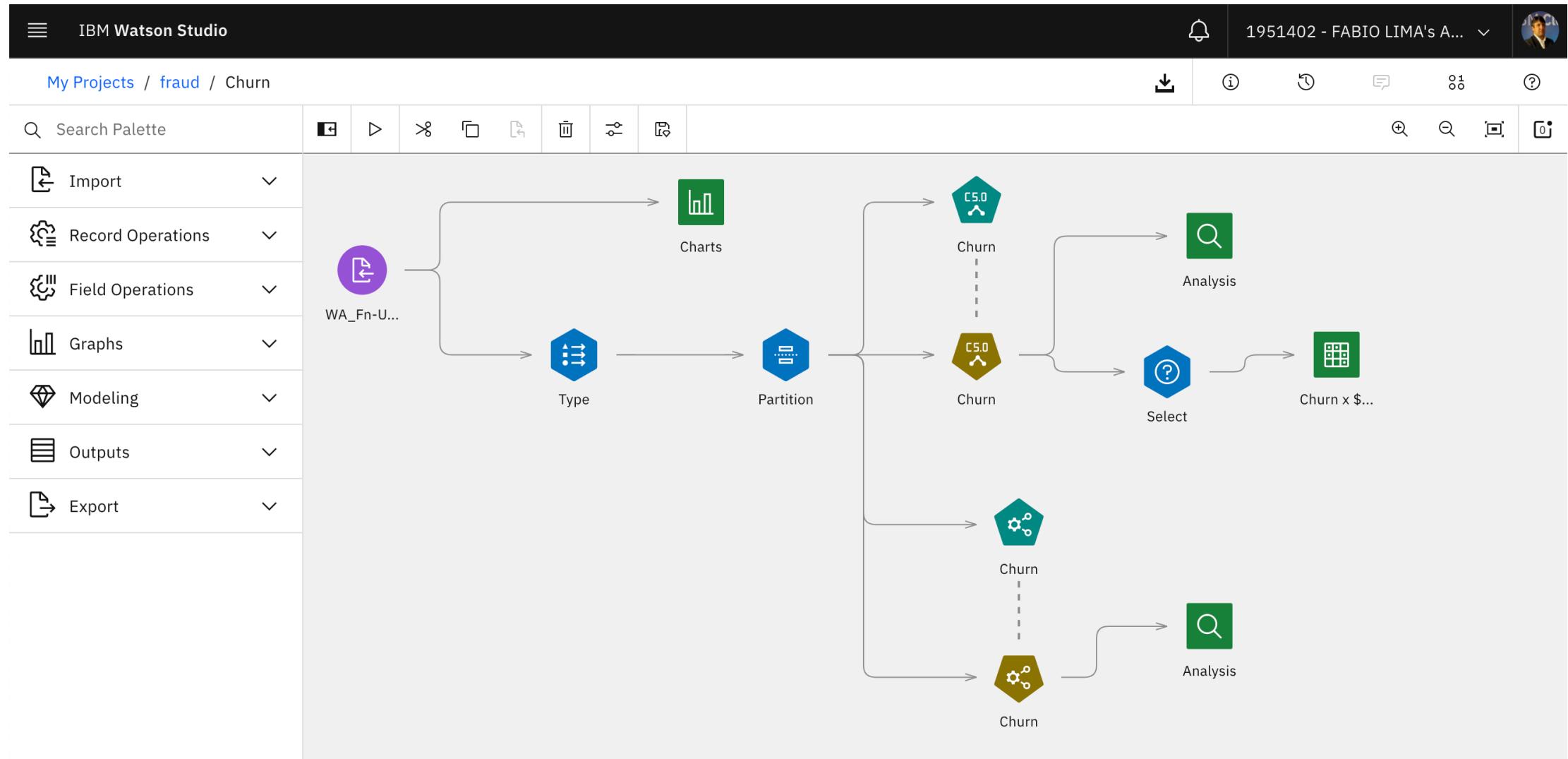
What is IBM Watson Studio & Machine Learning

