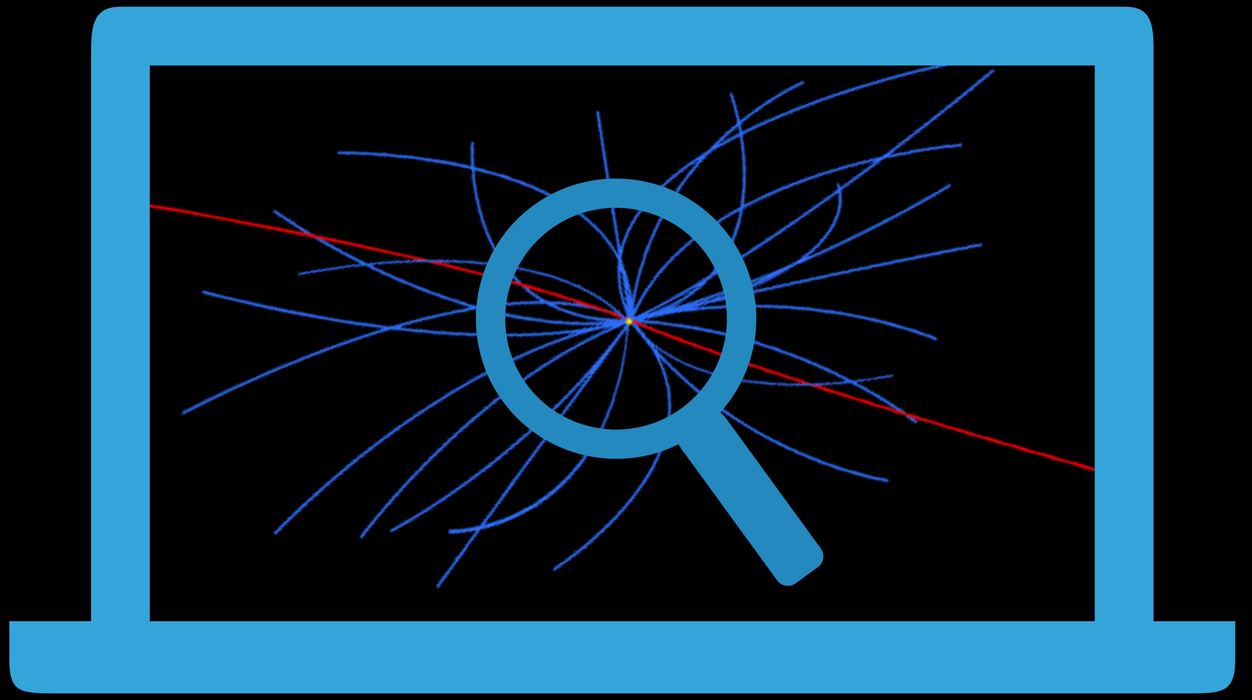


Teaching Machine Learning
with ATLAS Open Data

Kate Shaw, Tom Stevenson, Caley Yardley

University of Sussex

IOP Open Data Workshop for Higher Education

14 May 2026

Acknowledgements

Andrey Kukhmay (Sussex MSc Data Science)

Iago Rossetto (Sussex Intern, Physics)

The ATLAS Collaboration, and CERN Open Data Portal

Overview

Overview

UNIVERSITY OF SUSSEX

Open Data Goals

Open science principles underpin all Research activity to ensure it upholds the **UN Sustainable Development Goals** with respect to equality of education and growth.

CERN Open Science

Experiments at CERN have an obligation to make data **public** and **accessible**.

> Open research data refers to the publishing of the data underpinning scientific research results so that they have no restrictions on their access and usage.
>
> Openly sharing data opens it up to inspection and re-use, forms the basis for research verification and reproducibility, and opens up a path to broader collaboration.

UNESCO Open research data definition

Overview

Open Data Goals

Open science principles underpin all Research activity to ensure it upholds the **UN Sustainable Development Goals** with respect to equality of education and growth.

CERN Open Science

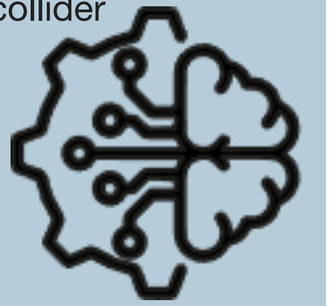
Experiments at CERN have an obligation to make data **public** and **accessible**.

> Open research data refers to the publishing of the data underpinning scientific research results so that they have no restrictions on their access and usage.
>
> Openly sharing data opens it up to inspection and re-use, forms the basis for research verification and reproducibility, and opens up a path to broader collaboration.

UNESCO Open research data definition

Machine Learning

Machine learning in ATLAS is used to search for rare particle events amid large backgrounds, leveraging many variables to detect subtle signals in high-dimensional collider datasets, enabling precise analyses and searches for new physics.



Education

Accessible, **multi-level training resources** developed to guide students from no prior experience to advanced machine learning, enabling **hands-on exploration** of ATLAS Open Data, skill development, and preparation for physics and data science careers.

Overview

Open Data Goals

Open science principles underpin all Research activity to ensure it upholds the **UN Sustainable Development Goals** with respect to equality of education and growth.

CERN Open Science

Experiments at CERN have an obligation to make data **public** and **accessible**.

> Open research data refers to the publishing of the data underpinning scientific research results so that they have no restrictions on their access and usage.
>
> Openly sharing data opens it up to inspection and re-use, forms the basis for research verification and reproducibility, and opens up a path to broader collaboration.
>
> *UNESCO Open research data definition*

Machine Learning

Machine learning in ATLAS is used to search for rare particle events amid large backgrounds, leveraging many variables to detect subtle signals in high-dimensional collider datasets, enabling precise analyses and searches for new physics.

