

Practical Data Mining With RapidMiner Studio

Duhai Alshukaili

Ibri College of Technology
Information Technology Department

Oct 23, 2018



الكلية التقنية بعبري
Ibri College of Technology

What RapidMiner?

RapidMiner

RapidMiner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics^a

^aen.wikipedia.org/wiki/RapidMiner

RapidMiner Ecosystem

- RapidMiner Studio
 - Free for non-commercial and academic use.
- RapidMiner Server
- RapidMiner Radoop (RapidMiner version of Hadoop).
- RapidMiner Cloud
- RapidMiner Marketplace
 - Extensions add new functionality to RapidMiner products.

Advantages

- Shallow learning curve.
- Extensive set of operators.
- No coding required.
- Free of cost (the Studio)

Disadvantages

- Deployment of models?!!!
- Commercial use, expensive.
- Pre-processing tasks can be very convoluted.

- The key building blocks are an operator.
- An operator is a functional unit that specified input and output.
- Work-flows are constructed by orchestrating different operators.

Training Videos

<http://bit.ly/2HWe211>

Online Documentation

<http://bit.ly/2HtfwmQ>

Let's dirty our hands




Acknowledgements

The material presented in the tutorial mixes some original material by the author as well as material adopted from

- [Brown, 2017]
- [Grimson, 2016]
- [Tan et al., 2005]

The author gratefully acknowledges the work of the authors cited while assuming complete responsibility for any mistake introduced in the adaptation.

References

-  [Brown, G. \(2017\).](#)
The COMP61011 Not-Very-Scary Guide to Machine Learning.
-  [Grimson, E. \(2016\).](#)
Introduction to machine learning.
[Accessed 22-Mar-2018.](#)
-  [Tan, P., Steinbach, M., and Kumar, V. \(2005\).](#)
Introduction to Data Mining.
[Addison-Wesley.](#)