

Taking Weka to the next level with ADAMS

No need to write code

Outline

- What is ADAMS?
- Weka Investigator
- Flow editor
- Other useful tools
- Demo

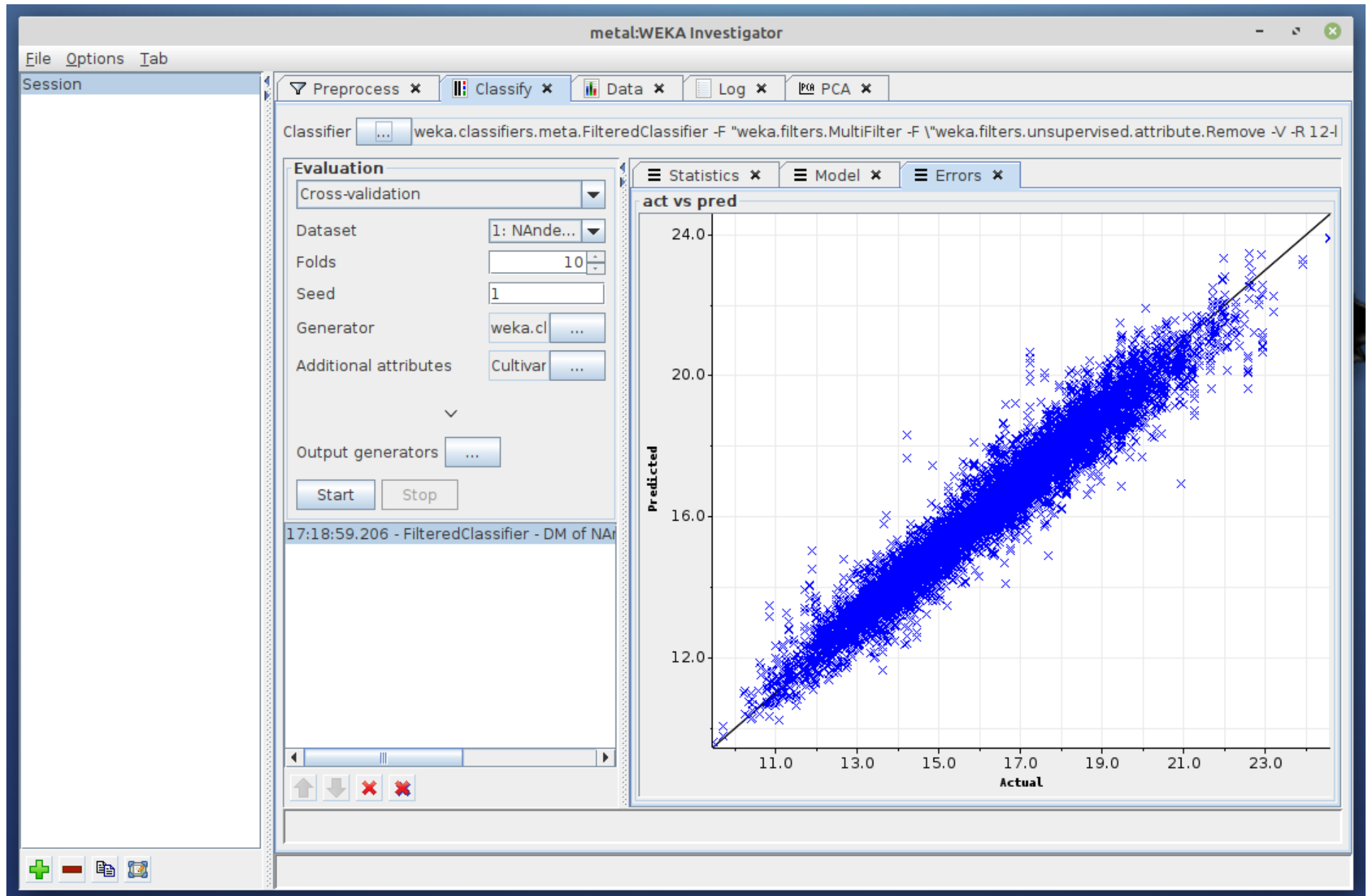
What is ADAMS?

