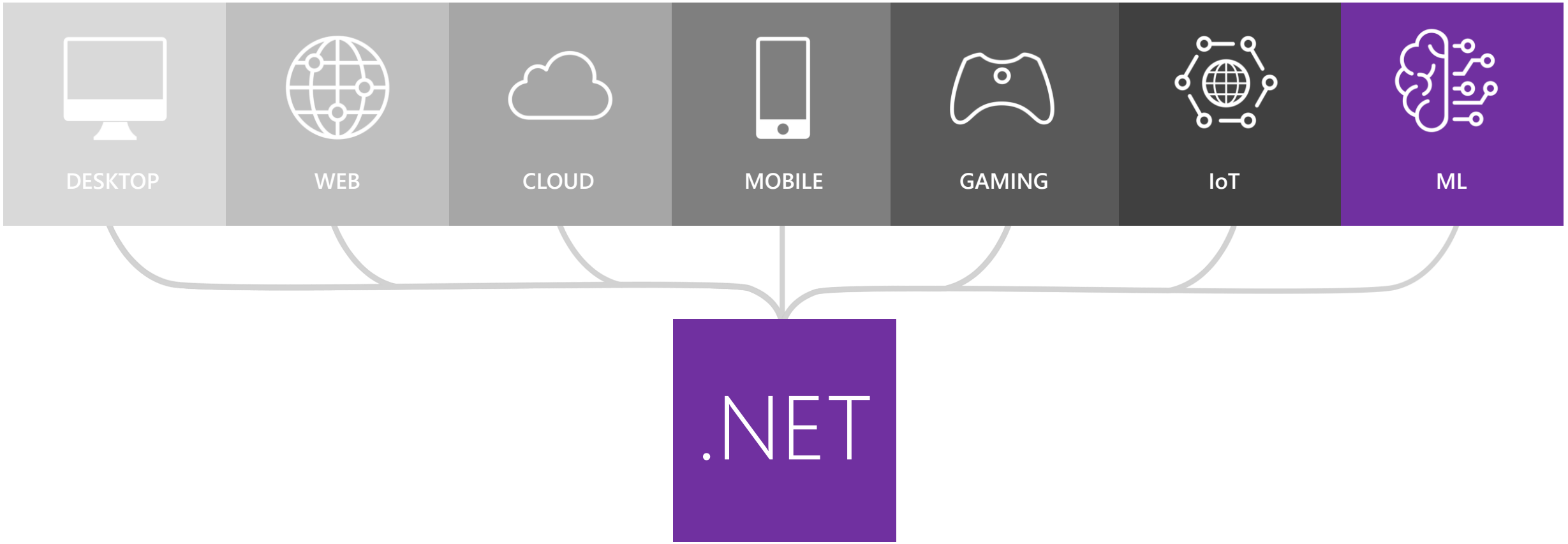
Microsoft

ML.NET: A cross-platform and open source machine learning framework

Toan Huynh

Your platform for building **anything**



Machine Learning
"Programming the UnProgrammable"

$f(x)$ { Is this a face?

$f(x)$ { Price of Shirt?

"It has **exquisite** buttons … with **long sleeves** …works for casual as well as **business settings**"

Machine Learning
"Programming the UnProgrammable"

Machine Learning creates a

$f(x)$

Model

Using this data

Face

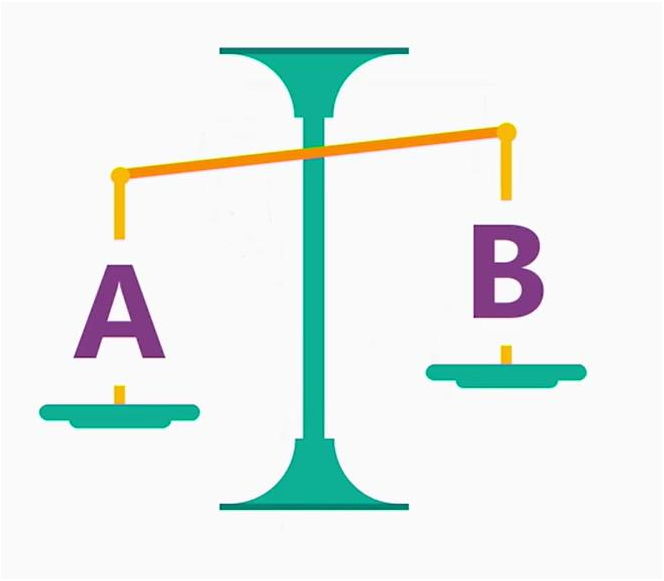
Face

Not a face

Not a face

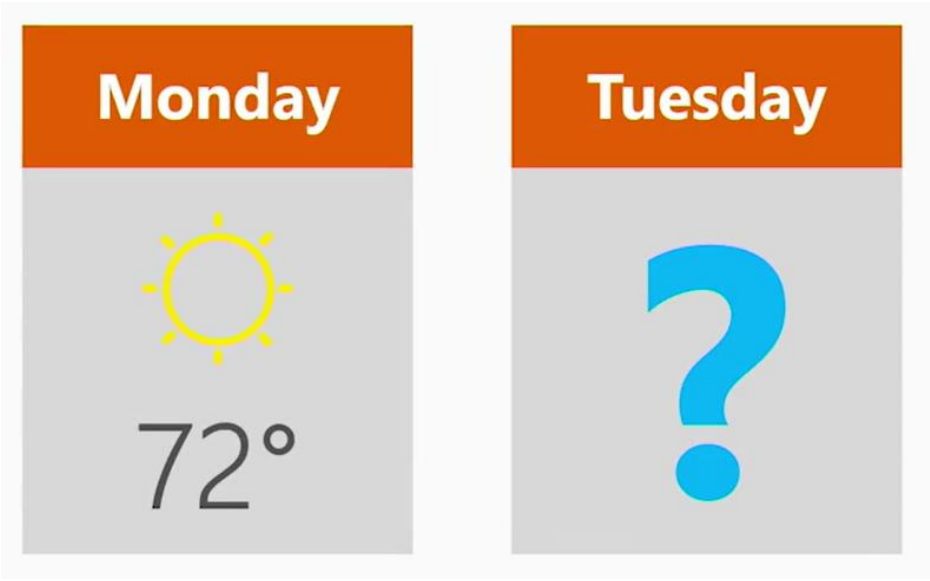
Many ML Tasks…

Is this A or B?