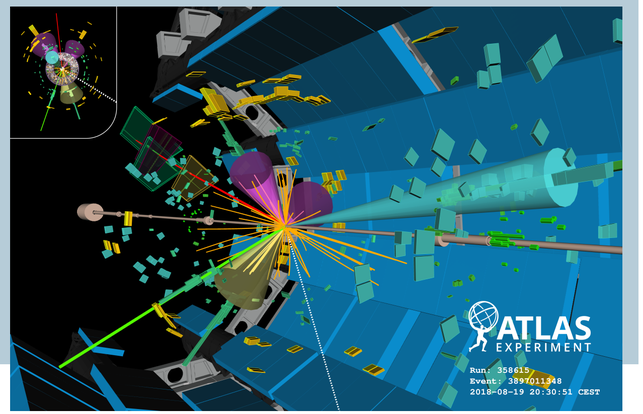
Education

Accessible, **multi-level training resources** developed to guide students from no prior experience to advanced machine learning, enabling **hands-on exploration** of ATLAS Open Data, skill development, and preparation for physics and data science careers.

ATLAS @ CERN

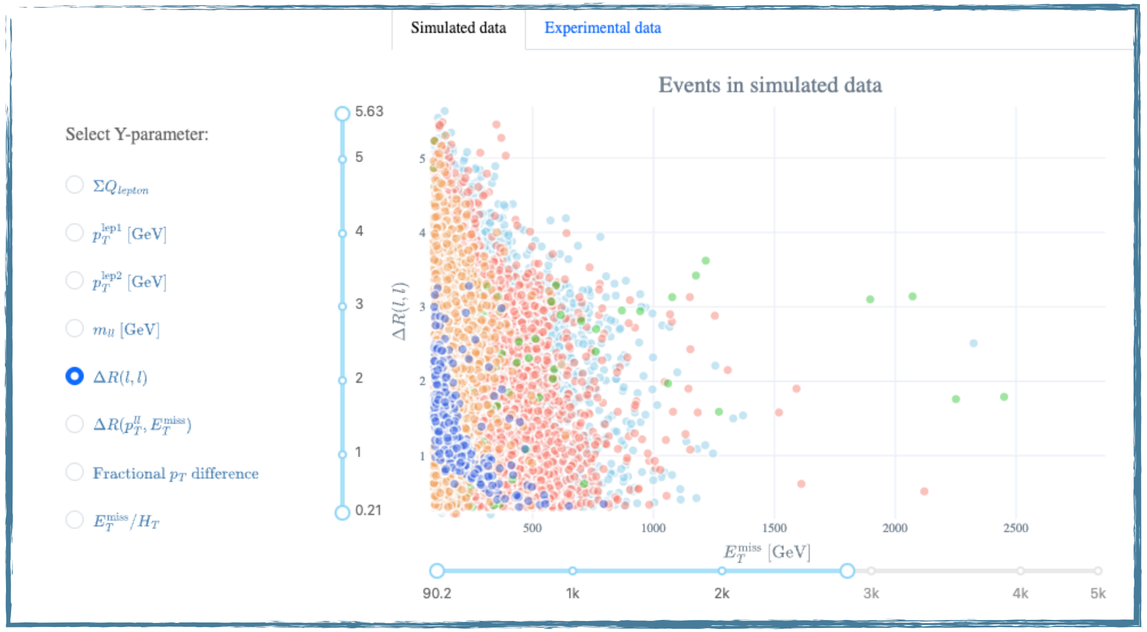
The ATLAS experiment at CERN's Large Hadron Collider is a particle detector exploring the building blocks of matter & the forces shaping our universe.

By studying proton-proton collisions at unprecedented energies, it probes the Higgs boson, rare processes, and potential new physics, aiming to unlock the Universe's deepest mysteries.

