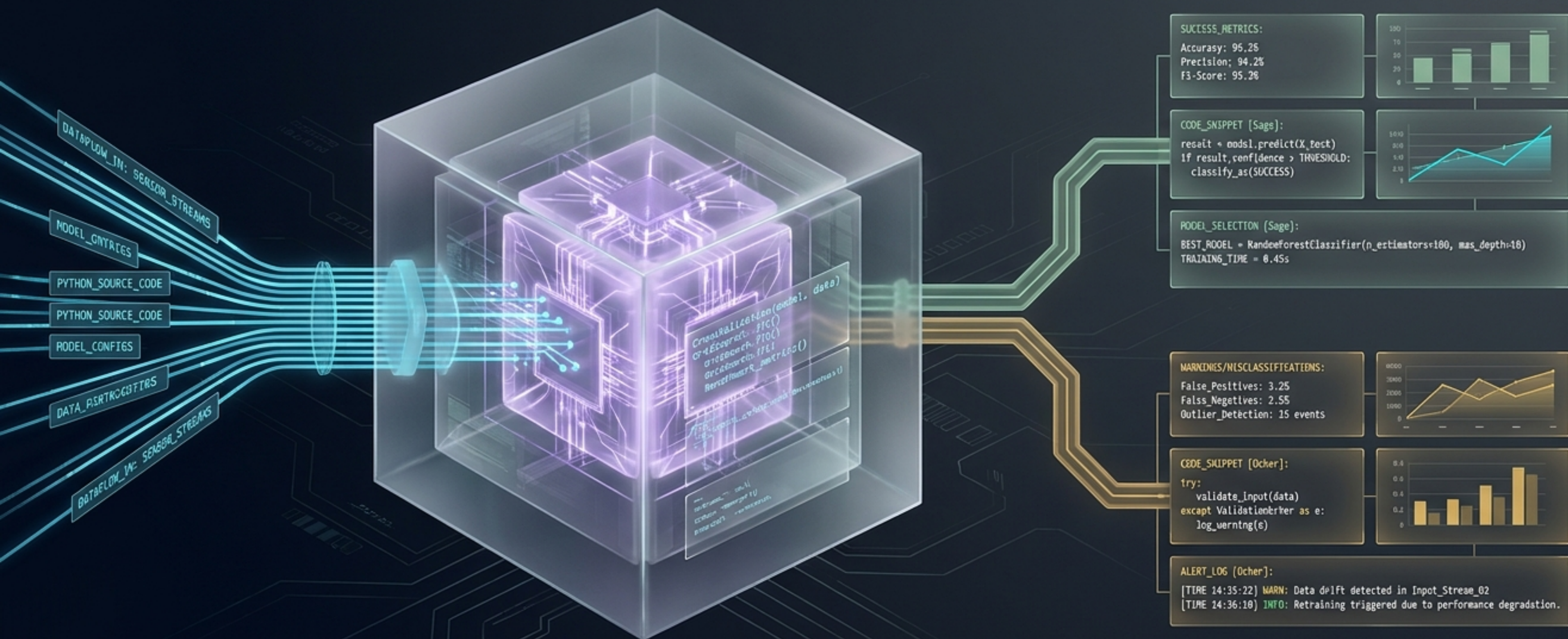


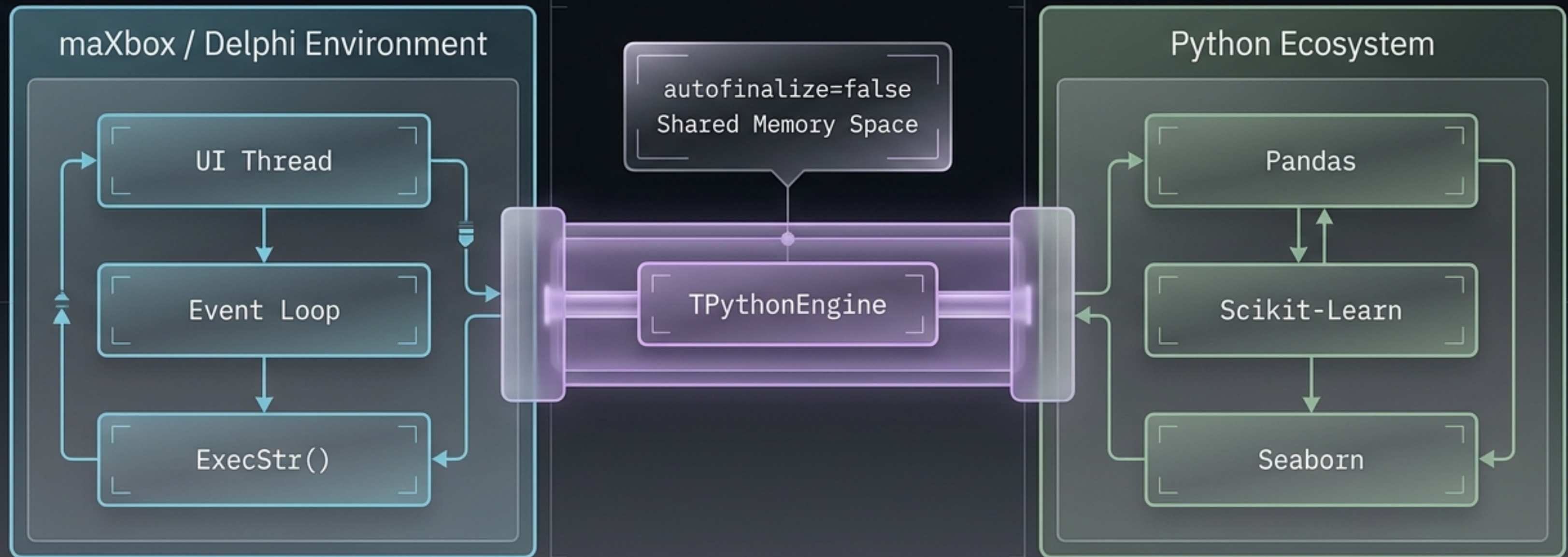
The Embedded ML Pipeline

Model Selection, Cross-Validation, and