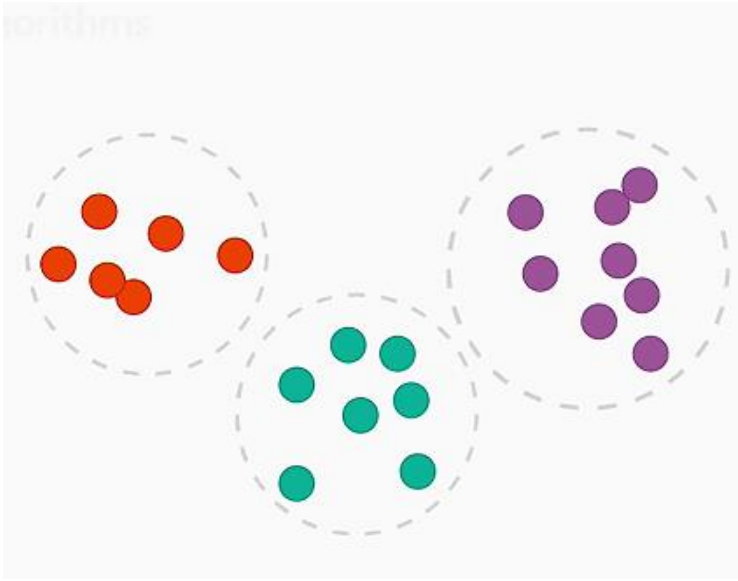
Classification

How much? How many?



Regression

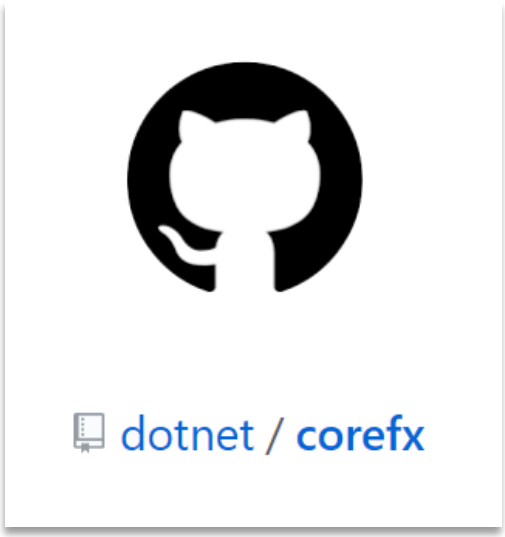
How is this organized?