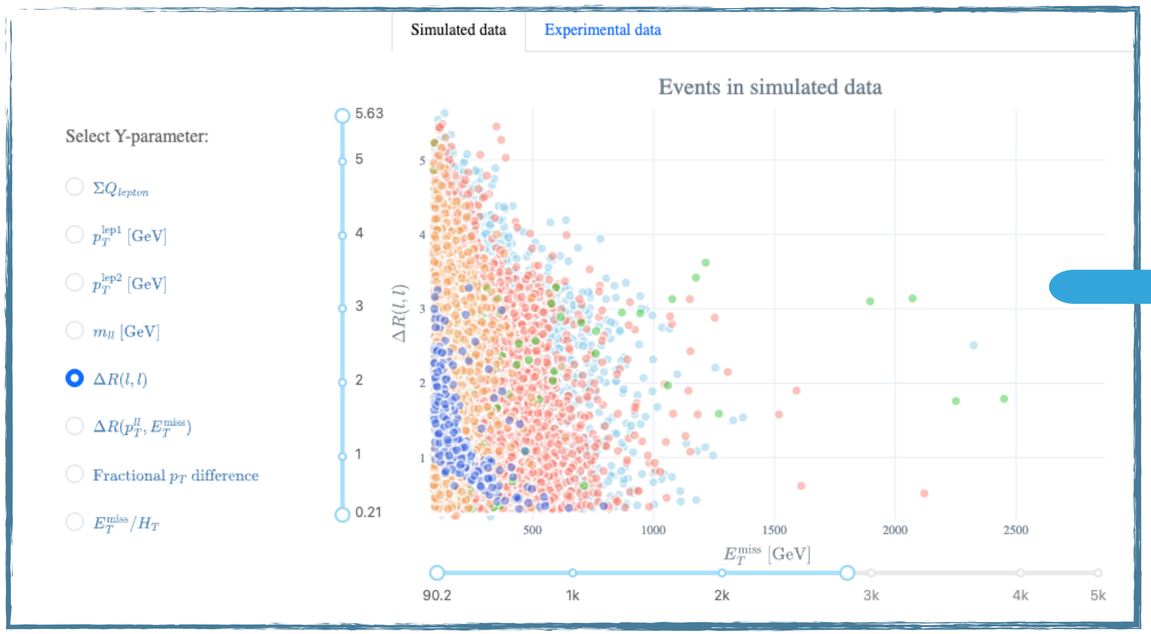


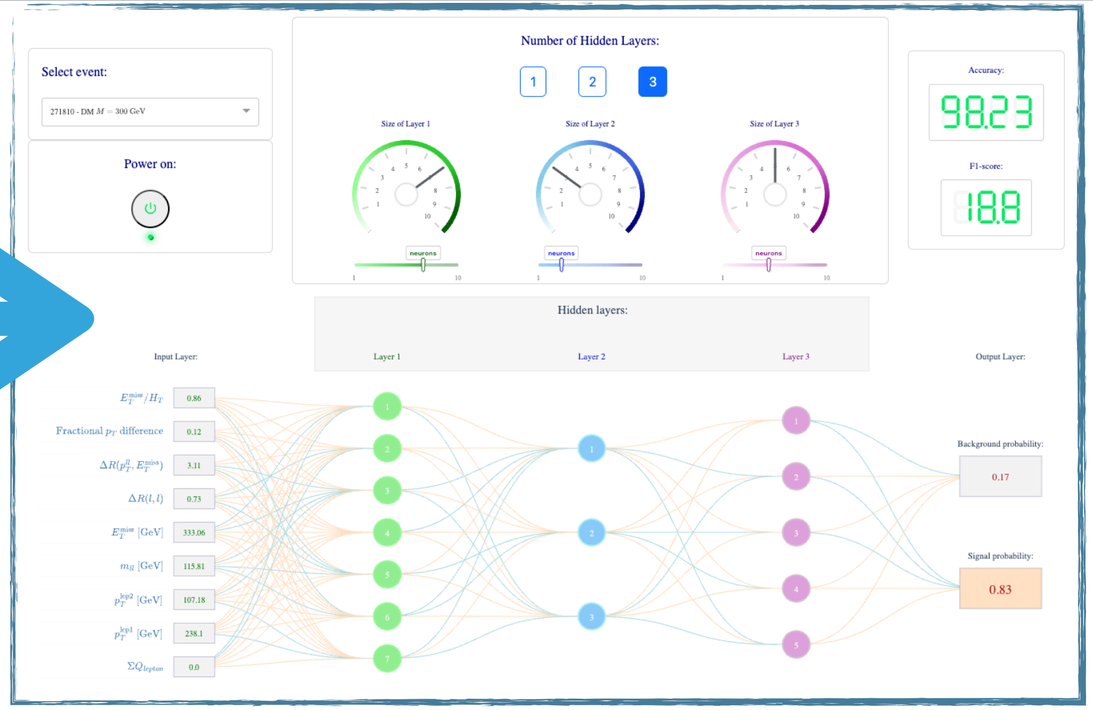
Training Resources and Learning Pathways