Benchmarking in maXbox via Python4Delphi

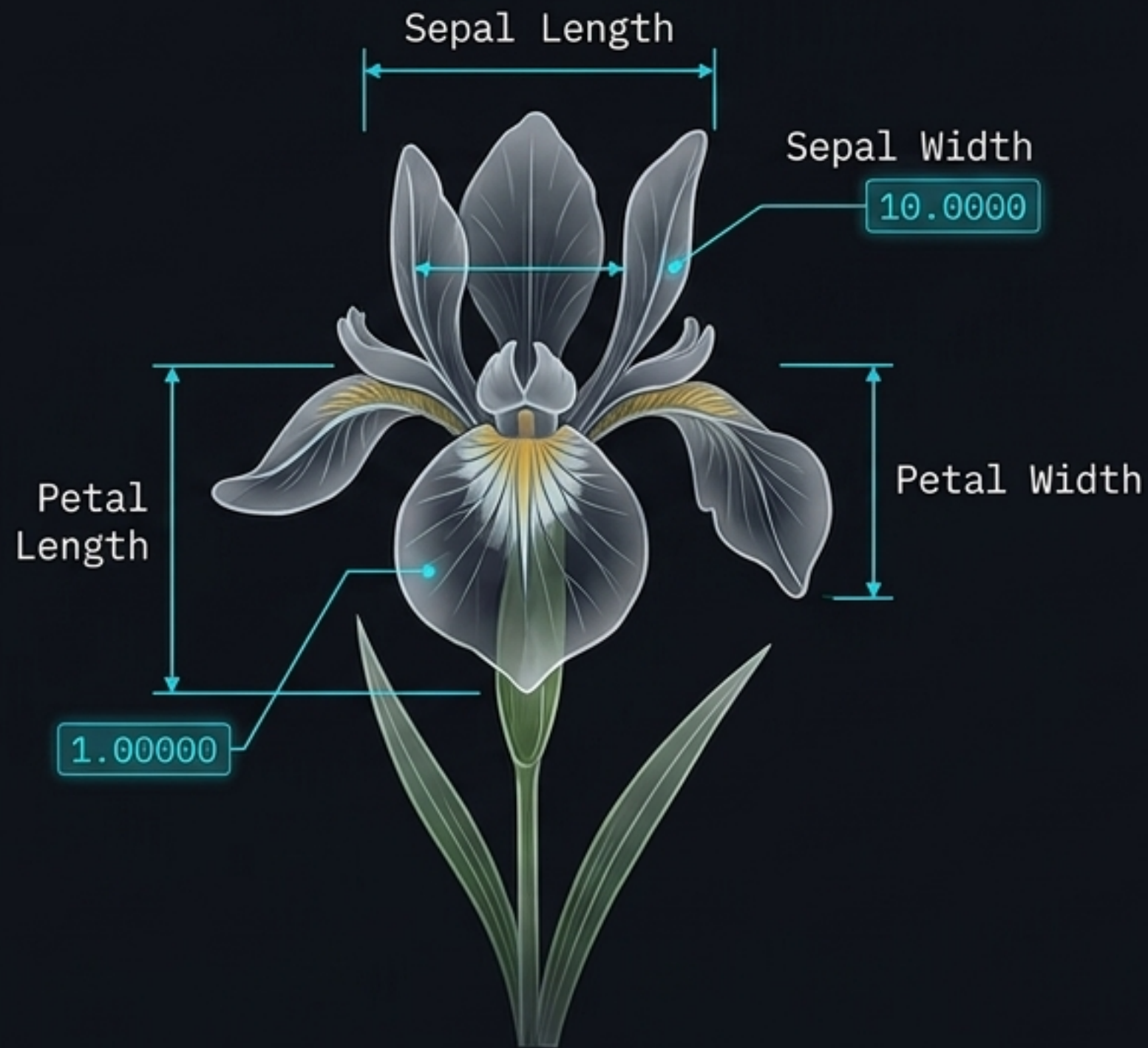


Two Ecosystems. One Continuous Pipeline.