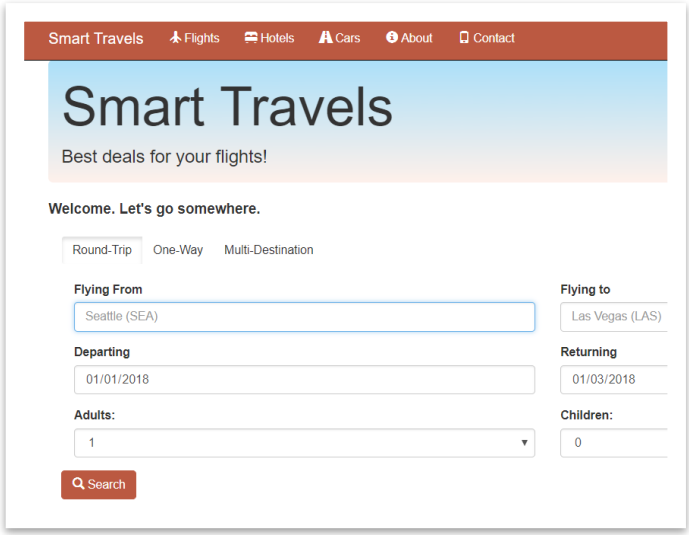
Clustering

Several Example Scenarios

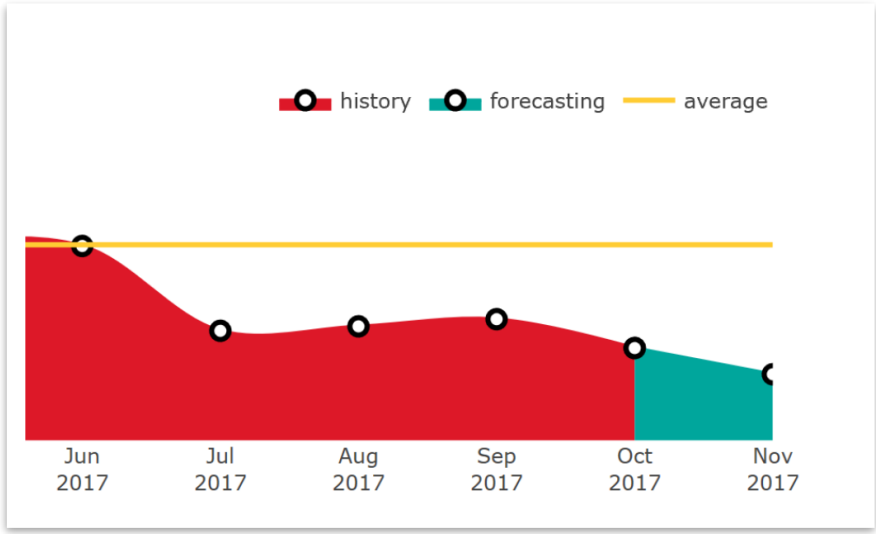
Solve three problems with ML.NET



Issue Classification

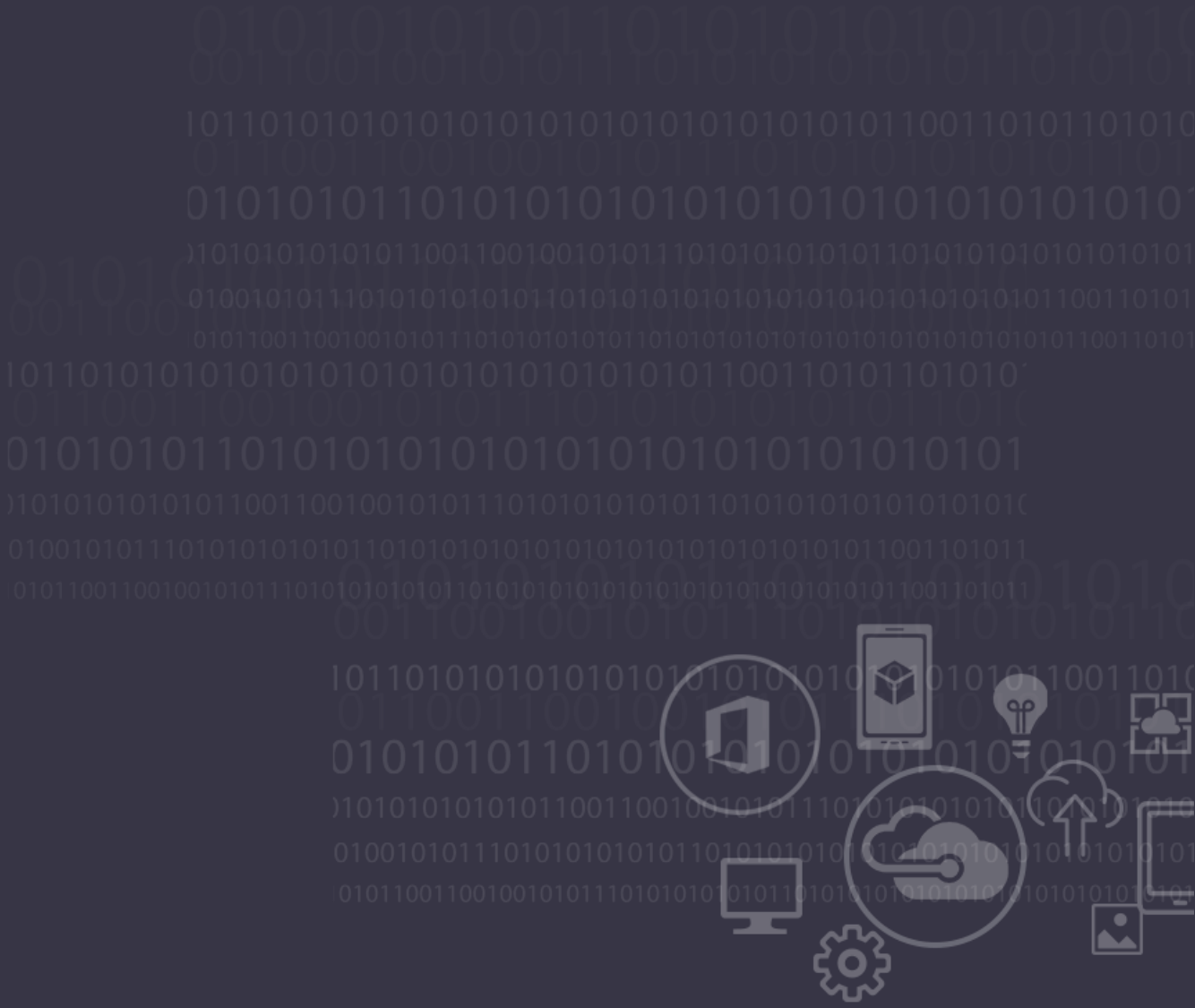


Flight Delays



Sales Forecasting

ML.NET



01 WHAT IS ML.NET?

ML.NET is a cross-platform, open source machine learning framework

ML.NET Components

https://github.com/dotnet/machinelearning

API	Transforms	Learners	Misc.
Training	Natural Text	Linear	ML Data framework
Inference	Image	Boosted Trees	Evaluators
Model management	Schema	Svm	Calibrators
	Missing values	K-Means	Data loaders
	Categorical	Etc.	Auto-inference
	Normalization		Etc.
	Feature Selection		
	Pre-trained models		

01 WHAT IS ML.NET?

ML.NET is for building your own ML models

Two **flavors** of Machine Learning

Pre-built ML Models (Cognitive Services)

Vision

Speech

Language

Labs

Knowledge

Search

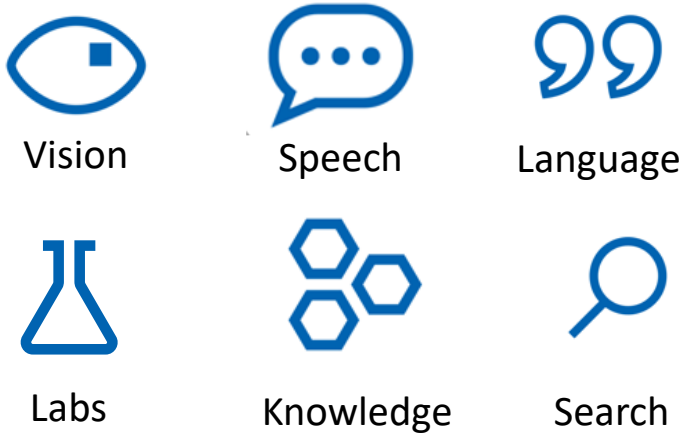
Custom ML Models

TensorFlow

Easy / Less Control

Full Control / Harder

Pre-built ML Models (Azure Cognitive Services)



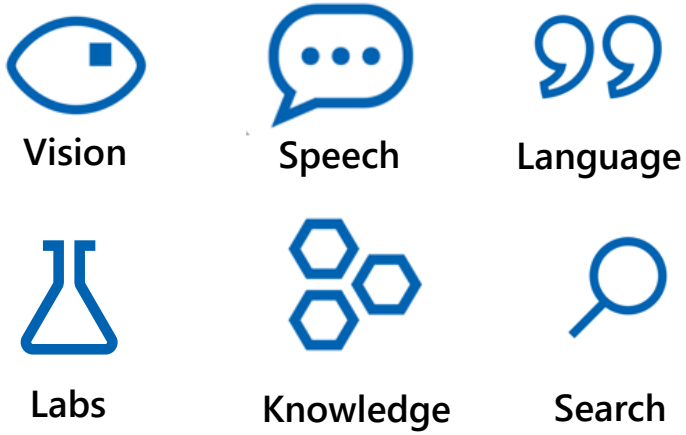
Consume (C#, VB, F#)

e.g. Sentiment Analysis using Azure Cognitive Services

```
TextAnalyticsAPI client = new TextAnalyticsAPI();
client.AzureRegion = AzureRegions.Westus;
client.SubscriptionKey = "1bf33391DeadFish";

client.Sentiment(
new MultiLanguageBatchInput(
new List<MultiLanguageInput>()
{
 new MultiLanguageInput("en","0",
     "This is a great vacuum cleaner")
}));
```

96% positive

Pre-built ML Models (Azure Cognitive Services)



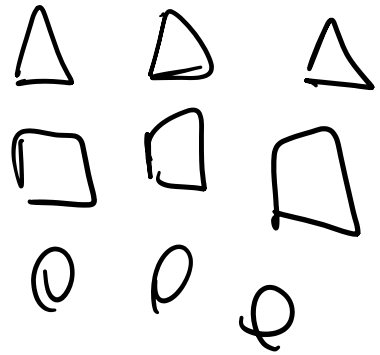
Consume (C#, VB, F#)

e.g. Sentiment Analysis using Azure Cognitive Services

```
TextAnalyticsAPI client = new TextAnalyticsAPI();
client.AzureRegion = AzureRegions.Westus;
client.SubscriptionKey = "1bf33391DeadFish";

client.Sentiment(
new MultiLanguageBatchInput(
new List<MultiLanguageInput>()
{
new MultiLanguageInput("en","0",
    "This vacuum cleaner sucks so much dirt")
}));
```

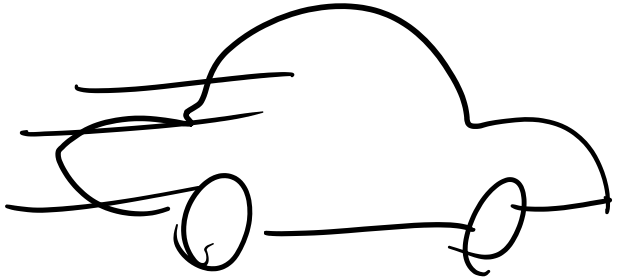
☹ 9% positive

Build your own (custom) ML Models



Prepare Your Data

Build & Train



Run

Build your own (custom) ML Models

Existing Solutions

- Python, R are great for ML and Data Science
- ML.NET is another way to do it with familiar tools
  - .NET currently lacks ML libraries and ML essentials
- ML.NET complements the experience that AML, CogSvcs provides
  - Build your own
  - Code First approach
  - AppLocal Model deployment