Web-Based Interactive Application for ML

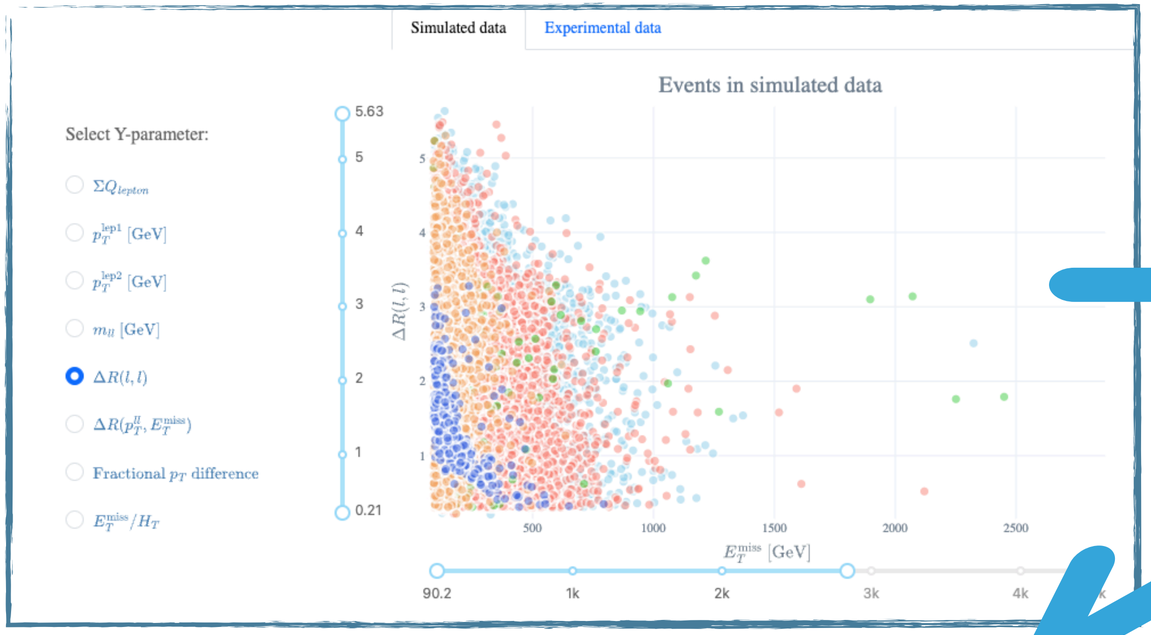


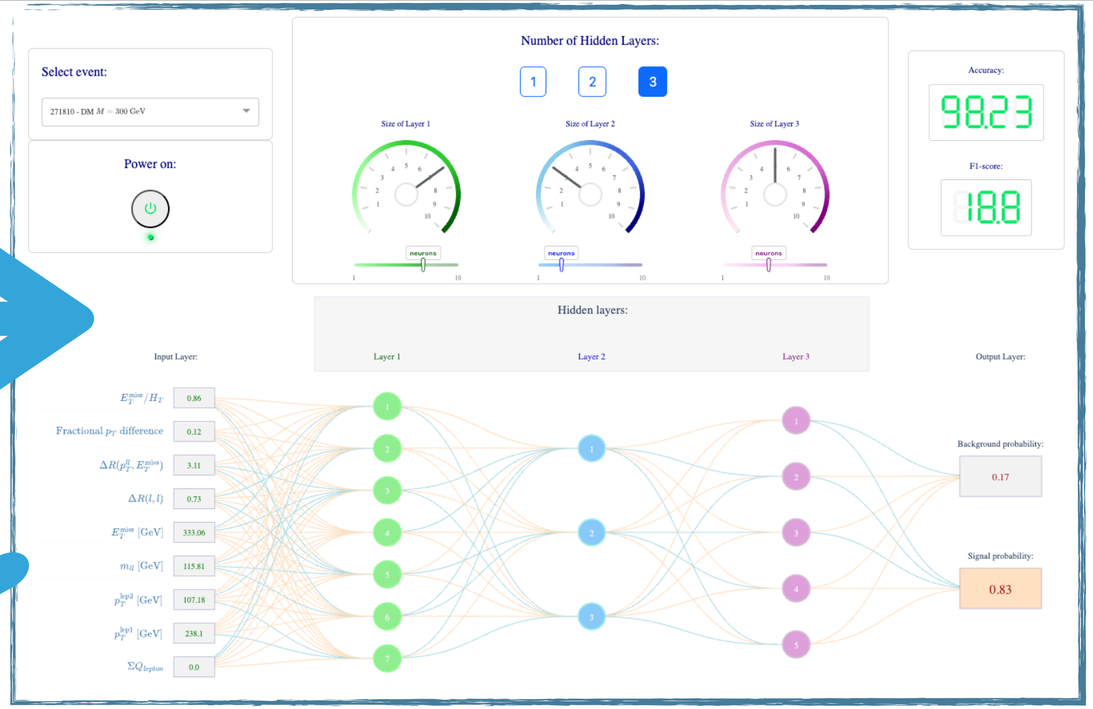
Web-Based Interactive Application for ML

