Visual Programming for Data Science

Doing Complex Things the Right Way

Michael Berthold

Spring Data Talks
March 24, 2021

Data Science Trends

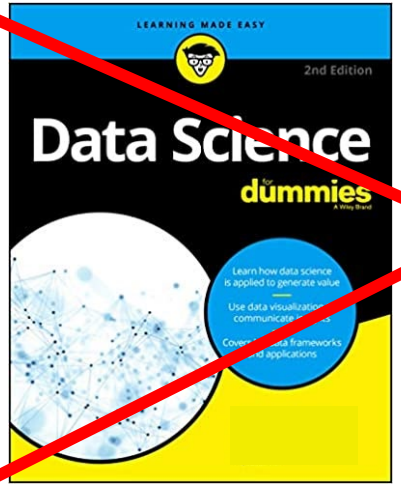
Democratizing Data Science

No-Code / Low-Code

Data Science Trends

Democratizing Data Science

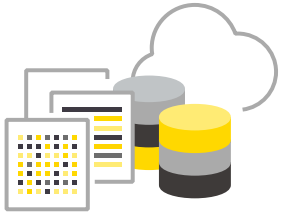
No-Code / Low-Code



"How to become a Data Scientist in One Weekend"

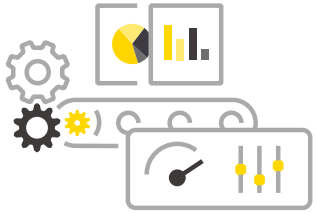
Common Data Science Requirements

Blend & Transform

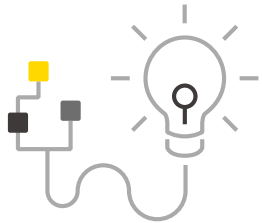


Model & Visualize

Deploy & Manage



Consume & Interact



Techniques:

- Joins
- Pivot
- Normalization
- PCA
- …

- Clustering
- Regression
- Deep Learning
- Visualizations
- …

- Model Monitoring
- Updating
- Validation
- …

- Active Learning
- Interactive Applications
- Services
- …

Technologies and Languages:

- SQL
- Data Lakes (AWS, Azure, Google…)
- …

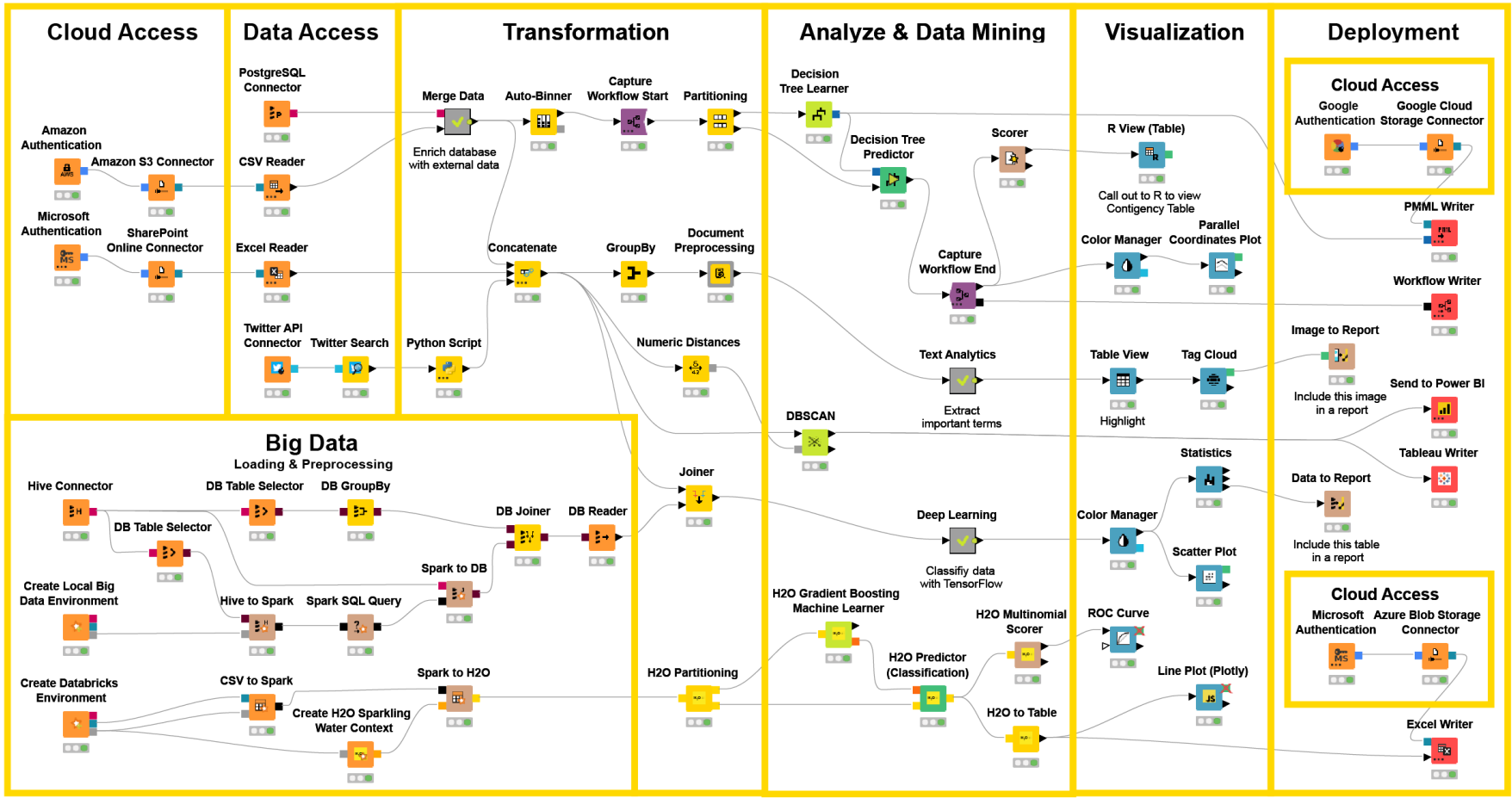
- R, Python
- JavaScript
- …

- Go
- Docker
- Grafana
- …

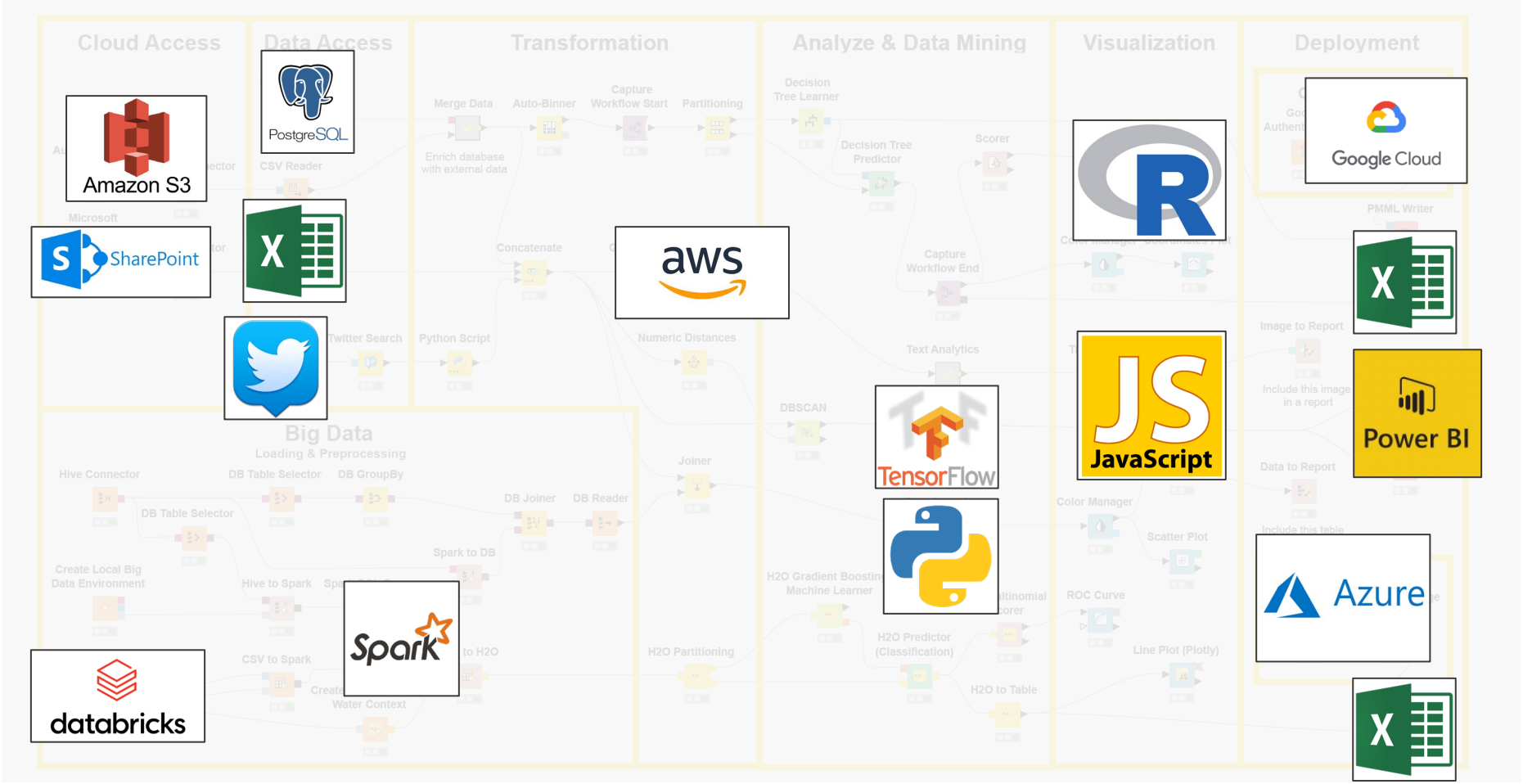
- Plotly, R-Shiny
- JavaScript
- Vue.js
- REST
- …