ML.NET is proven in large scale Microsoft products

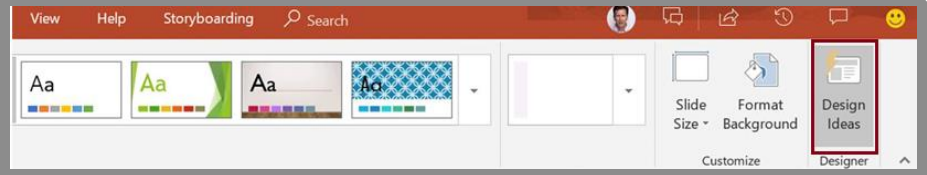
ML.NET Usage at Microsoft

"This was made through **Design Ideas and ML.NET**"

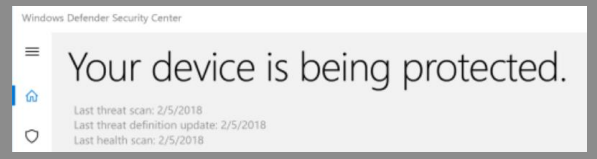
Bing Ads

Excel

Power Point



Windows 10



+ more!

ML.NET 0.1 (Preview)

The Machine Learning framework made by and for .NET developers

Supported on Windows, Linux, and macOS

Customizable

Developer Focused

Proven & Extensible

Open Source

Try it out @ **https://github.com/dotnet/machinelearning**

ML Primer

Machine Learning: Programming the Unprogrammable

Write a function that tells me…

The price of a shirt based on its description?

*It has **exquisite** buttons … with **long sleeves***
*…works for casual as well as **business settings***

Is this a face?