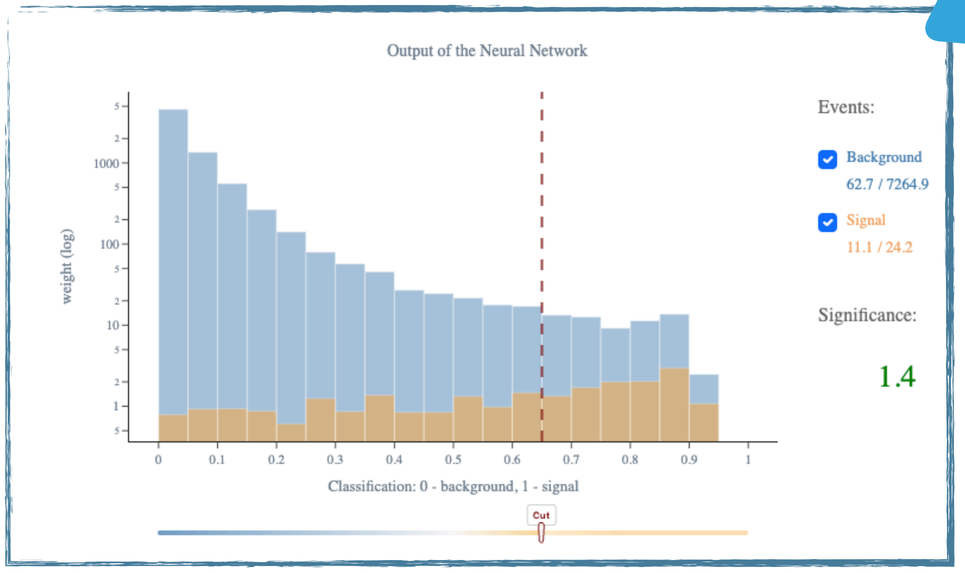




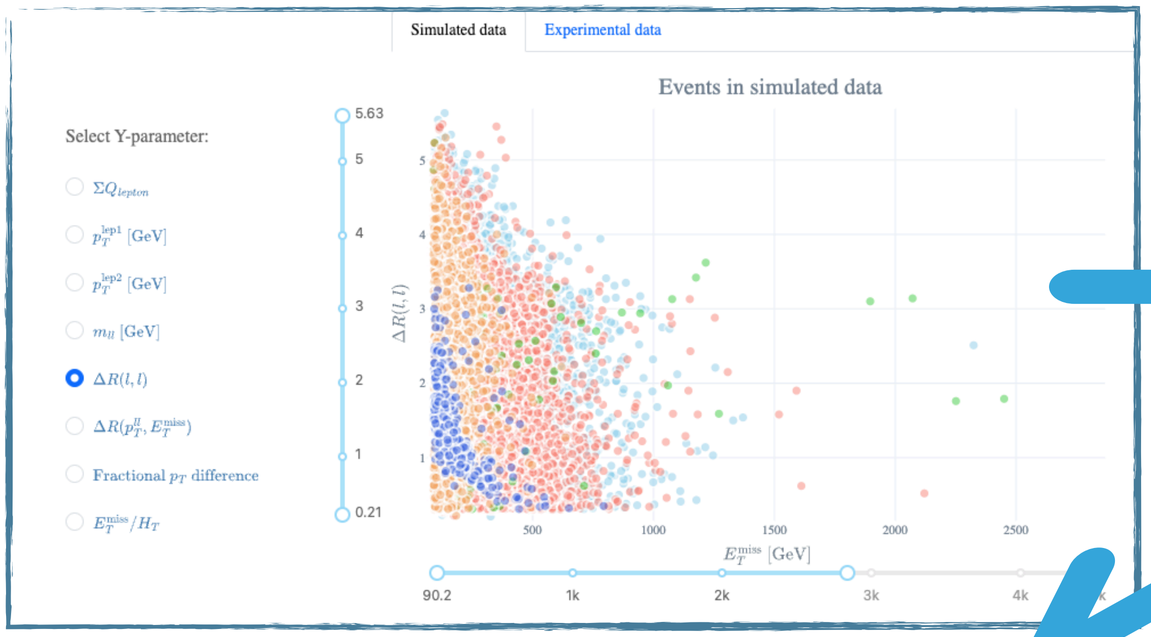
Web-Based Interactive Application for ML

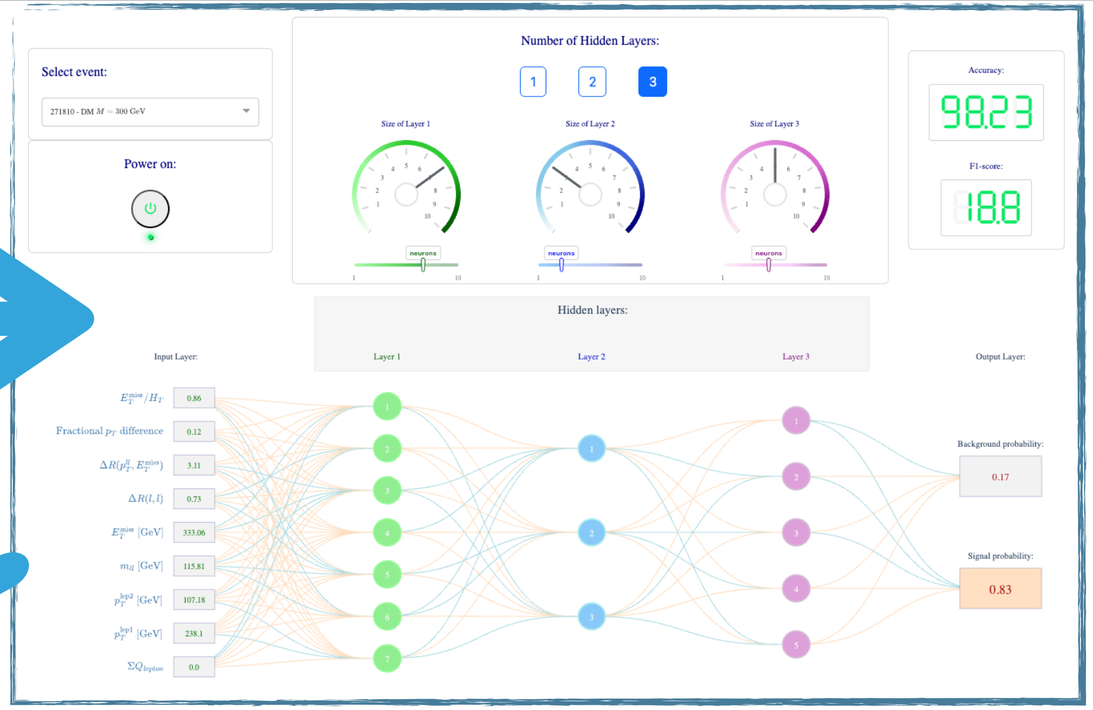


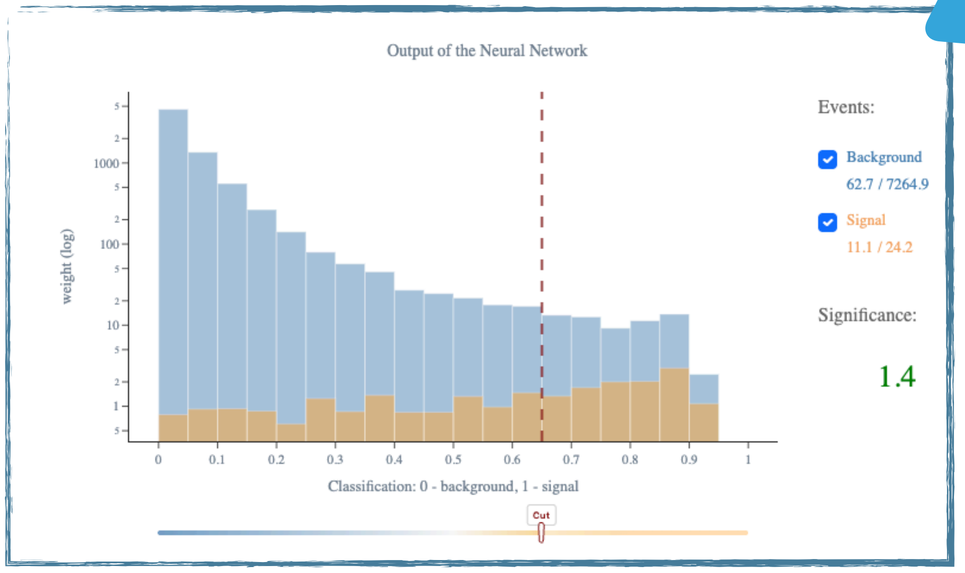




Web-Based Interactive Application for ML







Web-based interactive app provides guided introduction to ML and data analysis, accessible to students with no prior experience.

