

Machine Learning at Microsoft with ML.NET

Matteo Interlandi^m, Sergiy Matushevych^m, Misha Bilenko^y, Saeed Amizadeh^m, Shauheen Zahirazami^m, Markus Weimer^m



A machine learning framework made for .NET developers.

<https://dot.net/ml>

Open source

<https://github.com/dotnet/machinelearning>

Used at Microsoft for several years and by hundreds of services



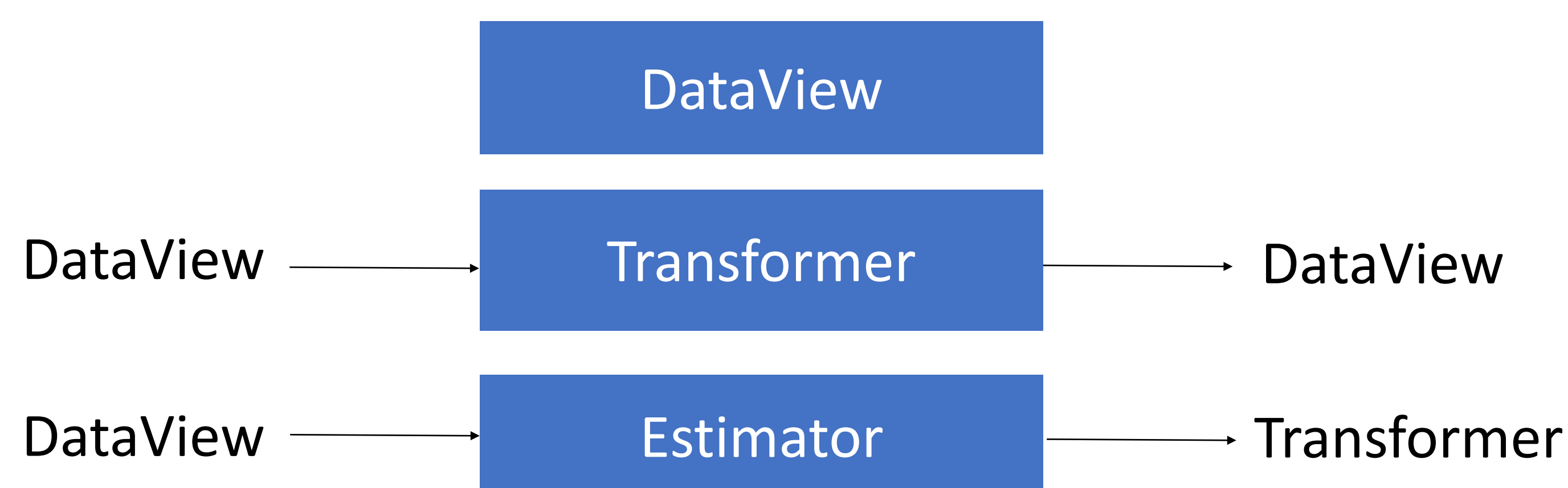
Comment	Toxic? (Sentiment)
==RUDE== Dude, you are rude ...	1
== OK! == IM GOING TO VANDALIZE ...	1
I also found use of the word "humanists" confusing ...	0
Ooooooh thank you Mr. DietLime ...	0

Features (input)

Label (output)

```
var context = new MLContext();
var loader = context.Data.TextReader(new
    TextLoader.Arguments { Column = new[] {
        new TextLoader.Column("Label", DataKind.Bool, 0),
        new TextLoader.Column("Text", DataKind.Text, 1)
    }
});
IDataView trainingData = loader.Reader(trainDataPath);
var featurizer = context.Transforms.Text
    .FeaturizeTex("Text", "Features");
var trainer = context.BinaryClassification.Trainers
    .FastTree("Label", "Features");
IEstimator pipeline = featurizer.Append(trainer);
ITransformer model = pipeline.Fit(trainingData);
```

ML.NET is a framework for building ML pipelines. Main abstractions:



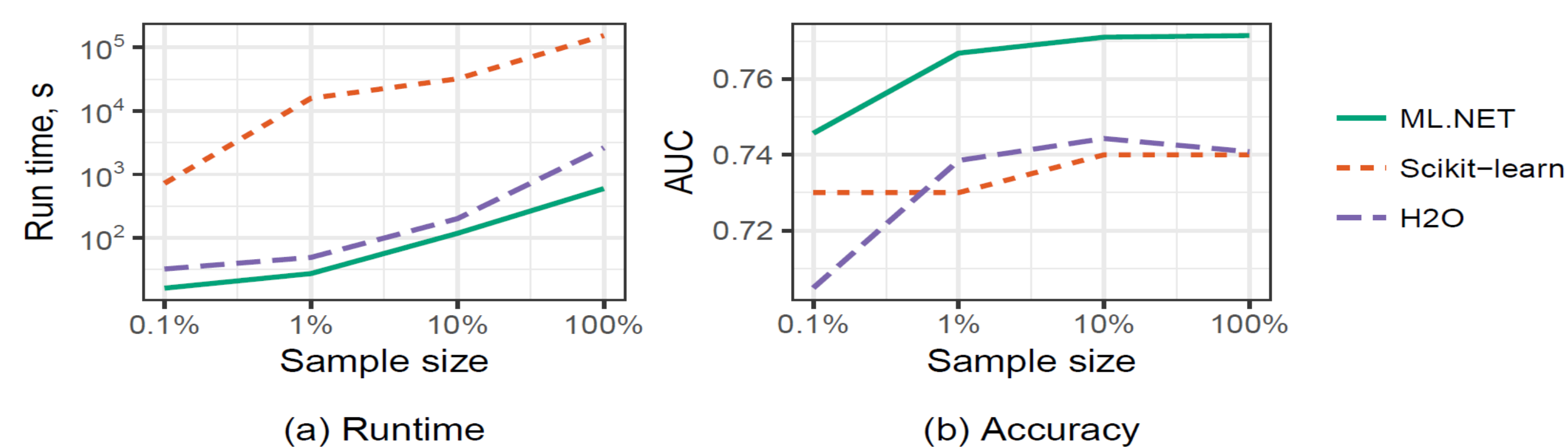
DataView: data abstraction defined over a schema which expresses a sequence of columns names with related types

Properties: DataViews are *composable*: new DataViews are formed by applying Transformations over other DataViews; DataViews are *virtual*, i.e., they can be *lazily* computed on demand from other DataViews without having to materialize any partial results; and since a DataView does not contain values, but computes values from its source, it is *immutable* and *deterministic*

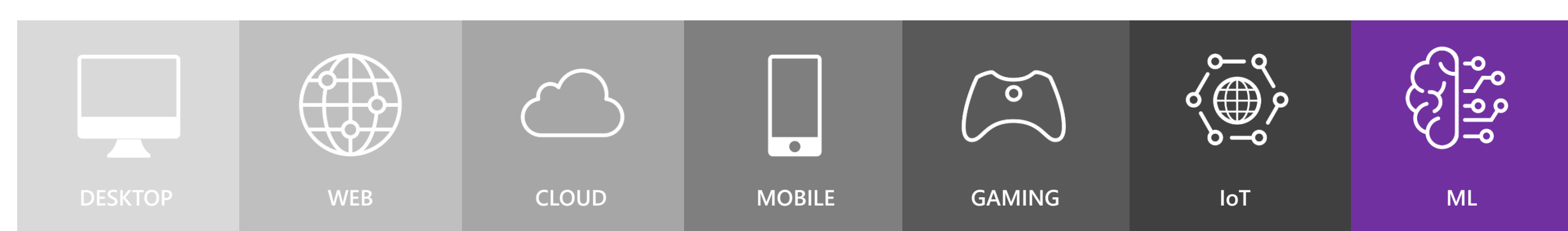
Transforms: data operators for data featurization. ML.NET provides hundreds of transform for structured, textual, or image data. ML.NET ships with ~50.

Learner: trainable ML models. ML.NET supports linear, trees, clustering and DNN models (through external bindings to TensorFlow and ONNX).

Performance (Criteo)



.NET is the platform for building anything



In any platform



With easy to use functional / object oriented languages



And native integration with C/C++ for best performance



25,000 CONTRIBUTORS	1700 COMPANIES	55 ACTIVE PROJECTS	462 RESOURCES
------------------------	-------------------	-----------------------	------------------

.NET Open Source Software