Data Science Teams shouldn't really have to worry about all of this…

Visual Programming for Data Science



Technologies & Languages under the Hood

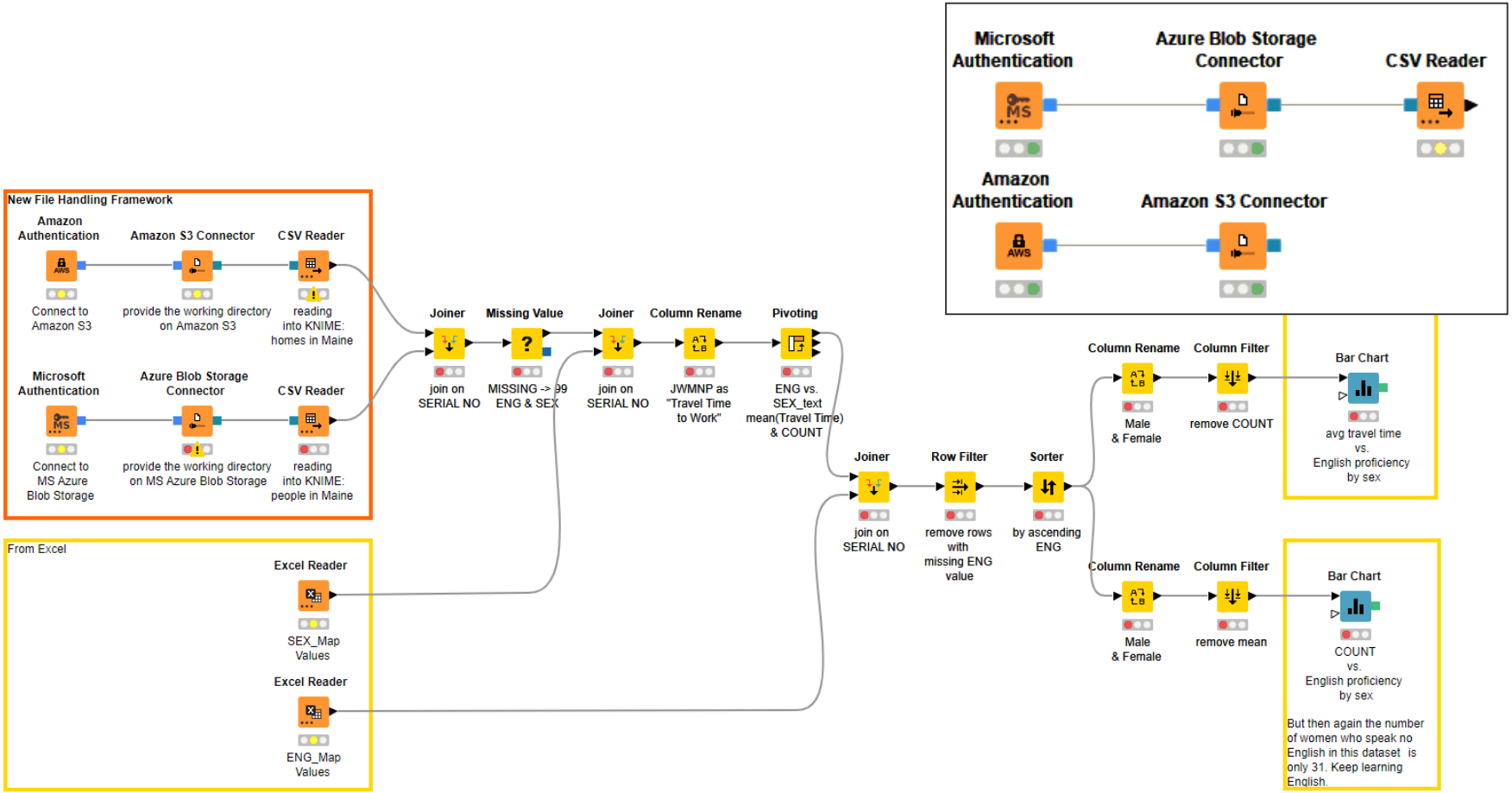


Open for Innovation KNIME

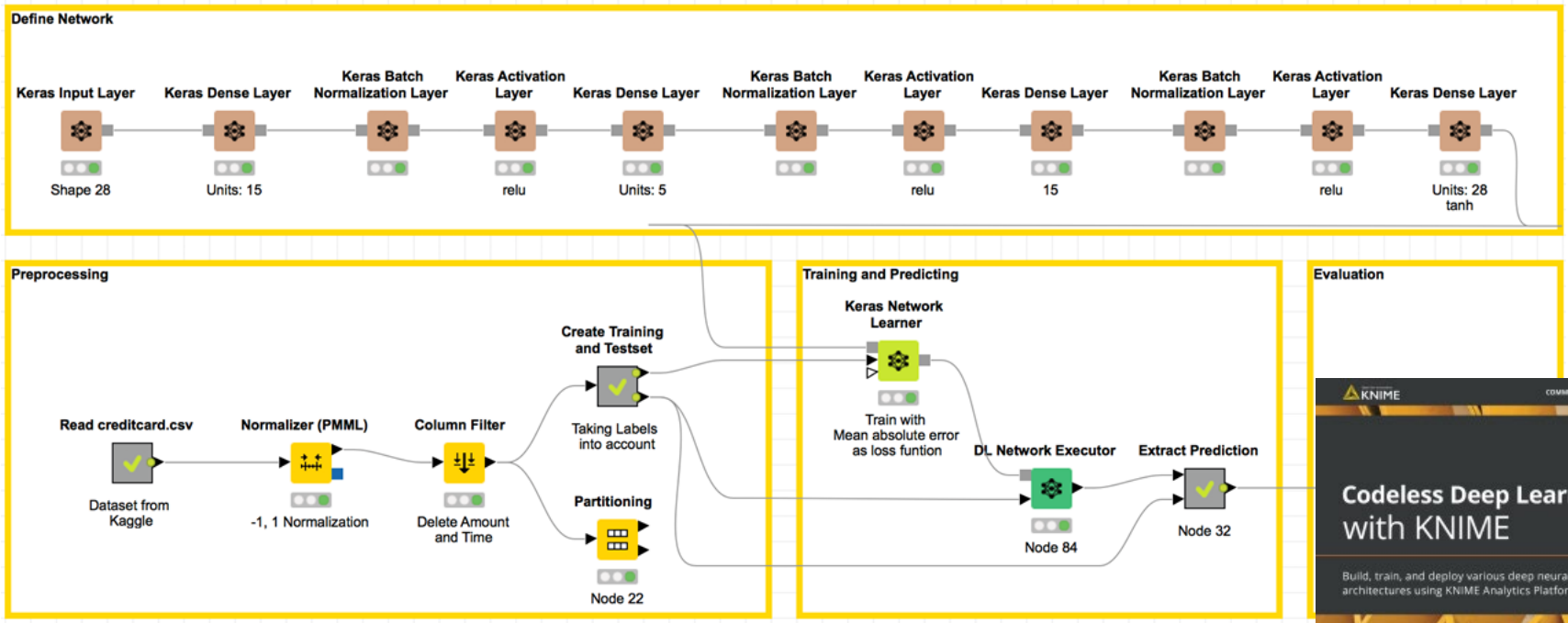
Data Science for Real

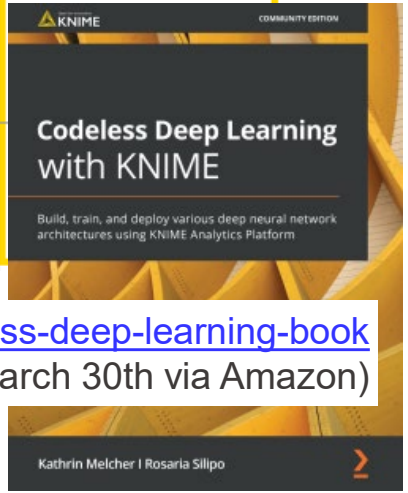
- Data Science *is* Complex (that's why it's called "…Science")
  - It requires a broad spectra of skills – often done in collaboration
  - It requires understanding and access to the inner wheels of an algorithm (but not to the algorithm itself)
  - No single language or tool has all the answers – it's about tool blending as well!

