Towards Plug-and-Play Machine Learning in Astronomy

[1]J. Riedel, [2]M. Taylor, [1]K. Polsterer
johanna.riedel@h-its.org m.b.taylor@bristol.ac.uk kai.polsterer@h-its.org
[1]HITS gGmbH, [2]University of Bristol

November 1–5, 2026 ADASS 2026, Perth, Australia

Abstract

As in many other disciplines, in astronomy Machine Learning (ML) is becoming more and more important. Current workflows often force researchers to repeatedly extract and preprocess data in specific formats for using them in their individual ML tools. We present a concept that extends the International Virtual Observatory Alliance (IVOA) vision of interoperability to ML models. Our goal is to enable an easy integration of ML models into existing astronomical tools. By using standardized representations such as ONNX, models can easily be tested on diverse datasets without significant additional effort. This standardization makes models interchangeable, allows fair comparisons, accelerates iterative development, and simplifies the extension of tools with individual AI functionalities. Finally, we demonstrate a prototype that integrates various custom-built ML models into TOPCAT, showcasing practical applications such as automatic classification or redshift estimation.

Mission Objective — in progress

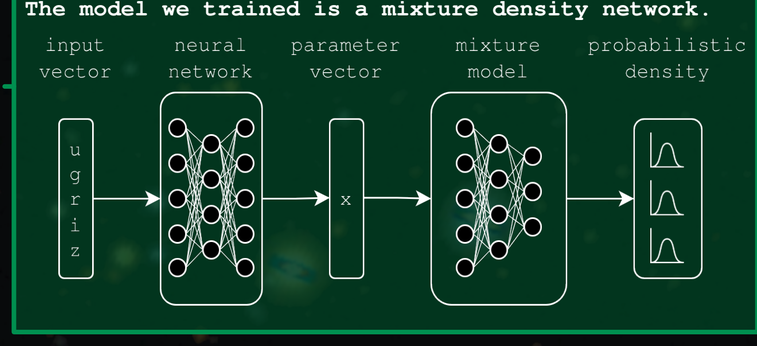
Our vision is to establish a culture of publishing models in a standardized representation alongside papers and having tools that are able to incorporate these models. This would lead to the following advantages:

1. *Interchangeability of Models*: Models can be changed based on the users' needs. Your model is better than a standard model? – Use and publish your own!
2. *Fair comparisons*: Models from different developers can be compared in the exact same environment. No need for re-implementing models from others to reproduce results. No need in knowing the exact parameters a model was trained with.
3. *Accelerated iterative development:* Being able to use your model with a bunch of different datasets makes it easier to figure out the weaknesses of your model.
4. *Simplified extension of tools*: The tool you prefer lacks a functionality and it would take years for the developers to implement it? – Create your own model for that and plug it in! You can even use your favorite ML framework.

Redshift Model — completed

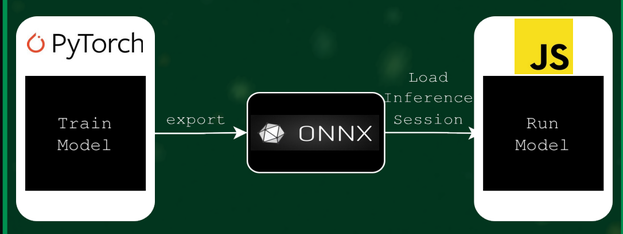
For prototyping and testing with Topcat, we trained a model to predict redshifts as probabilistic distributions. For an object we therefore not only know the estimated redshift, but also the uncertainty. This allows us to select instances based on uncertainties.

The model we trained is a mixture density network.



ONNX — completed

Open Neural Network Exchange (ONNX) provides an open-source framework with the goal to standardize the representation of machine learning algorithms and software tools.
It defines common building blocks and a common file format allowing machine learning models to be portable. You can use your preferred framework to train your model, export it and then import and use it in whatever framework/programming language you want.
With ONNX it is possible to use different frameworks, tools, runtimes and compilers. ONNX is developed by the Linux Foundation.



Metadata — in progress

A standardized representation needs standardized information about how to use a model. To export a model with ONNX we have to define the key names to access all input and output parameters. We also need to provide a dummy input in the correct shape for ONNX to know what it should work with.

```
# Export model
export_model.eval()
torch.onnx.export(
    export_model,
    dummy_input,
    "redshift_model_pdf.onnx",
    input_names = ['input'],
    output_names = ['pi', 'sigma', 'mean'],
    dynamic_axes = {
        'input': {0: 'batch_size'},
        'pi': {0: 'batch_size'},
        'sigma': {0: 'batch_size'},
        'mean': {0: 'batch_size'}
    },
    external_data = False
)
```



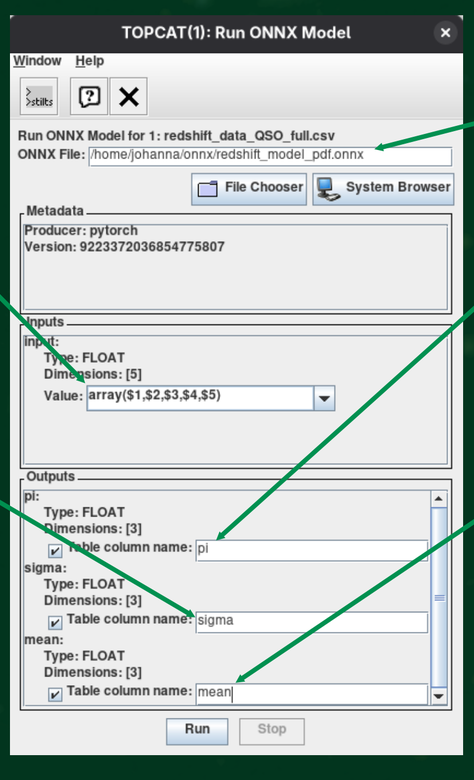
Registry — start now

How can we find models that are already published and ready for integration in a tool? We suggest creating an online registry to have a catalogue of models. A model entry should describe the metadata, provenance, its purpose, the storage location, and e.g. the tools that already successfully used the model.

IVOA standards — start now

Models take a fixed input format and produce a fixed output format. This needs standardization. Therefore, we propose to think about extending/developing IVOA standards for standardized usage of models. This could for example include standard input and output formats for models solving specific tasks.

Example: Topcat — in progress



Choose the Path to your model

Configure how Topcat can feed inputs into your model

Choose the column name for your first output dimension

Choose the column name for your second output dimension

Choose the column name for your third output dimension

Try yourself:

https://www.star.bristol.ac.uk/mbt/releases/topcat/pre/topcat-full_onnx.jar

Other tools? — start now

ASCL.net Astrophysics Source Code Library · SIMBAD · VizieR · ALADIN · CANFAR

Player Profile

Johanna Riedel (she/her)
johanna.riedel@h-its.org
Master's Student
Astroinformatics research group
Heidelberg Institute for Theoretical Studies