Demonstrates how ML separates rare signal events from large backgrounds in ATLAS. Users **explore variables** and **adjust neural network hyper-parameters** via point-and-click controls and evaluate performance using metrics and output distributions.

All activities usable **without coding**, offering a practical, engaging, and **research-focused** experience for students and all publics that introduces students to core **ML concepts** and their applications in experimental physics.

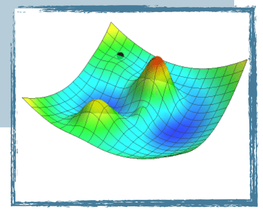
Undergraduate Jupyter Notebooks for ML

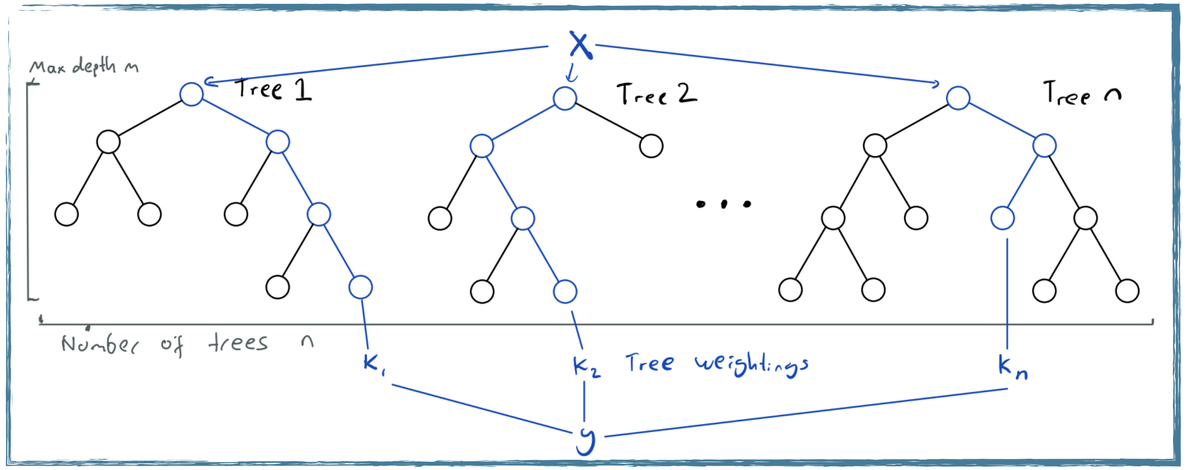
UNIVERSITY OF SUSSEX

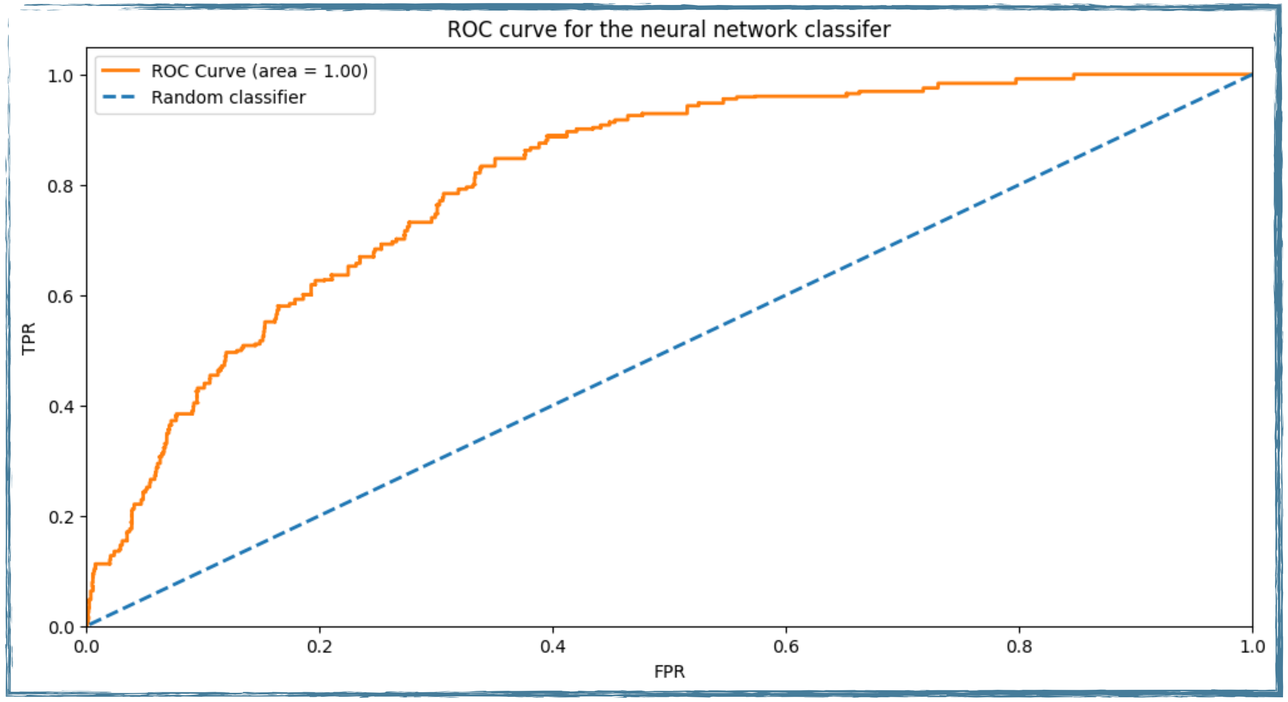
Jupyter Notebook introduces machine learning fundamentals using Python's **scikit-learn** to create classifiers.