Data Science *is* complex: Data Wrangling



Data Science *is* complex: Deep Learning

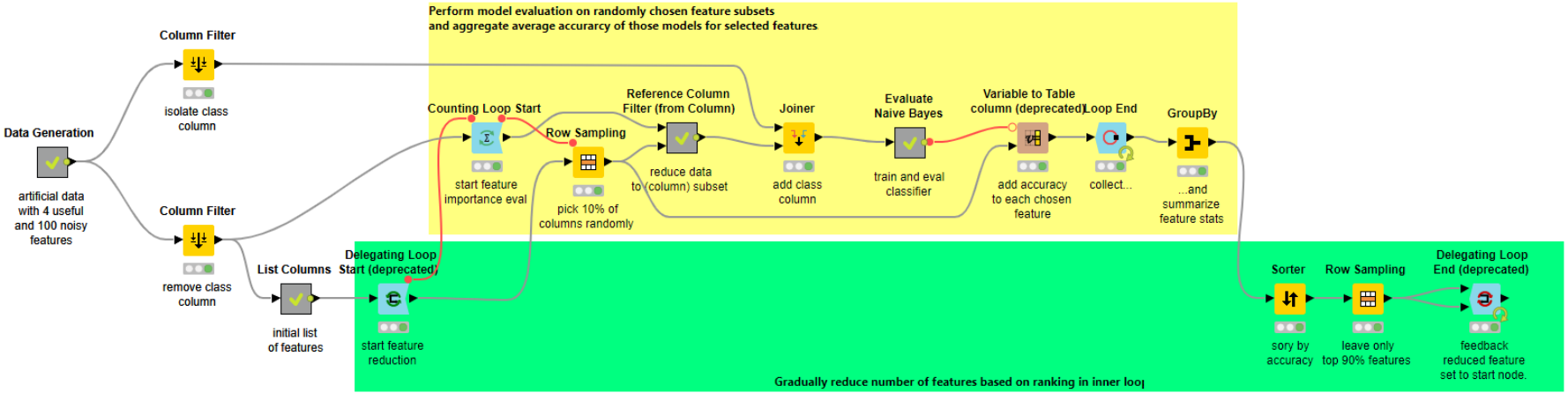




www.knime.com/codeless-deep-learning-book
(40% off until March 30th via Amazon)

Data Science *is* complex: Visual Programming

Visual Data Science Algorithms: Recursive Feature Elimination:



Data Science *is* complex: More Topics

- Data Blending:
  - Feature Selection
  - Feature Engineering
  - Anonymization
  - …
- Modeling
  - Model Selection
  - Parameter Optimization
  - Ensemble Creation
  - Active Learning
  - Transfer Learning
  - …
- Compliance
  - Best Practices
  - Required Standards
  - Bias Detection and Removal
  - Data Privacy / GDPR
  - Explainable “AI”
  - …
- Governance
  - Integration / Dependencies
  - Traceable / Lineage
  - Documented
  - Reproducibility
  - …

Data Science for Real

- Data Science *is* Complex (that's why it's called "…Science")
  - It requires a broad spectra of skills – often done in collaboration
  - It requires understanding and access to the inner wheels of an algorithm (but not to the algorithm itself)
  - No single language or tool has all the answers – it's about tool blending as well!

- Visual Programming (if done right)
  - provides the appropriate level of abstraction to allow focus on modeling a data process
  - allows access to all relevant nuts and bolts
  - removes tool interface complexity
  - enables abstraction / encapsulation to safely reuse
  - provides transparency
  - enables governance

The Data Scientist Top-5

What Should a Visual Programming Environment for Data Science provide?

1. Consistent UI
   (similar look&feel no matter what the underlying technology)
2. One flexible Framework
   (interface the underlying technologies)
3. Uniform access to all relevant Parameters
   (translating different tool APIs to data science configurations)
4. Handle all Dependencies and Versions
   (guarantee (backwards) compatibility)
5. Access to the Underlying Technology when desired
   (inventing, writing, optimizing new algorithms is usually not part of the job)

The Organizational Top-5

What does my Organization get?

1. Collaboration
   (No technology breaks inhibiting experts working together)
2. Reusability
   (Expertise easy to encapsulate & reuse)
3. Governance
   (The environment handles all SW integration and dependency nightmares)
4. Compliance
   (The environment enables best practices and permission handling)
5. Documentation and Meta Information
   (Capture, reproduce, and communicate the flow of data & information)

“Computer programming is an art, because it applies accumulated knowledge to the world, because it requires skill and ingenuity, and especially because it produces objects of beauty.”

Donald Knuth 1974