



# RAPIDMINER OVERVIEW

ML FOR THE MASSES?

STEVEN FOUSKARINIS

NOVEMBER 14, 2020

# DISCLAIMER

- I AM NOT AN EXPERT IN RAPIDMINER. THE NOTES AND VIEWS EXPRESSED IN THIS OVERVIEW ARE VERY HEAVILY INFLUENCED BY MY OWN SUBJECTIVE EXPERIENCE AND USE OF THE PLATFORM.

# AGENDA

- TARGET AUDIENCE
- PLATFORM OVERVIEW
- OVERVIEW OF CAPABILITIES
  - DATA PREP
  - AUTOML
  - ML
  - ML OPS
- DEMO THROUGHOUT
- Q&A

# TARGET AUDIENCE

- SUPPOSEDLY EVERYONE!
  - BUSINESS USERS & STUDENTS
    - NO PROGRAMMING/SOFTWARE ENGINEERING BACKGROUND
    - NO DATA SCIENCE BACKGROUND
  - PRACTITIONERS
    - DATA ENGINEERS
    - DATA SCIENTISTS
- BUT MUST UNDERSTAND SOME ML CONCEPTS, ESPECIALLY ML PIPELINE STRUCTURE, TRAINING VS VALIDATION, AND PERFORMANCE METRICS

# PLATFORM OVERVIEW

- RAPIDMINER STUDIO (LOCAL)
  - DESIGNER
  - TURBO PREP
  - AUTO MODEL
  - DEPLOYMENTS
- AI HUB (REMOTE)
  - RAPIDMINER SERVER
  - GRAFANA (DASHBOARDS)
  - WEB APP DESIGNER
  - JUPYTERHUB NOTEBOOKS (EG USING PYTHON API)
  - REAL TIME SCORING EXTENSION

# OVERVIEW OF CAPABILITIES: DATA PREP

- VISUALIZE DATA (SPARKLINE HISTOGRAM)
- RENAME, COPY, DELETE, SAMPLE & FILTER
- NORMALIZE DATA (REMOVE DUPLICATES, REPLACE MISSING, NORMALIZATION, ETC)
  - REMOVE LOW QUALITY
  - DUMMY ENCODING, DISCRETIZATION
  - PCA
- TRANSFORM DATA TYPE
- GENERATE NEW COLUMNS BASED ON EXISTING COLUMNS
- HEAVILY TABLE ORIENTED (CSV, XLS, DATABASE)
- GENERATES MODEL OF ALL TRANSFORMATIONS THAT CAN BE APPLIED TO ANOTHER DATASET

# OVERVIEW OF CAPABILITIES: AUTOML

- AUTOML TYPES
  - CLASSIFICATION (CLUSTERS)
  - REGRESSION (PREDICT)
  - OUTLIERS
- PROCESS
  - SELECT TARGET (LABEL)
  - WEIGHT LOSS FUNCTION (COST MATRIX)
  - AUTOMATIC SUGGESTION OF COLUMNS TO EXCLUDE (STATISTICAL ANALYSIS)
  - CHOOSE MODEL TYPES

# OVERVIEW OF CAPABILITIES: ML

- VISUAL WORKFLOW EDITING
- COMPONENT OPERATORS
  - DATA LOADING & DATA MANIPULATION
  - MODELS
  - PERFORMANCE MEASUREMENT (VALIDATION)
  - [MODEL SIMULATOR & PRESCRIPTIVE ANALYTICS](#)
- EXTENSIONS
  - NEURAL NETWORKS
  - IMAGE HANDLING
  - PYTHON



# OVERVIEW OF CAPABILITIES: ML OPS

- AI HUB
  - RAPIDMINER SERVER
    - RUN PROCESSES ON SERVER
      - VERTICAL SCALING
      - HORIZONTAL SCALING
    - REMOTE COLLABORATION ON SERVER MODELS/PROCESSES
      - PERMISSIONING
      - VERSIONING
      - PROGRAMATIC CONTROL (REST API & PYTHON API)
    - CREATE & PUBLISH WEB SERVICES
    - REAL TIME SCORING
  - DEVELOP DASHBOARDS (GRAFANA)
  - DEVELOP WEB APP (WEB APP DESIGNER)
  - CREATE JUPYTER NOTEBOOKS (JUPYTERHUB) (EG USING PYTHON API)

# DEMO

- DATA PREP
- AUTOML
- ML
- ML OPS

# IN MY HUMBLE OPINION

- GOOD FOR QUICK INVESTIGATION OF POSSIBLE ML APPLICATION
  - TABULAR, SQL, TEXT DATA
  - CLASSIFICATION
  - REGRESSION (LINEAR & NON-LINEAR)
  - TIME SERIES
- GOOD FOR NO/LOW CODE
- GOOD FOR GRAPHICAL INTERACTIVE MODEL
- GOOD FOR VISUALIZING (MODEL SIMULATOR & DASHBOARDS)
- GEARED TOWARDS STRUCTURED DATA (TABULAR)
- JUST ADDED SUPPORT FOR IMAGE LOADING/CONVERSION IN OCTOBER
- VERY LIMITED NEURAL NETWORK MODELLING
- NOT CONVENIENT FOR PROGRAMMING (ALTHOUGH EXTENSIBLE)

QUESTIONS  
?



# DESCRIPTION

- RAPIDMINER IS A PLATFORM THAT IS USED BY OVER 4,000 UNIVERSITIES TO TEACH MACHINE LEARNING CONCEPTS TO STUDENTS WHILE NOT PRESUMING THOSE STUDENTS CAN PROGRAM. HOWEVER THE PLATFORM IS ALSO USED BY OVER 40,000 ORGANIZATIONS WORLDWIDE FOR COMMERCIAL APPLICATIONS.
- COME AND TAKE A BRIEF BUT INFORMATIVE GUIDED TOUR OF THE RAPIDMINER COAL MINES FROM A USER NOT A SALESPERSON.