

GitHub Issue

AutoML

What's in this session?

1. Install ML.NET Model Builder
2. Create new .NET CORE console project and add Machine Learning job
3. Pick a Scenario / Price Prediction
4. Set Data File
5. Train 60 seconds
6. Understand Train result
7. Understand evaluation result
8. Generate Code
9. Examine Code

Install ML.NET Model Builder

<https://marketplace.visualstudio.com/items?itemName=MLNET.07>



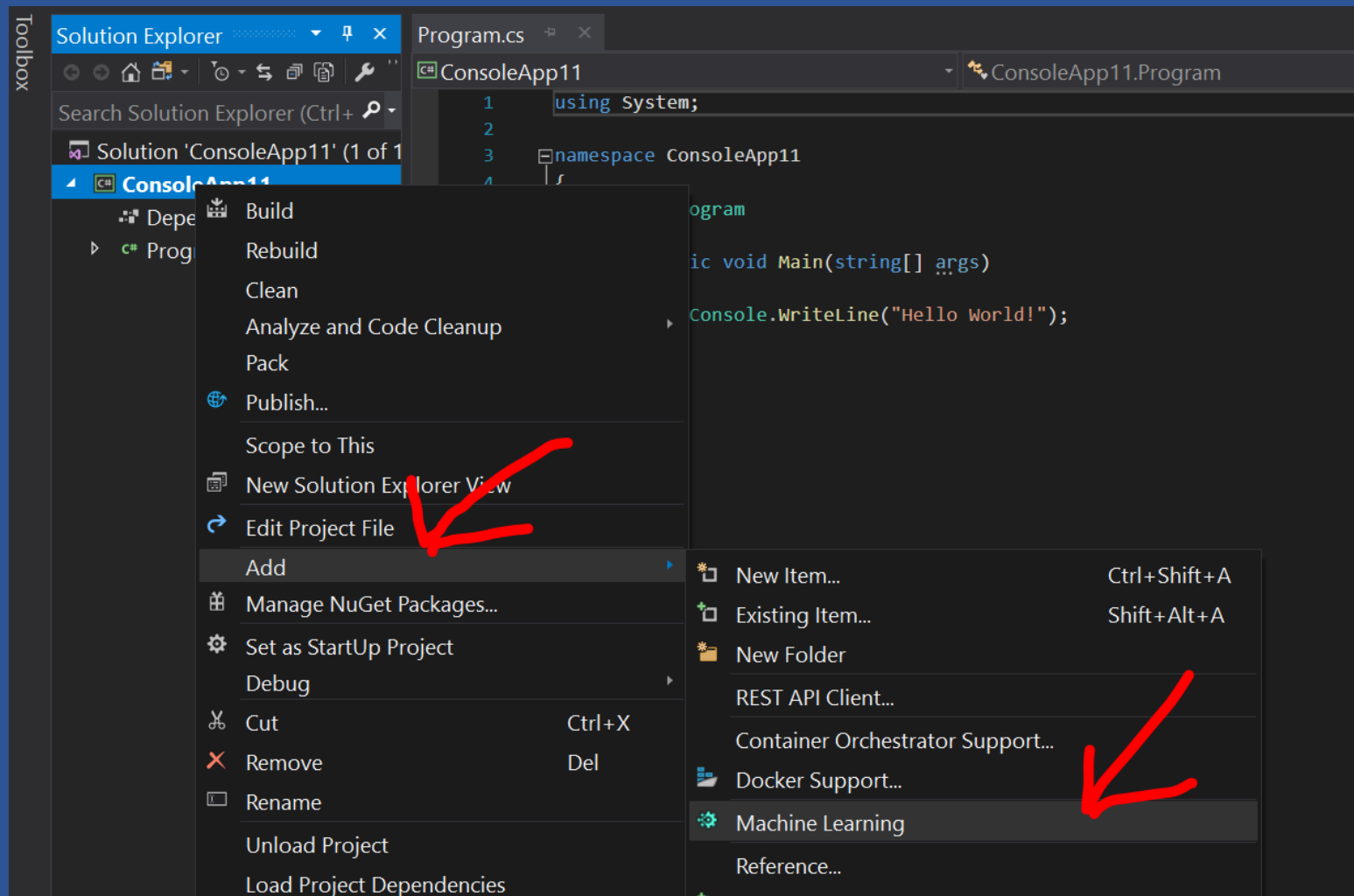
ML.NET Model Builder (Preview)

Microsoft | 24,428 installs |  39,194 downloads | ★★★★★ (12) | Free

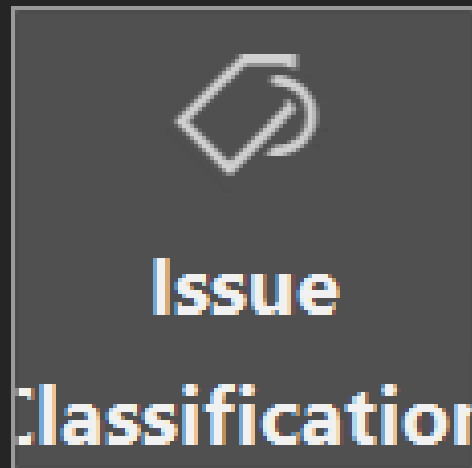
Simple UI tool to build custom machine learning models.