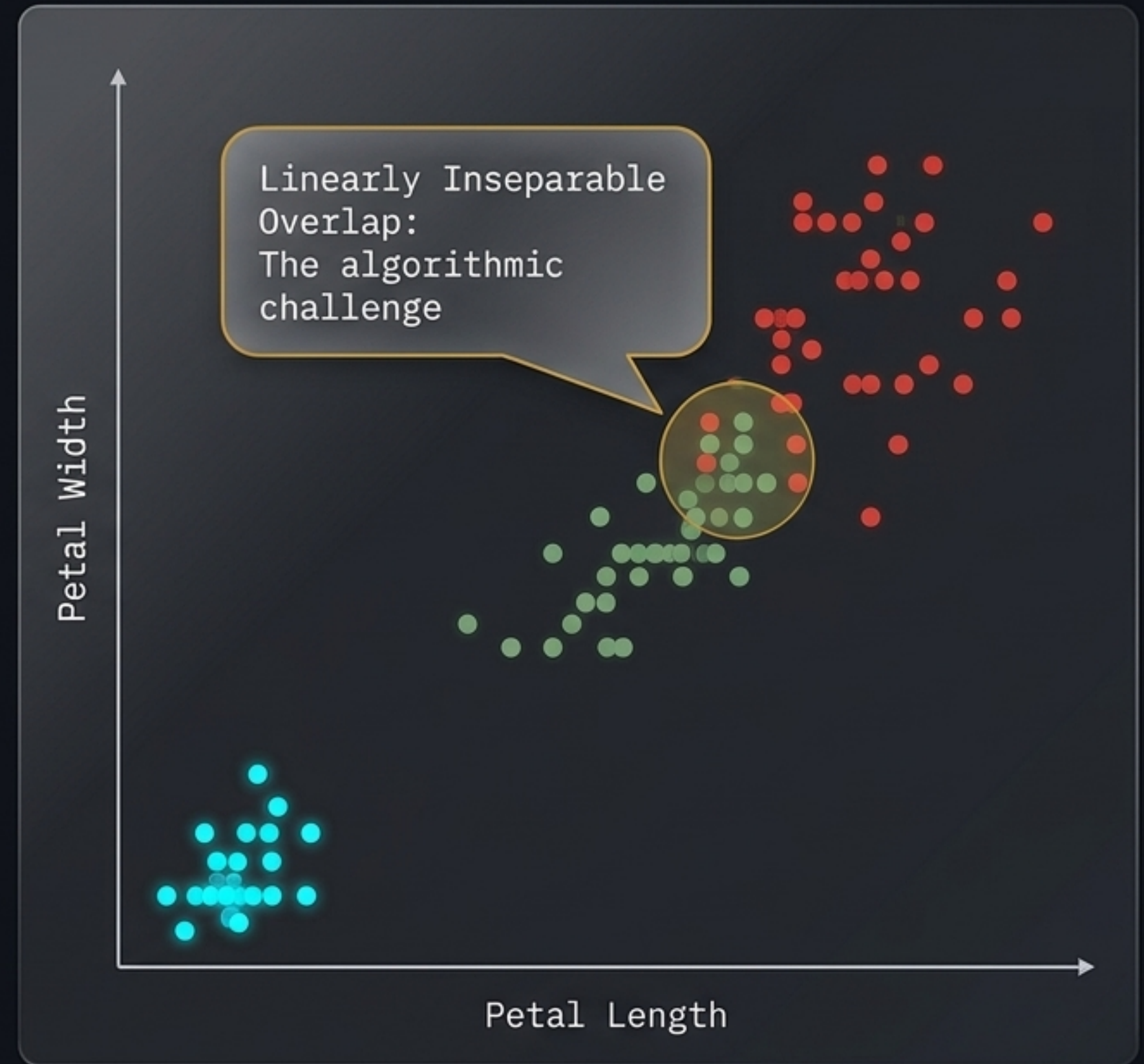
Bypassing traditional API network overhead by embedding Python directly into Delphi desktop applications. This native integration enables real-time execution of complex data science workflows entirely within a compiled GUI.



The Proving Ground: Iris Dataset



The dataset originated from the Iris Flower Dataset, which was originally composed of Sepal Width. In this version, the features and their values are classified.