Machine Learning: Programming the Unprogrammable

We do have examples of inputs and outputs

Face

Face

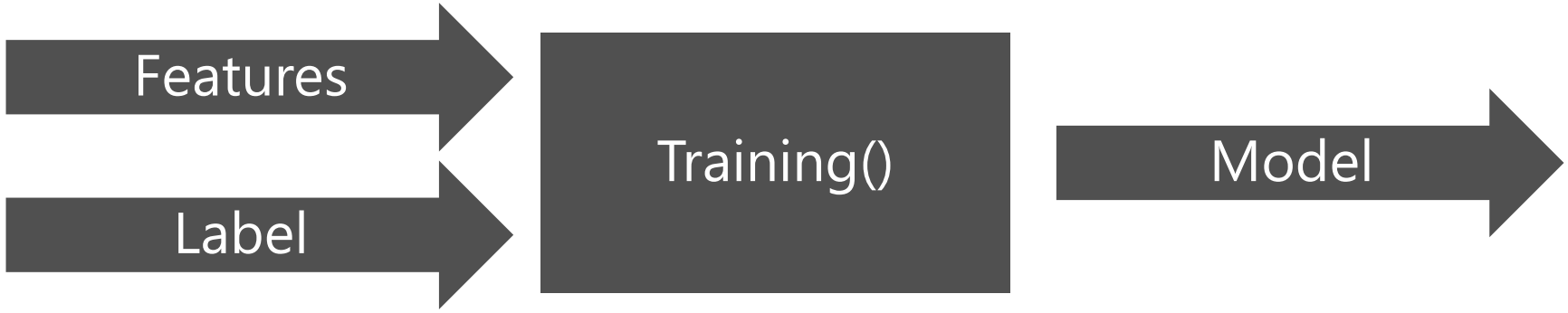
Not a face

Not a face

Machine Learning creates a function (model) using this data

ML Models: Supervised Learning

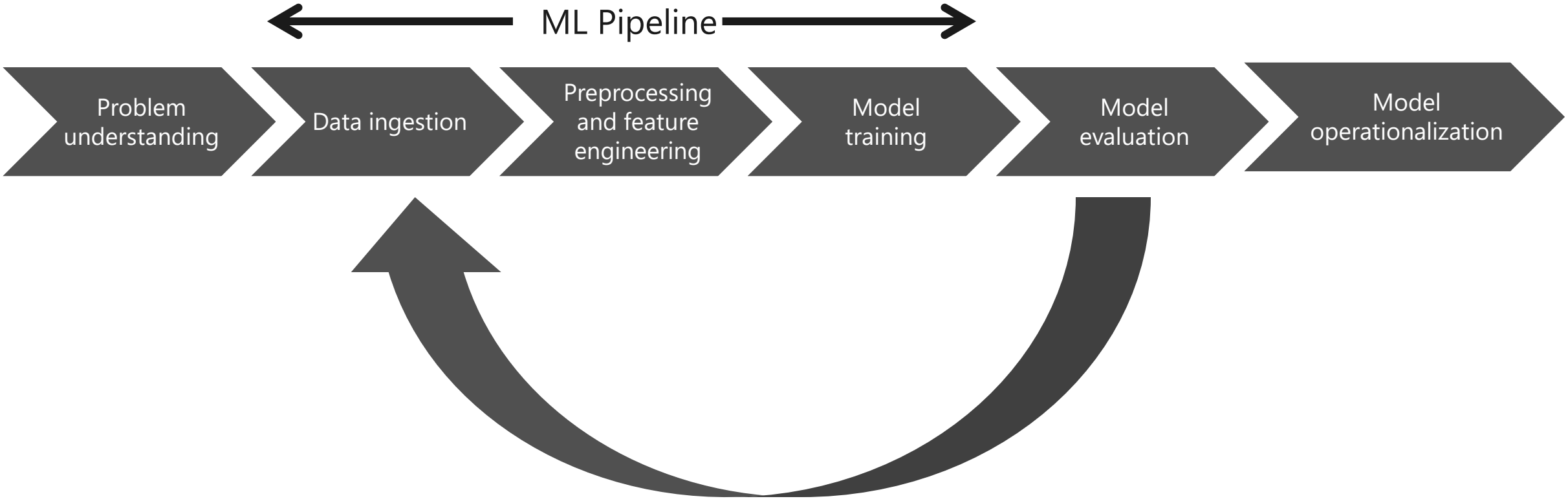
Training



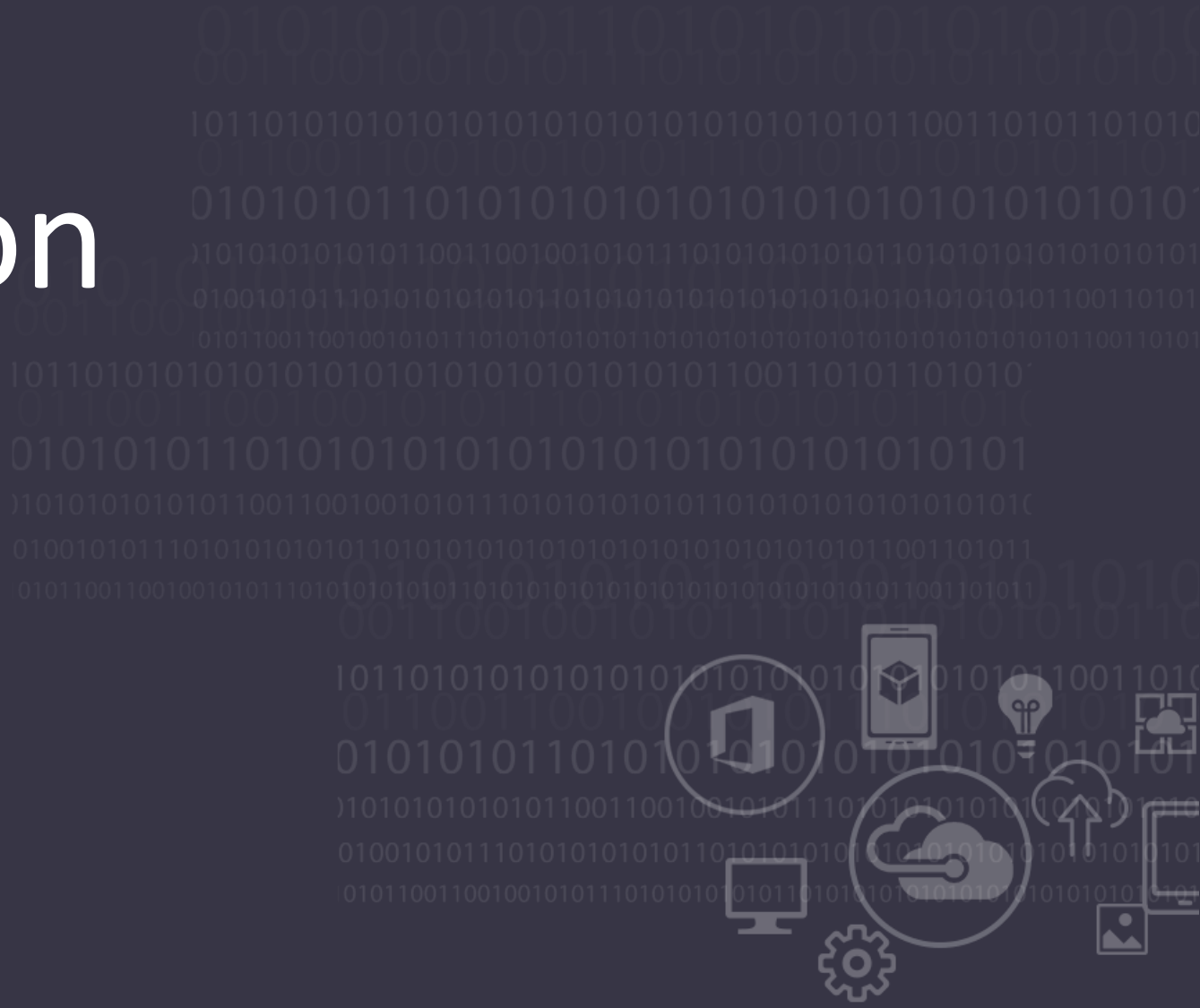
Prediction

Evaluation: Is the prediction close to the label (true value)?

Machine Learning Workflow



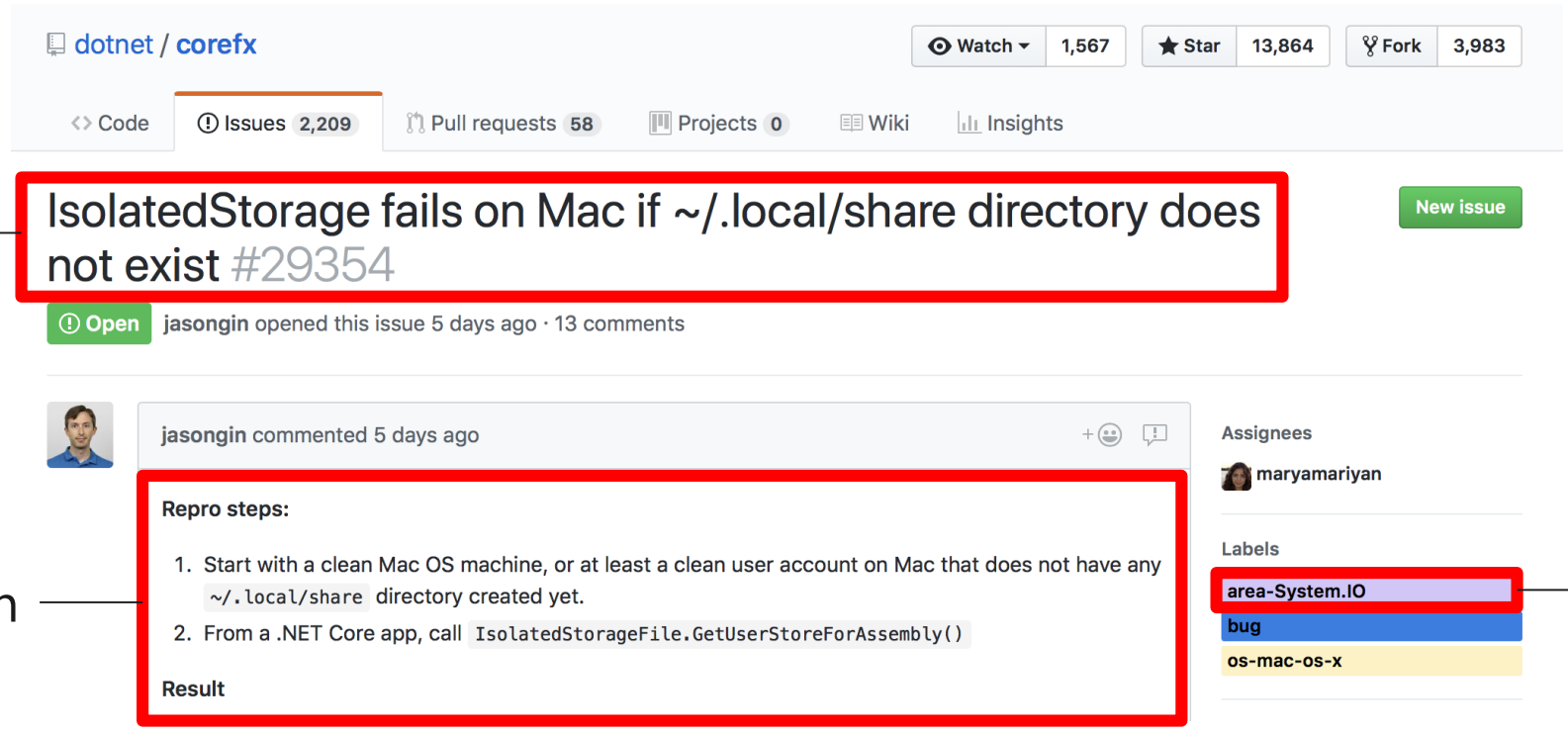
Text classification



Problem Understanding

Feature: Title

Feature: Description



Label

Problem understanding → Data ingestion → Preprocessing and feature engineering → Model training → Model evaluation → Model operationalization