Download

Create new .NET CORE console project and add Machine Learning



Pick a Scenario / Issue Classification



Classify data into 3+ categories (multi-class classification), e.g. predict labels of GitHub issues.

Set Data File

Data / File / issues train.tsv
Label column name = Area

Input

Choose input data source from either SQL Server or File:


File

Select a file:

E:\ml\issues train.tsv



Supported file formats: csv or tsv. Maximum file size: 1 GB.

Column to Predict (Label): 


Area

Data Preview

ID	Area (Label)	Title	Desc
17	area-System.Xml	Some XPath.XDocument tests are failing	Som
20	area-System.Xml	2 XPath.XDocument tests fail because of lacking feature	XPat
22	area-System.Numerics	Two Numerics Tests are failing only on our CI server	Two
36	area-System.Numerics	SIMD test failures on non-ENU configurations.	After
41	area-System.Numerics	Quaternion operator overloads should be using the respective methods	Quat
49	area-Infrastructure	Add Linux/Mac build script	A `b
50	area-System.Numerics	Made Quaternion's operator overloads use their respective methods	Ope

Train 600 seconds

Input

Time to train (seconds): 

Cancel training

Progress

Start training to see progress and results

Status: 582 seconds remaining

Best Accuracy:

Best Algorithm:

Last Algorithm:

Understand Train result

Progress

Start training to see progress and results

Status:	Done
Best Accuracy:	85.23%
Best Algorithm:	SdcaMaximumEntropyMulti
Last Algorithm:	SymbolicSgdLogisticRegressionOva

Understand evaluation result

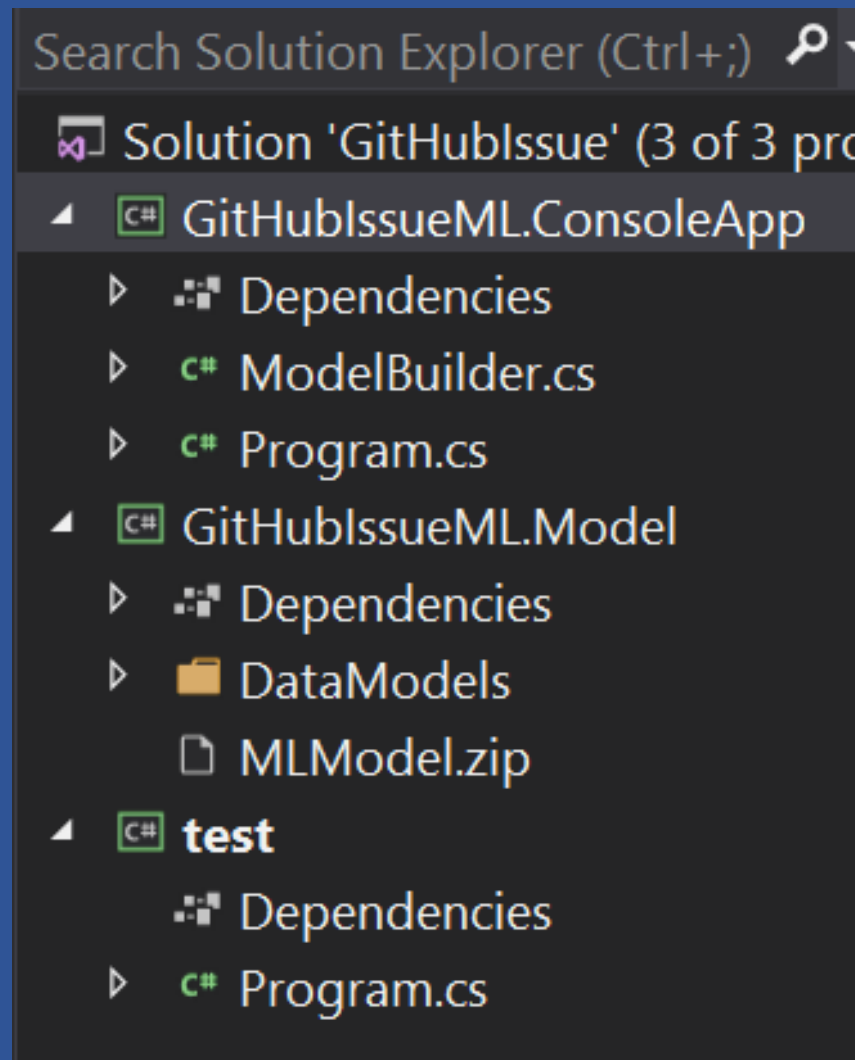
Output

ML Task: multiclass-classification
Dataset: issues_train.tsv
Column to Predict (Label): Area
Best Model: SdcaMaximumEntropyMulti
Best Model Accuracy: 85.23%
Training Time: 600.76 seconds
Models Explored (Total): 4

Top 4 models explored

Rank	Trainer	MicroAccuracy	MacroAccuracy	Duration
1	SdcaMaximumEntropyMulti	0.8523	0.8159	60.5
2	LightGbmMulti	0.8414	0.7893	306.7
3	AveragedPerceptronOva	0.8327	0.7927	18.2
4	SymbolicSgdLogisticRegressionOva	0.7784	0.7353	17.1

Examine Code



Next Step

Write Code to build, train, evaluate, and use ML model