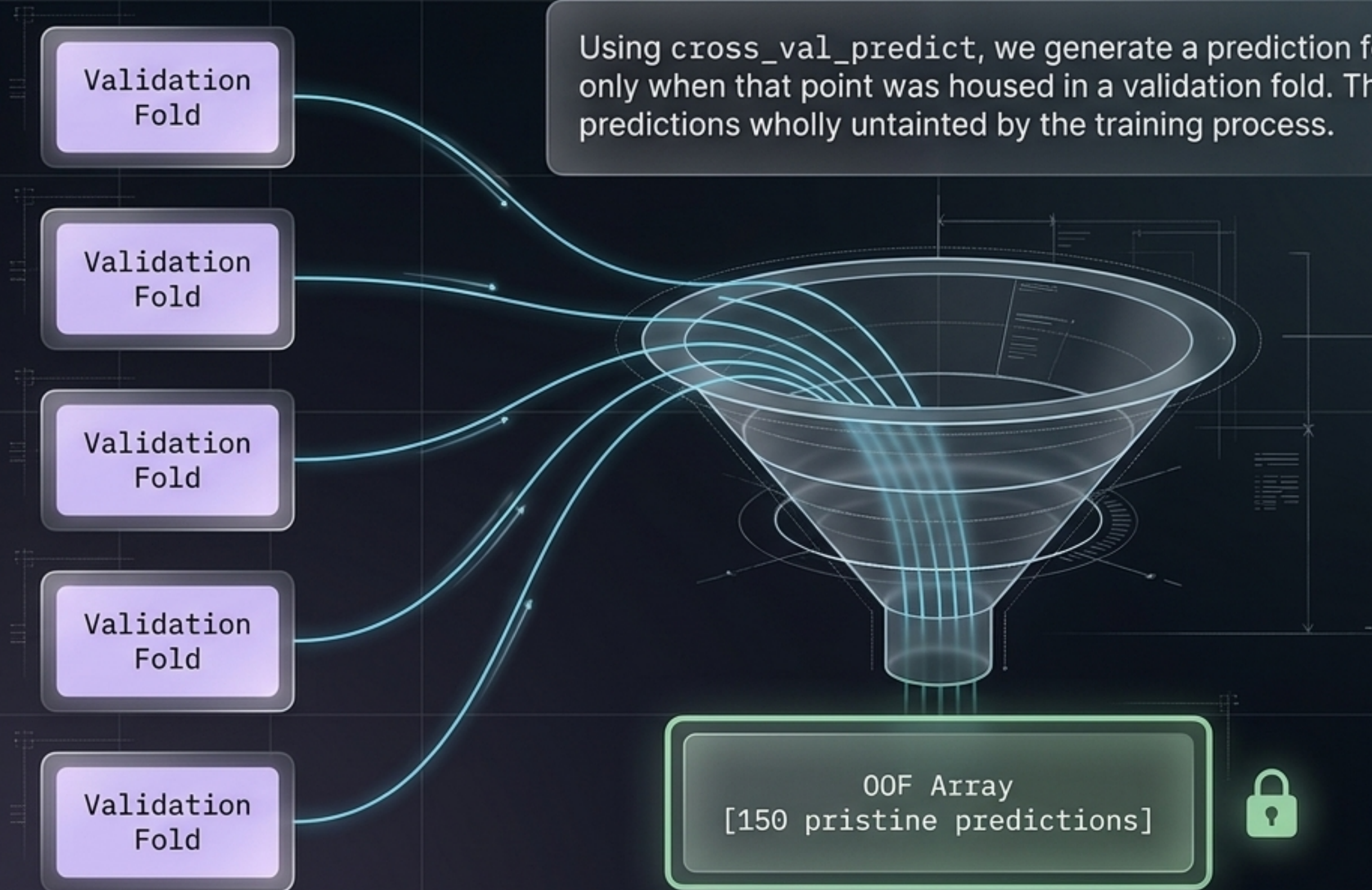
Rigorous Validation: Stratified K-Fold

To prevent biased benchmarking, we abandon simple train/test splits. Stratified K-Fold ensures every model is tested across multiple unique data permutations while strictly preserving the underlying class distribution in every fold.



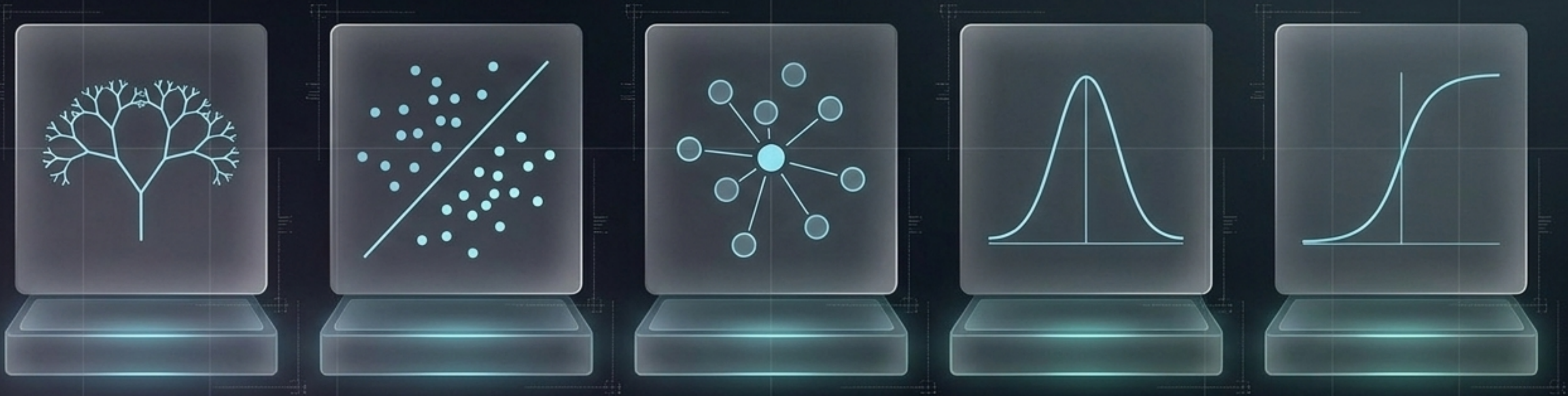
Eliminating Overfit: Out-of-Fold (OOF) Methodology

Using `cross_val_predict`, we generate a prediction for every single data point—but only when that point was housed in a validation fold. The result? A pristine array of predictions wholly untainted by the training process.