Problem understanding

Feature: Title

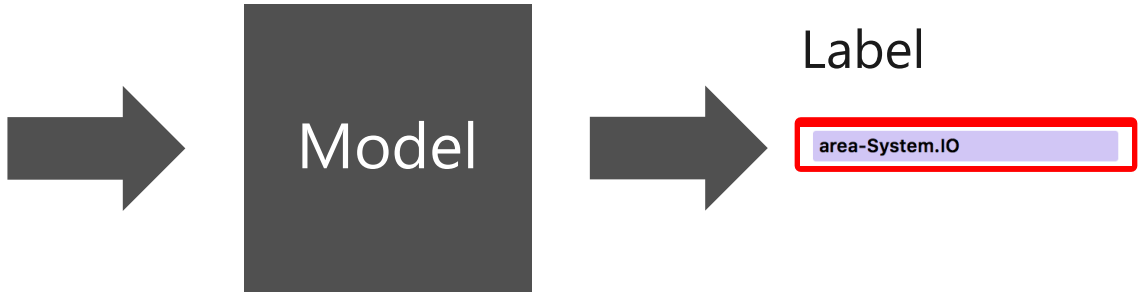
IsolatedStorage fails on Mac if ~/.local/share directory does not exist #29354

Feature: Description

Repro steps:

1. Start with a clean Mac OS machine, or at least a clean user account on Mac that does not have any `~/.local/share` directory created yet.
2. From a .NET Core app, call `IsolatedStorageFile.GetUserStoreForAssembly()`

Result



Demo:
GitHub Issue Classification

Problem understanding: Classification

- Training data: issues belong to different areas (classes)
- Predict the class of a new issue

area-Microsoft.CSharp

area-Microsoft.VisualBasic

area-Microsoft.Win32

area-Serialization

area-System.AppContext

area-System.Buffers

area-System.CodeDom

Data ingestion

- Collect GitHub issues using the Octokit library

ID	Area	Title	Description
13235	area-System.Runtime	Bring remaining System.Runtime members to netsta	Adds src, ref, and tests for the remaining members of System.Runti
28639	area-System.Net.Security	Stop blocking explicit opt-in to old SslProtocols in Ssl	By default, .NET Core defaults to allowing TLSv1, TLSv1.1, and TLSv
7899	area-System.Net	Failing network tests in the CI	Of the past dozen or so PRs I've done I've noticed the following test
20234	area-System.Collections	Provide consistent support for null elements in immu	Provide consistent support for null elements in ImmutableHashSet<
22689	area-System.Numerics	Add Span-based APIs to BigInteger	Adds: ```C# public BigInteger(ReadOnlySpan<byte> value); publ

Data ingestion: dataset understanding

- 16286 issues from Nov 2014 to Apr 2018
- 22 Area values (Labels)
- Min-mix description length

ID	Area	Title	Description
13235	area-System.Runtime	Bring remaining System.Runtime members to netsta	Adds src, ref, and tests for the remaining members of System.Runti
28639	area-System.Net.Security	Stop blocking explicit opt-in to old SslProtocols in Ssl	By default, .NET Core defaults to allowing TLSv1, TLSv1.1, and TLSv
7899	area-System.Net	Failing network tests in the CI	Of the past dozen or so PRs I've done I've noticed the following test
20234	area-System.Collections	Provide consistent support for null elements in immu	Provide consistent support for null elements in ImmutableHashSet<
22689	area-System.Numerics	Add Span-based APIs to BigInteger	Adds: ```C# public BigInteger(ReadOnlySpan<byte> value); publ

Data ingestion

```
var pipeline = new LearningPipeline();

pipeline.Add(new TextLoader<CoreFxIssue>(dataPath, header: true));

class CoreFxIssue
{
        public string ID;
        public string Area;
        public string Title;
        public string Description;
}
```

Preprocessing and feature engineering

- Get data into format suitable for the learner (often numeric vector)
- Extract signal to make it easier for the learner

Preprocessing and feature engineering

- Example: Text featurization
  - Text preprocessing (e.g. remove punctuation)
  - N-gram extraction (extract consecutive word sequences)

"To be or not to be."

"to be or not to be"

to	be	or	not	to be	be or	or not	not to
2	2	1	1	2	1	1	1

Preprocessing and feature engineering

0	Original data
1	Dictionarizer
2	TextFeaturizer (on Title)
3	TextFeaturizer (on Description)
4	Concatenate

Area	Title	Description
area-System.Net	Failing network tests in the CI	Of the past dozen or so PRs I've done I've noticed the following tests …
5	Failing network tests in the CI	Of the past dozen or so PRs I've done I've noticed the following tests …
5	**<0.45,0.34,1.43, …>**	Of the past dozen or so PRs I've done I've noticed the following tests …
5	<0.45,0.34,1.43, …>	**<0.27, 0.43, 0.24, 0.54, … >**
5	**<0.45,0.34,1.43, …, 0.27, 0.43, 0.24, 0.54, …>**	

Problem understanding | Data ingestion | Preprocessing and feature engineering | Model training | Model evaluation | Model operationalization

Model training

```
pipeline.Add(new StochasticDualCoordinateAscentClassifier());

PredictionModel<CoreFxIssue, CoreFxIssuePrediction> model = pipeline.Train<CoreFxIssue,
CoreFxIssuePrediction>();

public class CoreFxIssuePrediction
{
        [ColumnName("PredictedLabel")]
        public string Area;
}
```