- Java plugin framework (Linux, OSX, Windows)
- Modules with various functionality
 - eg modules for Weka, Meka, MOA
- Interfaces for various tasks
- Workflow engine
- Started in 2009 to work with GC-MS data

Weka Investigator

- Exploration tool like Weka Explorer
- But it supports
 - multiple sessions
 - multiple datasets loaded at the same time
 - batch filtering
 - arbitrary number of tabs
 - predefined output generators
 - different/additional visualizations
 - easy exports of results

Weka Investigator



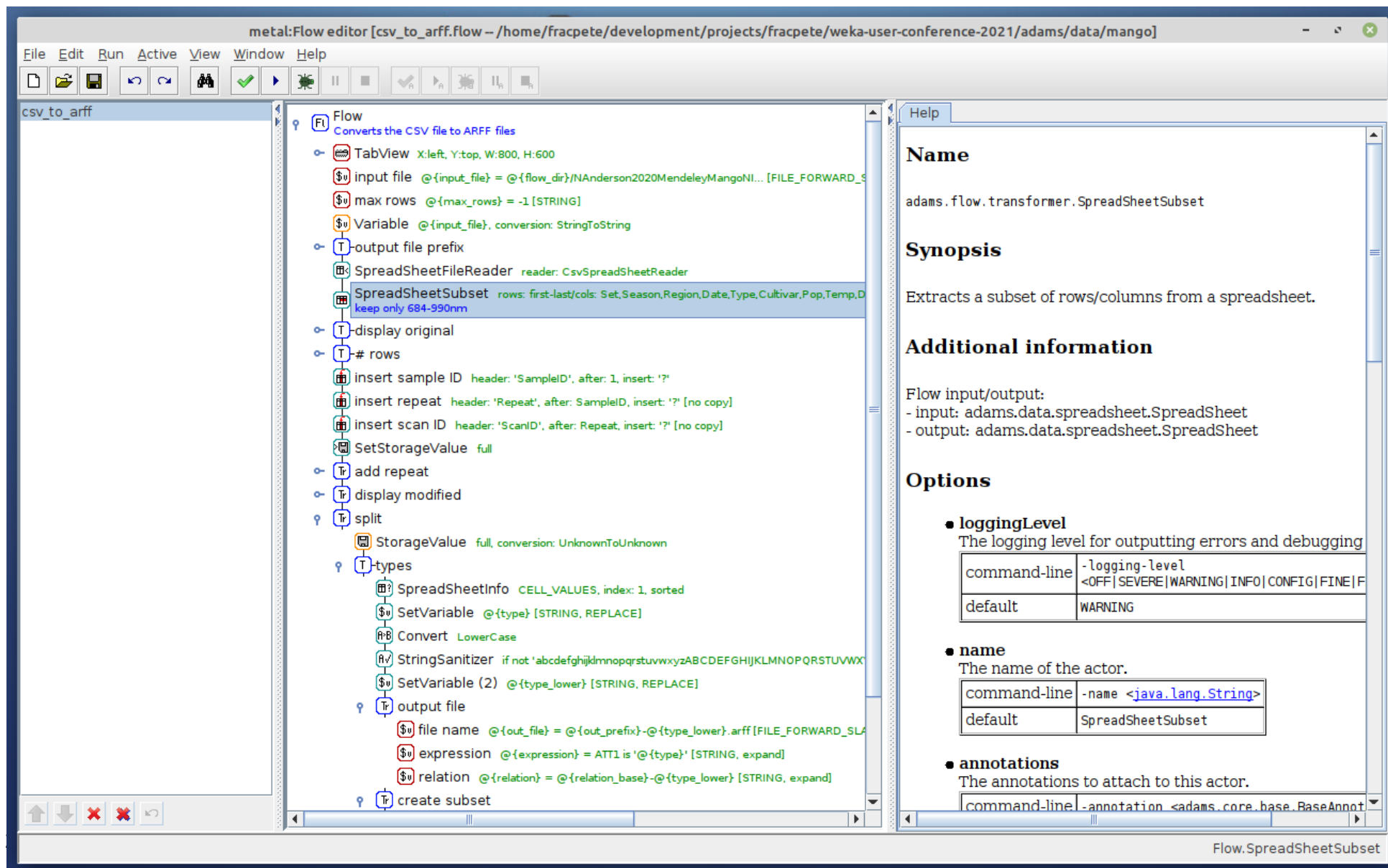
Flow editor

- Why another workflow system?
 - Canvas-based set up is tedious
 - Doesn't scale well (when using 100s or 1000s of operators)
 - KnowledgeFlow too Weka-centric
- How is it different/better?
 - Actors (= operators) arranged in tree (collapsible)
 - Actor handlers nest other actors (eg sequence)
 - Control actors control data flow (eg tee, branch, switch)
 - Input/output defines:
 - standalone (no I/O), source (only O), transformer (I/O), sink (only I)
 - Offers debugging (incl. breakpoints and data inspection)

Flow editor

- How does it work?
 - Data-driven system (actor output triggers next one)
 - Event-based triggers available (eg cron, web-services)
- But: Tree only supports 1-to-n connections
- Simulating n-to-m semantics
 - Containers (combine multiple outputs)
 - Variables (change actor options on-the-fly)
 - Internal storage (key-value storage)
 - Callable actors (feed data into named actor)

Flow editor



The screenshot displays the Flow editor interface. The main workspace shows a flow diagram for a process titled "Converts the CSV file to ARFF files". The flow includes several actors: TabView, input file, max rows, Variable, output file prefix, SpreadSheetFileReader, SpreadSheetSubset (highlighted), display original, # rows, insert sample ID, insert repeat, insert scan ID, SetStorageValue, add repeat, display modified, split, StorageValue, types, SpreadSheetInfo, SetVariable, Convert, StringSanitizer, SetVariable (2), output file, and create subset. The right-hand panel shows the help documentation for the "SpreadSheetSubset" actor.

Name
adams.flow.transformer.SpreadSheetSubset