The 5-Algorithm Showdown

With our pristine evaluation protocol established, we introduce five distinct mathematical philosophies to the maXbox integration pipeline.



Random Forest

Support Vector
Machines

k-Nearest
Neighbors

Naive Bayes

Logistic
Regression

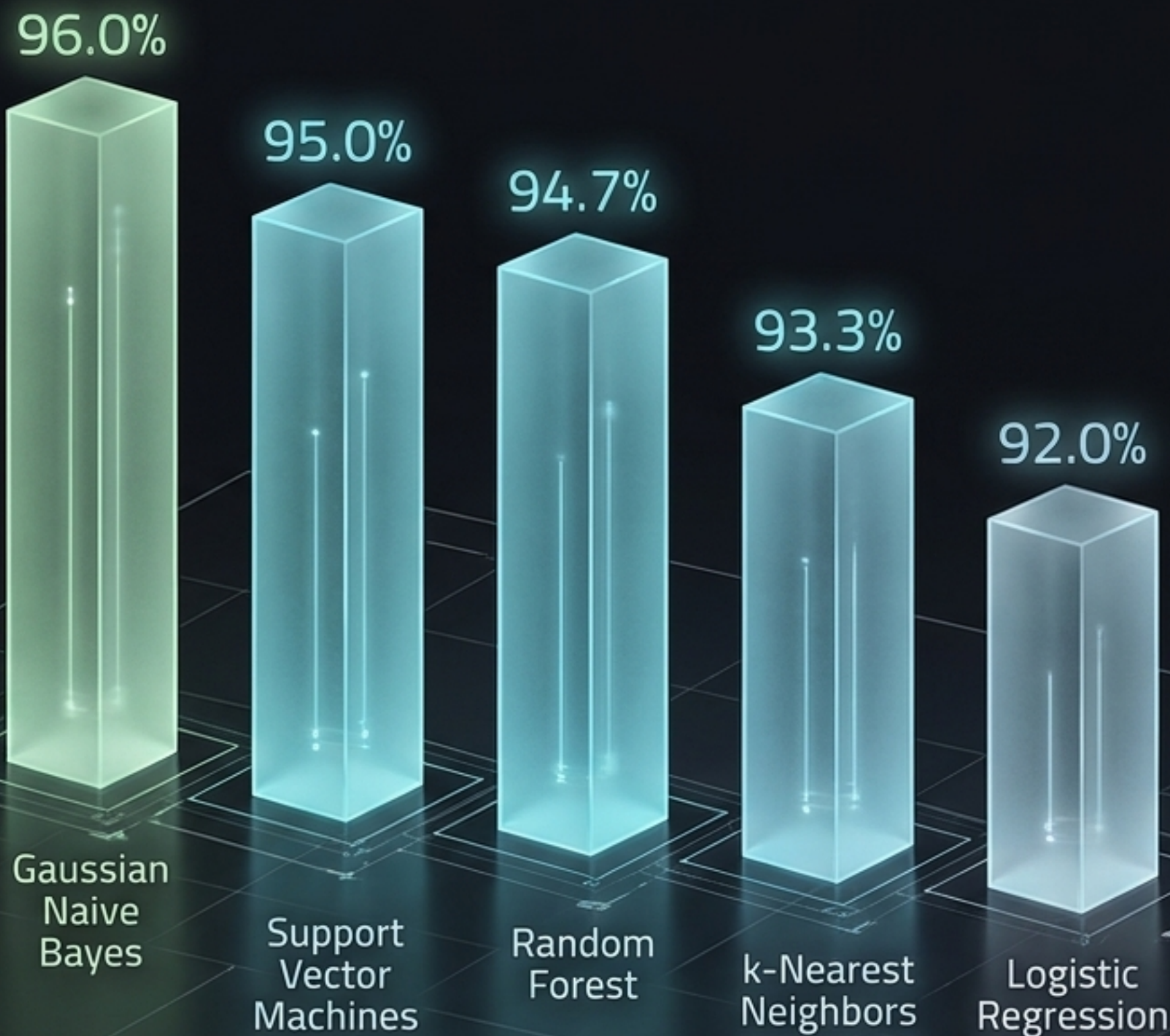
Architectural Philosophies

Classifier	Core Mechanism	Operational Strength
Random Forest	Ensemble Decision Trees	High accuracy, handles non-linear data overlaps
Support Vector Machines	Hyperplane Margin Maximization	Excellent edge-case definition in high dimensions
k-Nearest Neighbors	Distance-based voting (k=5)	Non-parametric, highly intuitive local baseline
Naive Bayes	Probabilistic / Gaussian	Blisteringly fast, effective with independent features
Logistic Regression	Sigmoid Probability	Supremely interpretable, solid linear boundary

```
execstr('rf.fit(X_train, y_train)')  
"writeln('Classify accuracy= ' +  
evalstr('accuracy_score(y_test,  
rf.predict(X_test))'))");
```

Naive Bayes Takes the Accuracy Crown

Processed via maXbox's Visual Machine Learning Studio, the probabilistic approach of Gaussian Naive Bayes marginally outperforms the ensemble strength of Random Forest on this specific feature configuration.