Model evaluation

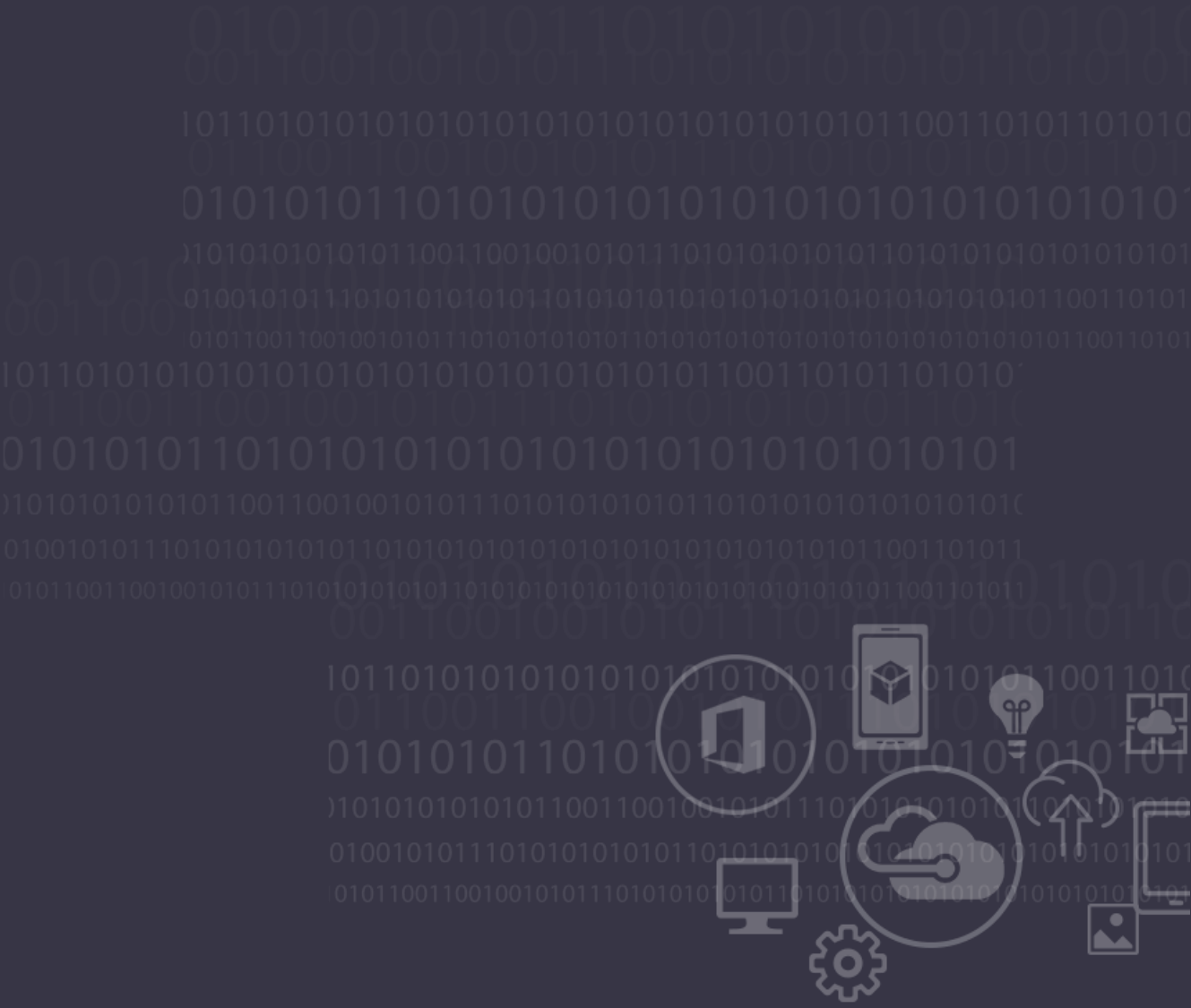
- Model must be valuable on new data (generalization)
- Evaluate on test data: labeled data not used for training
- Micro-accuracy: what percent of examples in test data are predicted correctly?

```
var evaluator = new ClassificationEvaluator();
ClassificationMetrics metrics = evaluator.Evaluate(model, testData);
Console.WriteLine("Micro-Accuracy is {0}", metrics.AccuracyMicro);
```

Model operationalization

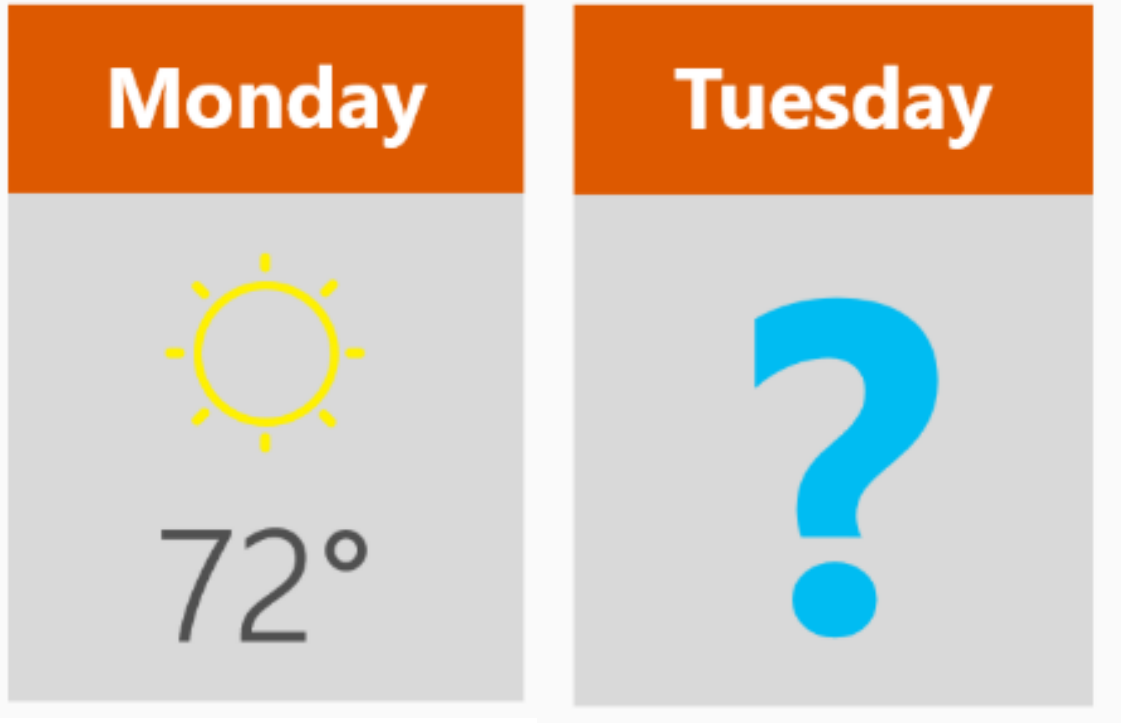
```
CoreFxIssue issue = new CoreFxIssue
{
        ID = "17475",
        Title = "Failing network tests in the CI",
        Description = @"Of the past dozen or so PRs I've done I've noticed the
following tests …"
};

CoreFxIssuePrediction prediction = model.Predict(issue);

Console.WriteLine("Area-Tag is " + prediction.Area);
```

Regression



Predicting Airline Delay

How much? How many?



Regression algorithms make numerical predictions, such as:

- What will the temperature be next Tuesday?
- What will my fourth quarter sales be?