Synopsis
Extracts a subset of rows/columns from a spreadsheet.

Additional information
Flow input/output:
- input: adams.data.spreadsheet.SpreadSheet
- output: adams.data.spreadsheet.SpreadSheet

Options

- **loggingLevel**
The logging level for outputting errors and debugging

command-line	-logging-level
default	<OFF SEVERE WARNING INFO CONFIG FINE F

- **name**
The name of the actor.

command-line	-name <java.lang.String>
default	SpreadSheetSubset

- **annotations**
The annotations to attach to this actor.

command-line	-annotation <adams.core.base.BaseAnnot
--------------	--

Flow.SpreadSheetSubset

Real world applications

- Annotate deep learning datasets
 - object detection, image/instance segmentation
- S3000
 - commercial tool based on ADAMS to analyze soil/plant samples (NIR, MIR, XRF)
 - User only manages configuration flows, generators create low-level worker flows
 - Eurofins Agro: ~450 models, ~3000 samples/day
 - Largest (production) worker flow: ~44,000 actors

Other useful tools

- Multi-Experimenter
- Spreadsheet file viewer
- SQL Workbench
- File commander (ftp, smb, sftp)
- Preview browser
- Append/merge datasets
- Check dataset compatibility

Available functionality

- Databases (MySQL, SQLite, PostgreSQL, HSQL, MSSQL, Sybase, MS Access)
- XML/XSLT/XPath, HTML, JSON
- LaTeX (generate/execute), PDF (read/extract/generate)
- spreadsheets (CSV, TSV, Gnumeric, ODF, MS Excel, fixed column, ...)
- webservices (SOAP/REST)
- scripting (Groovy/Jython/Python)
- Weka, Meka, MOA, R-Project, parameter optimization
- Natural language processing (parsing, word clouds)
- OCR, barcodes
- spectral data (AniML, CAL, CML, DPT, EEM, JCampDX, MPS, NIR, Opus, Relab, SPA, SPC, SpecLib, spreadsheet-based)
- scatter/line plots, gnuplot, control charts
- images, heatmaps, audio, video, webcams, ffmpeg
- de-/compression (tar, zip, bzip2, gzip, lzma, xz, zstd)
- remote access (sftp, ssh, scp, rsync)
- Java code generation
- [Generate custom ADAMS application](#)
- [Docker images and image generation](#)
- frontend (user interaction), backend (Linux daemon/Windows Service)

Demo

Enough talking, let's see ADAMS in action!