The Danger of Pure Accuracy



Precision:

Of all the flowers we labeled Virginica, how many actually were?



Recall:

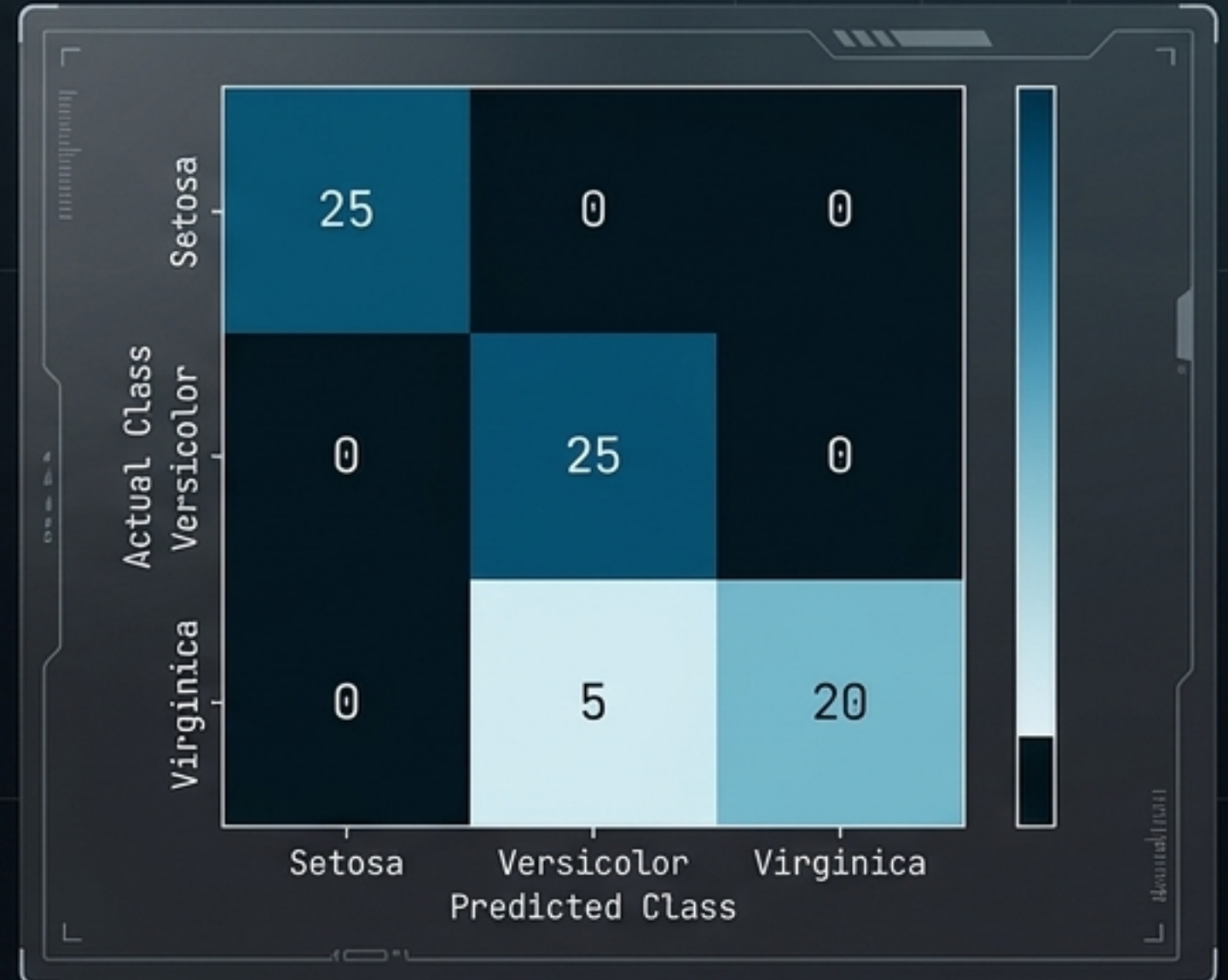
Of all the actual Virginica flowers, how many did we successfully detect?

An overall 93% accuracy rating can mask fatal operational failures. If you don't recognize a target, it's bad recall. If you confuse one target for another, it's bad precision.