Accelerate your time to value with Data Science



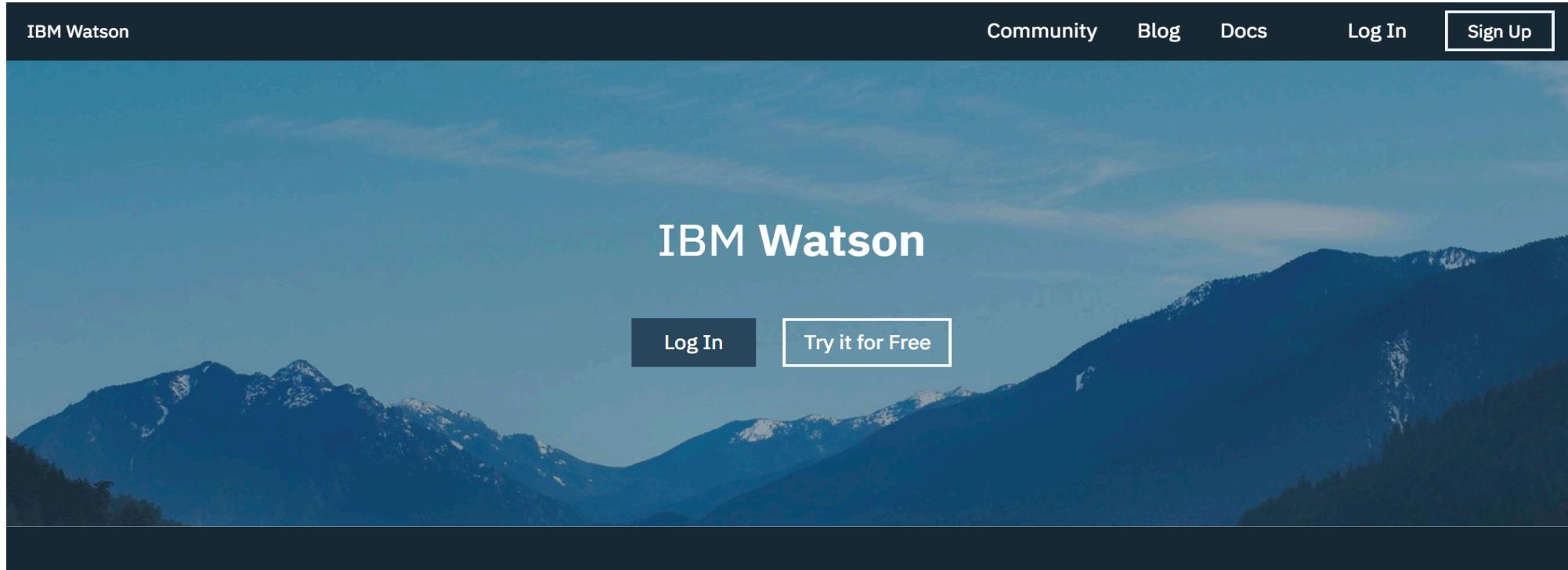
Modeler flows

SPSS Modeler



Access the IBM Watson Studio

<https://dataplatform.cloud.ibm.com/>



You will need to download and unzip this GitHub repository:

https://github.com/fabioflima/mao_massa

The screenshot shows the GitHub repository page for the repository `fabioflima / mao_massa`. The page includes the following elements:

- Header:** Includes the repository name, a watch count of 1, a star count of 0, and a fork count of 0.
- Navigation:** A navigation bar with links to Code, Issues (0), Pull requests (0), Actions, Projects (0), Wiki, Security, Insights, and Settings.
- Title:** The repository title is "Workshop Mão na Massa".
- Contributor Information:** Shows 1 commit, 1 branch, 0 packages, 0 releases, and 1 contributor.
- Action Buttons:** Buttons for Branch: master ▾, New pull request, Create new file, Upload files, Find file, and Clone or download ▾.
- Recent Activity:** A section showing a recent upload by `fabioflima` and the latest commit at `9cf47dd`.



<https://www.kaggle.com/blastchar/telco-customer-churn>

The image shows the dataset landing page for 'Telco Customer Churn' on Kaggle. At the top left is the 'Dataset' icon. On the right is a yellow circular badge with a checkmark and the number '902'. Below the title 'Telco Customer Churn' is the subtitle 'Focused customer retention programs'. The author is 'BlastChar', updated 2 years ago (Version 1). The page includes navigation links for Data, Tasks, Kernels (329), Discussion (8), Activity, and Metadata. There are download options for 'Download (955 KB)' and 'New Notebook'. A vertical ellipsis menu is also present.

Dataset

Telco Customer Churn
Focused customer retention programs

BlastChar • updated 2 years ago (Version 1)

Data Tasks Kernels (329) Discussion (8) Activity Metadata

Download (955 KB) New Notebook

Usability 8.8

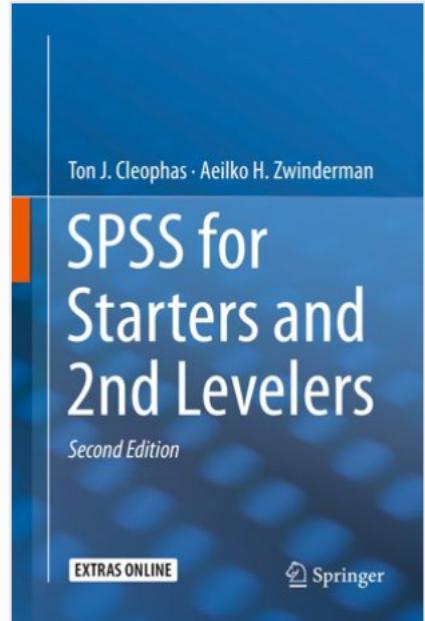
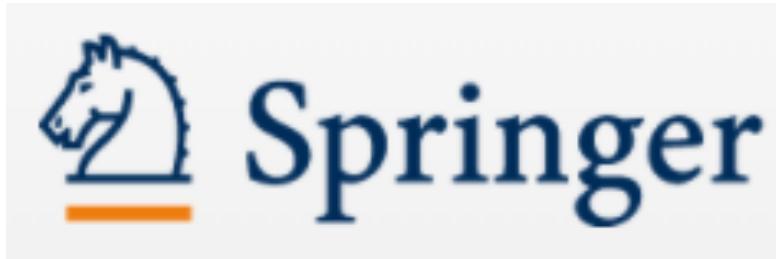
License Data files © Original Authors

Tags business, computing, business services, service providers, telecommunications and 1 more





Surpresa



SPSS for Starters and 2nd Levelers

• Authors

- Ton J. Cleophas
- Aeilko H. Zwinderman

<http://doi.org/10.1007/978-3-319-20600-4>

A Beginners Guide to Python 3 Programming

• Authors

- John Hunt

<http://doi.org/10.1007/978-3-030-20290-3>