Students actively explore **ROC curves**, **gradient descent**, **random forests**, and **neural networks**, with real ATLAS physics applications.

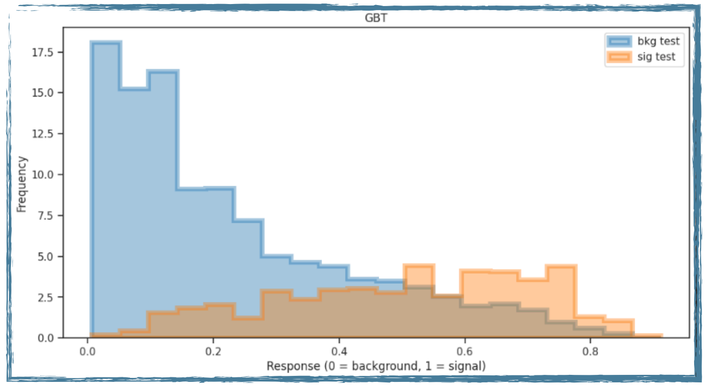
Also learning material about **more advanced libraries** such as PyTorch, TensorFlow, and Keras.

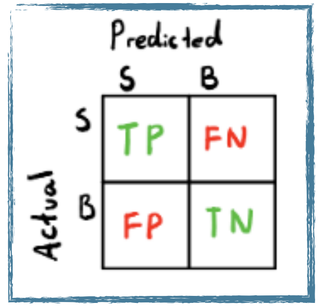







```
In [ ]: from keras import Sequential
        from keras.layers import Input,Dense
        from keras.utils import plot_model
        NN_model = Sequential()
        NN_model.add(Input(shape=(len(ML_inputs),)))
        NN_model.add(Dense(20, activation='relu'))
        NN_model.add(Dense(20, activation='relu'))
        NN_model.add(Dense(1, activation='sigmoid'))

        opt = keras.optimizers.Adam(learning_rate=0.005)

        NN_model.compile(loss='binary_crossentropy', optimizer=opt, metrics=['accuracy'])
```





Advanced / Early Career Jupyter Notebooks

UNIVERSITY OF SUSSEX

Machine Learning Workbook

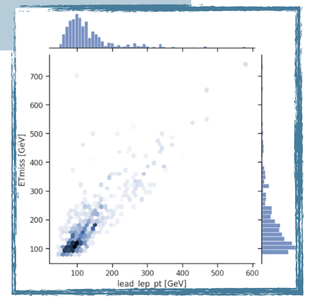
ACCESS WORKBOOK HERE

This online workbook, created by Thomas Stevenson, will introduce the basics of Machine Learning and how it can be exploited in an experimental physics context. Experience with python is required, however no knowledge of ML is necessary. Warning: it may take a long time to load the workbook and associated repositories! A preview is available by clicking on the image.

Jupyter Notebook, accessible via Binder, trains students in applied machine learning **using hands-on exercises**, including k-Fold Cross-Validation and ROC curve analysis.

Demystifies common ML terminology, making concepts approachable for Python users with no prior experience.

```
In [ …  from sklearn.model_selection import ShuffleSplit
        ABDT = AdaBoostClassifier(tree.DecisionTreeClassifier(max_depth=6),
                                  algorithm="SAMME",
                                  n_estimators=20)

        cv = ShuffleSplit(n_splits=2, test_size=0.2, random_state=0)
        errorVsSize(ABDT,cv,X_train,y_train,4)
```



Advanced / Early Career Jupyter Notebooks

Machine Learning Workbook

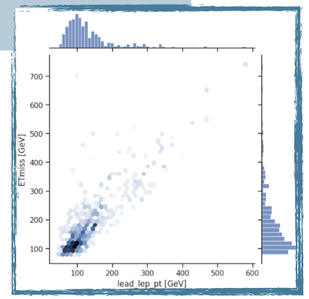
ACCESS WORKBOOK HERE

This online workbook, created by Thomas Stevenson, will introduce the basics of Machine Learning and how it can be exploited in an experimental physics context. Experience with python is required, however no knowledge of ML is necessary. Warning: it may take a long time to load the workbook and associated repositories! A preview is available by clicking on the image.

Jupyter Notebook, accessible via Binder, trains students in applied machine learning **using hands-on exercises**, including k-Fold Cross-Validation and ROC curve analysis.

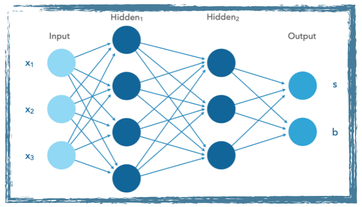
Demystifies common ML terminology, making concepts approachable for Python users with no prior experience.