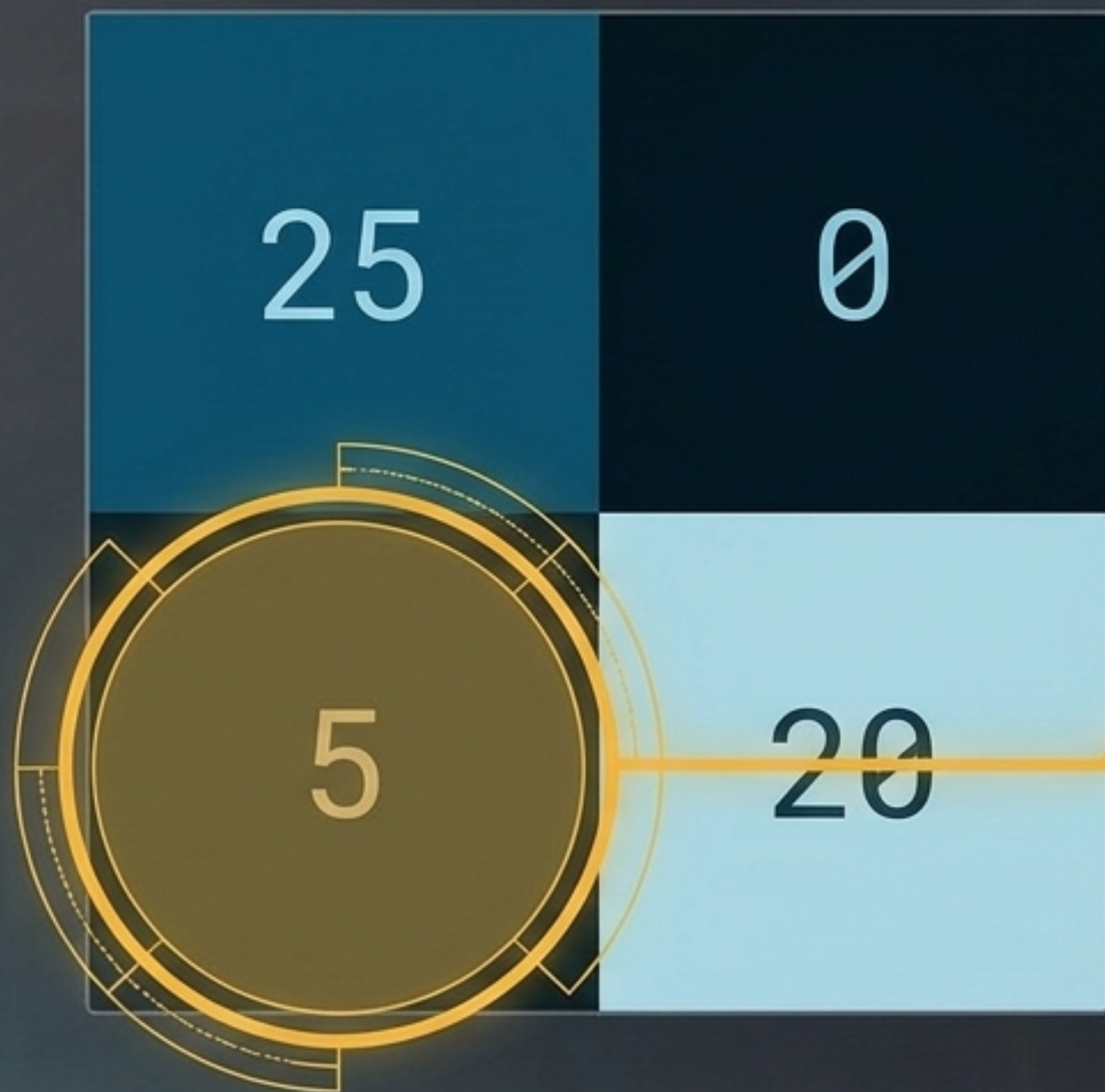
Visualizing the Fault Lines

Through Python4Delphi, we leverage Seaborn to generate rich, annotated heatmaps directly within the maXbox environment. This matrix pinpoints exactly where the model's logic breaks down.

```
execstr('import seaborn as sns');  
execstr('cm = pd.DataFrame(confusion_matrix(y_test, rf.predict(X_test)))');  
execstr('sns.heatmap(cm, annot=True)');
```