Predicting Airline Delay

Flight Dataset

- How to get better results with the flight dataset
- What are some interesting insights from this data

  e.g. https://www.kaggle.com/fabiendaniel/predicting-flight-delays-tutorial
- Interesting charts or something insightful and interesting for our audience

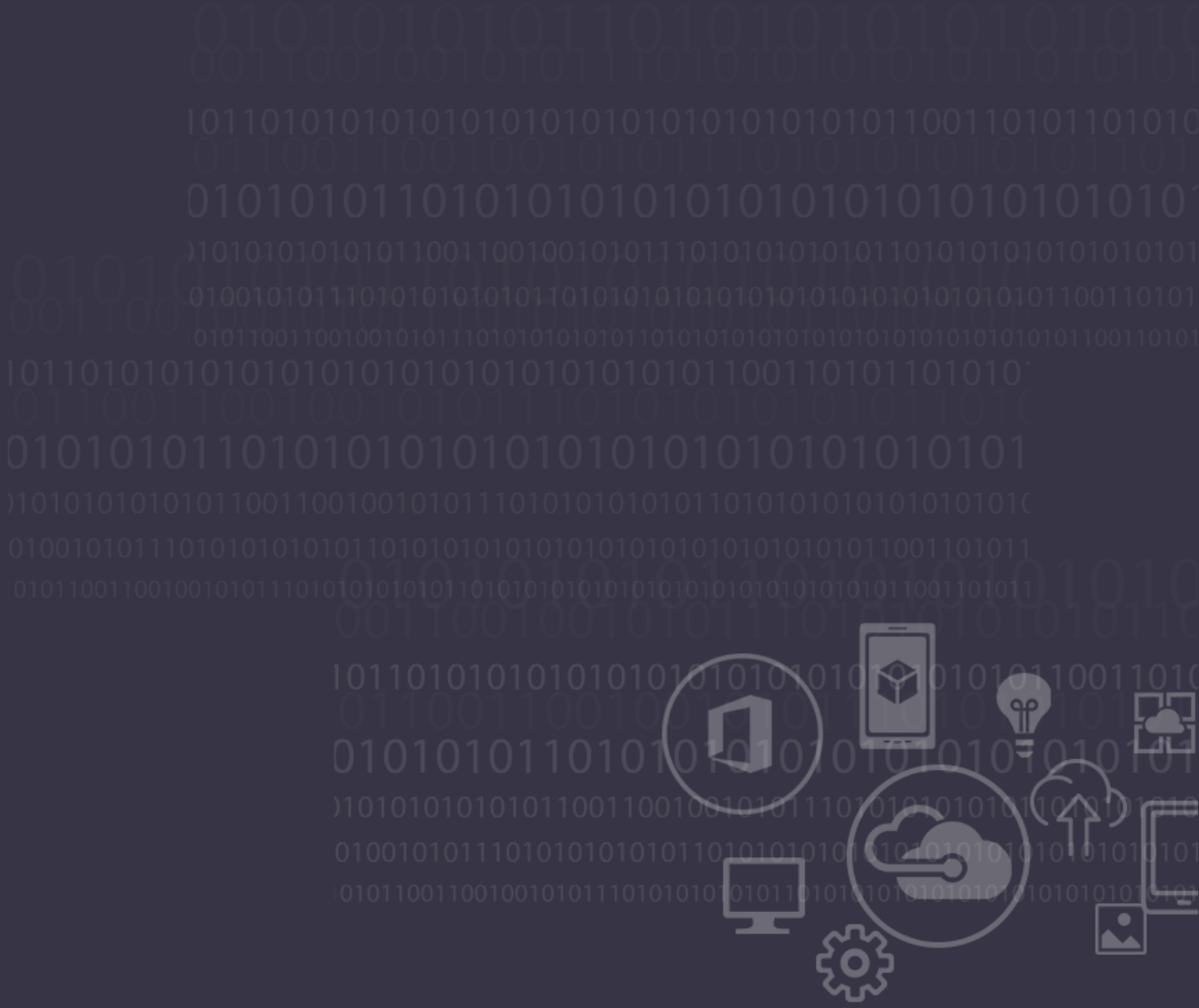
Machine learning is iterative

- Gather more or different data
  - User profiles
  - Other similar repositories
  - Recent issues
  - Issue date
- Improve the machine learning pipeline
  - Different learners
  - Different transforms
  - Different hyperparameters

Machine learning is iterative

- *Insert chart showing how accuracy and compute time change*
- Talk about how there are many nice learners and components in ML.NET. Shows the power.

Next Steps



What's next with ML.NET

- Additional ML Tasks and Scenarios
- Deep Learning
- ONNX support
- Scale-out on Azure
- Better GUI to simplify ML tasks
- Integration with VS Tools for AI
- Language Innovation for .NET

Close on what we can actually commit to.

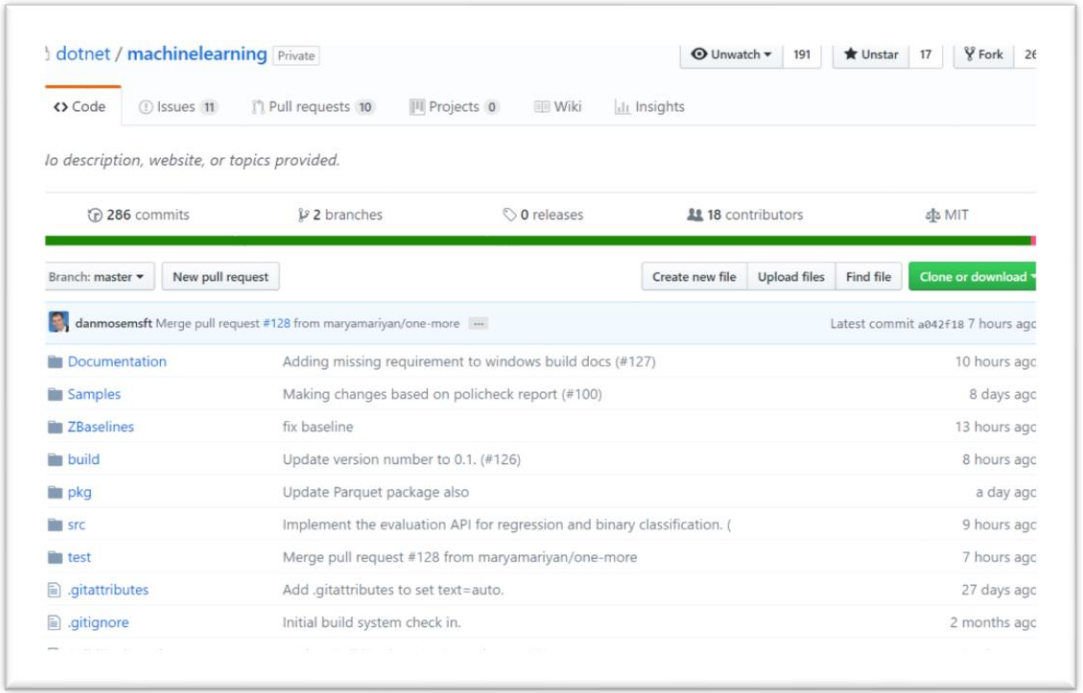
Add prettier APIs

Integration with other ML.NET toolkits

Help us make ML.NET Great!

Get Involved

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



Get Started

_____ LINK