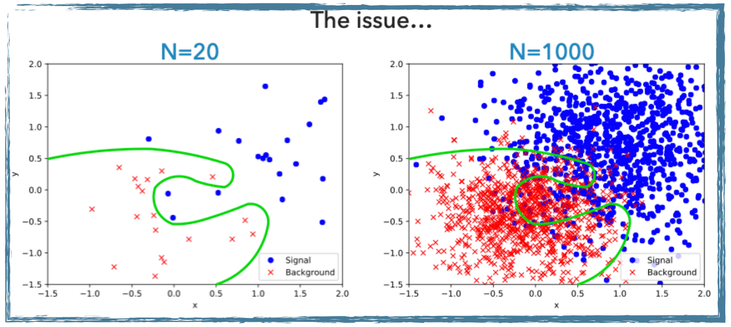
```
In [ … from sklearn.model_selection import ShuffleSplit
ABDT = AdaBoostClassifier(tree.DecisionTreeClassifier(max_depth=6),
                          algorithm="SAMME",
                          n_estimators=20)

cv = ShuffleSplit(n_splits=2, test_size=0.2, random_state=0)
errorVsSize(ABDT,cv,X_train,y_train,4)
```

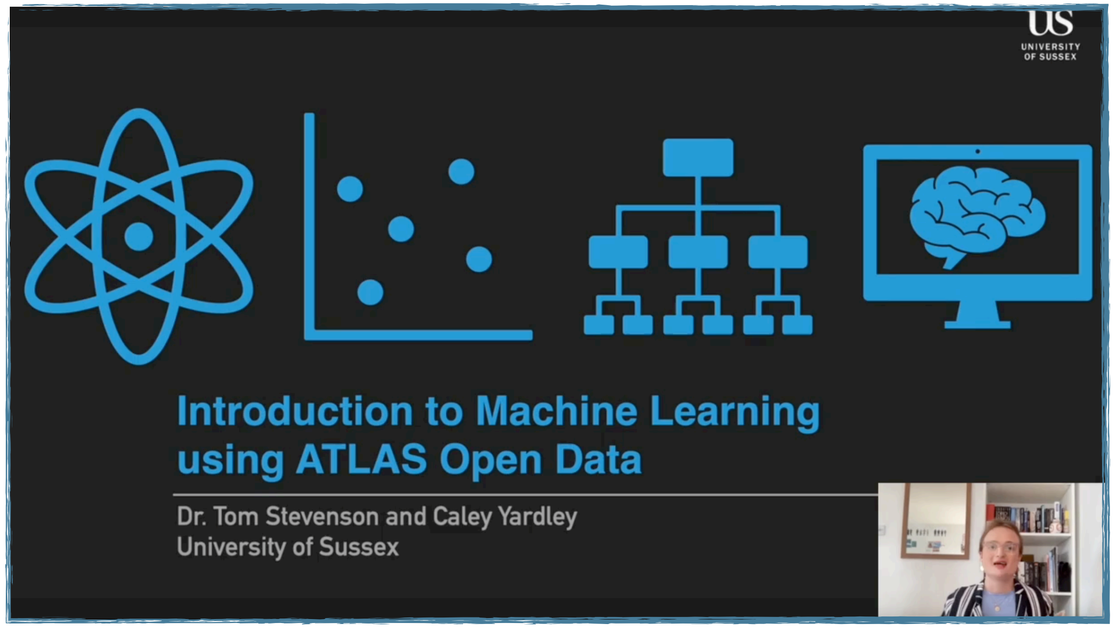


An accompanying presentation, available as downloadable slides and an explanatory video, provides additional guidance.

Together, they allow students to **explore concepts** at their own pace, gain practical skills, and **build research-focused expertise** applying machine learning to high-dimensional physics datasets.

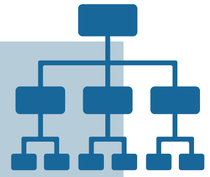






Outlook

Expand

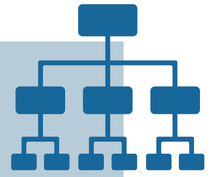


Expand resources to include advanced ML techniques and more complex ATLAS analyses

Outlook

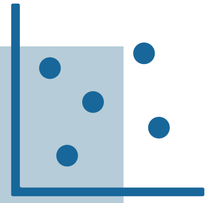
Expand

Expand resources to include advanced ML techniques and more complex ATLAS analyses



Integrate

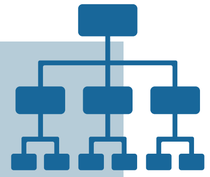
Integrate additional LHC datasets to provide broader, real-world data experiences



Outlook

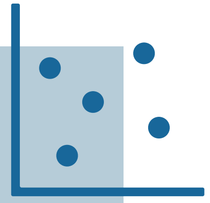
Expand

Expand resources to include advanced ML techniques and more complex ATLAS analyses



Integrate

Integrate additional LHC datasets to provide broader, real-world data experiences



Train

Train the next generation of physicists and data scientists in practical, hands-on high-dimensional data analysis skills

Thanks!

- **References and links:**
  - **ATLAS Collaboration (2020). ATLAS 13 TeV samples collection exactly two leptons (electron or muon), for 2020 Open Data release. CERN Open Data Portal. DOI:** *10.7483/OPENDATA.ATLAS.OHZ7.RFNH*
  - *https://opendata.atlas.cern/docs/category/web-apps-for-education*
  - *https://github.com/atlas-outreach-data-tools/ml-visual-dashboard*
  - *https://github.com/atlas-outreach-data-tools/Histogram_Analyser_Dark_Matter*