The Versicolor/Virginica Overlap



Virginica

Recall dragged down to

80.0%

Versicolor

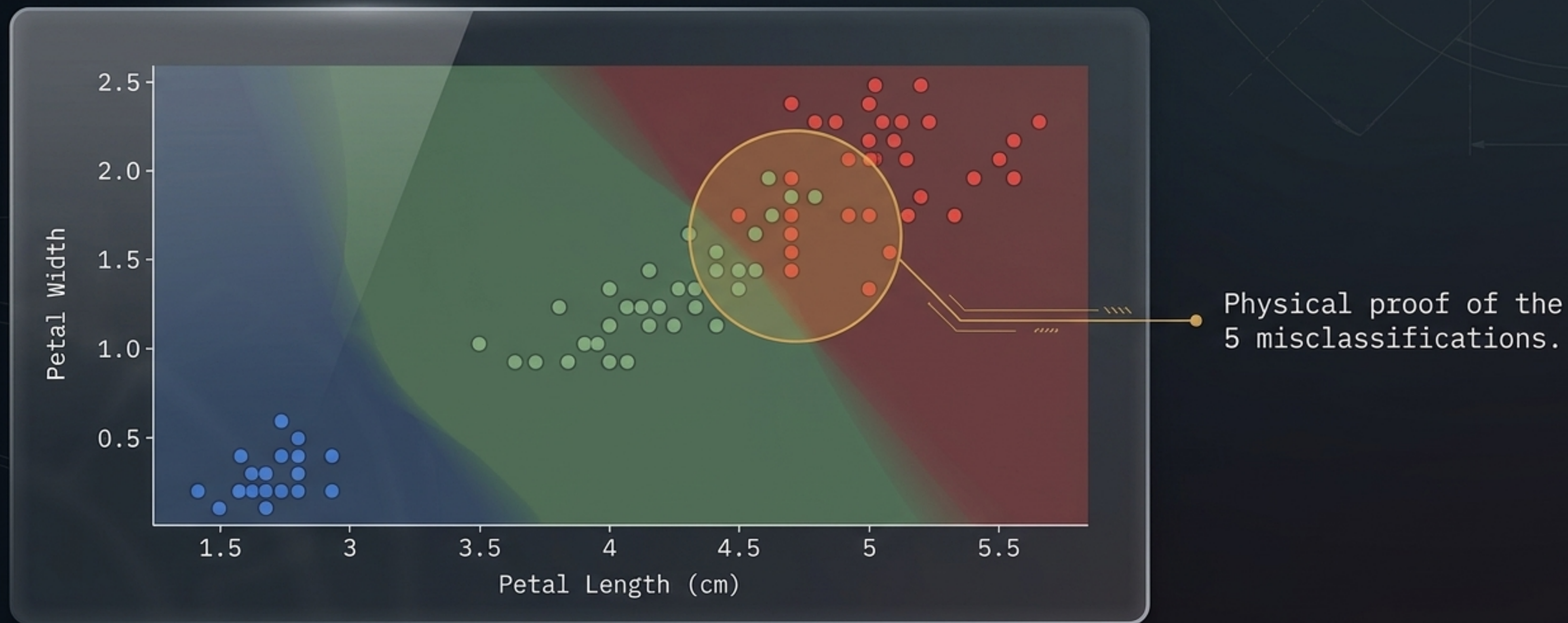
Precision falls to

83.3% (25/30)

The matrix reveals the nuance hidden by top-line accuracy. 20% of Virginica samples are misclassified.

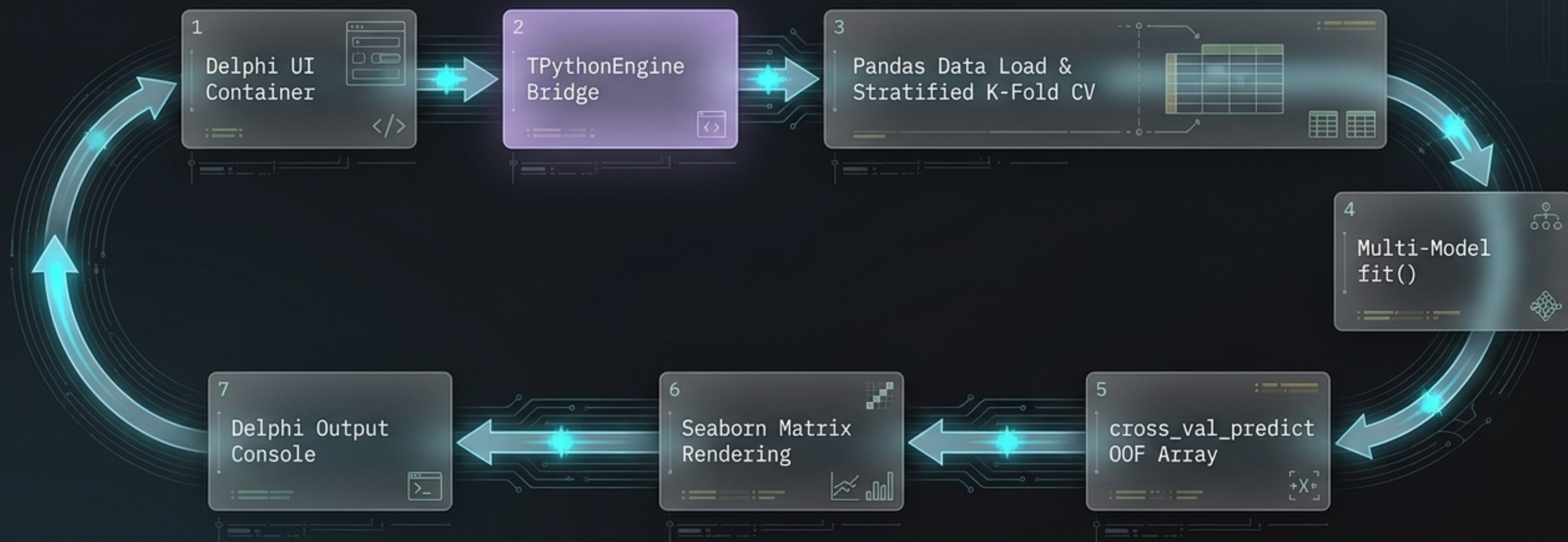
Seeing the Logic (Interactive Studio)

By projecting the out-of-fold predictions back onto the original feature space, we physically see the border disputes between classes.



The Embedded ML Blueprint

A cohesive, enterprise-ready pipeline. Rapid prototyping UI powered by the deepest machine learning libraries available, executed entirely natively.



Unlocking Analytical Power in Native Apps



Seamless Integration

`TPythonEngine` bridges compiled GUIs and dynamic Python without network overhead or API latency.

maXbox4exe

Digitally signed by
maXbox4exe
Date: 2026.10.09 16:31:06
+02'00'



Rigorous Validation

`Stratified K-Fold CV` and out-of-fold predictions ensure mathematical trust in benchmark results.



Actionable Insights

Embedding `Seaborn` matrices inside Delphi uncovers the "why" behind the accuracy, mapping direct paths to operational improvement.