



Az COMMUNITY

—— Conference 2022 ——

Asia's Largest Azure Community Conference

#AzConfDev



Praveen Raghuvanshi

Principal Engineer, Harman International



Applying Data Science and Machine Learning Using Microsoft .Net

Praveen Raghuvanshi

Acknowledgements



- Dataset – [Cricksheet.org](https://cricksheet.org)
- Organizers – AzConf 2022
- Sponsors

Identify Language



Maidin mhaith

Guten Morgen

おはよ ございます

शुभ प्रभात

Good Morning

Programming Languages



Data Science – Exploratory Data Analysis(EDA)

As per Wiki...

"In statistics, exploratory data analysis (EDA) is an approach of analyzing data sets to summarize their main characteristics, often with visual methods."

- Buy a House
 - No of bedrooms
 - Type (Apartment, Villa)
 - Locality
 - Crime Rate



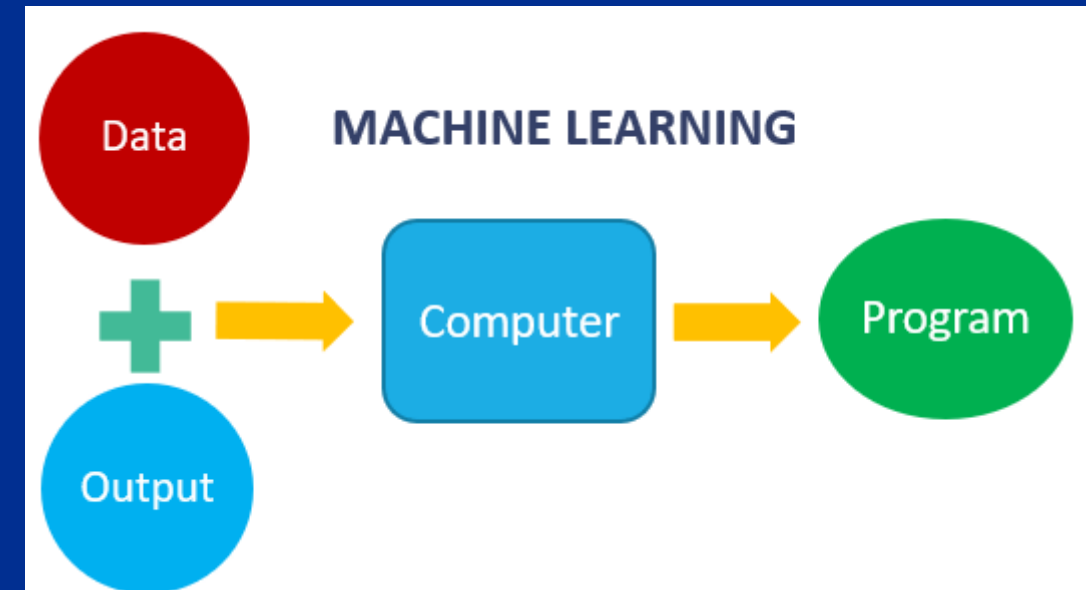
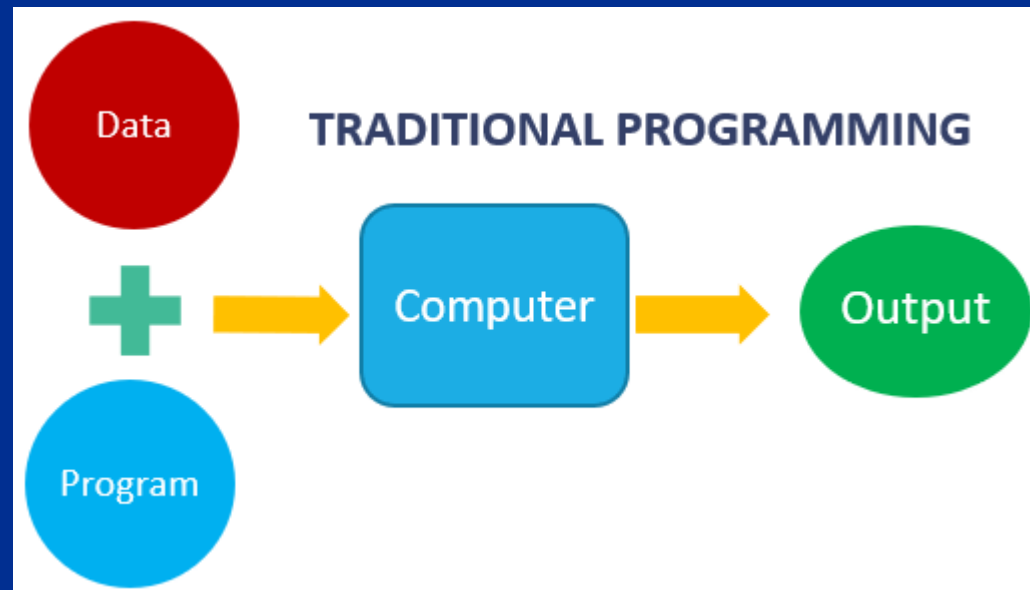
EDA Datastructure : DataFrame

	NAME	AGE	DESIGNATION
1	a	20	VP
2	b	27	CEO
3	c	35	CFO
4	d	55	VP
5	e	18	VP
6	f	21	CEO
7	g	35	MD

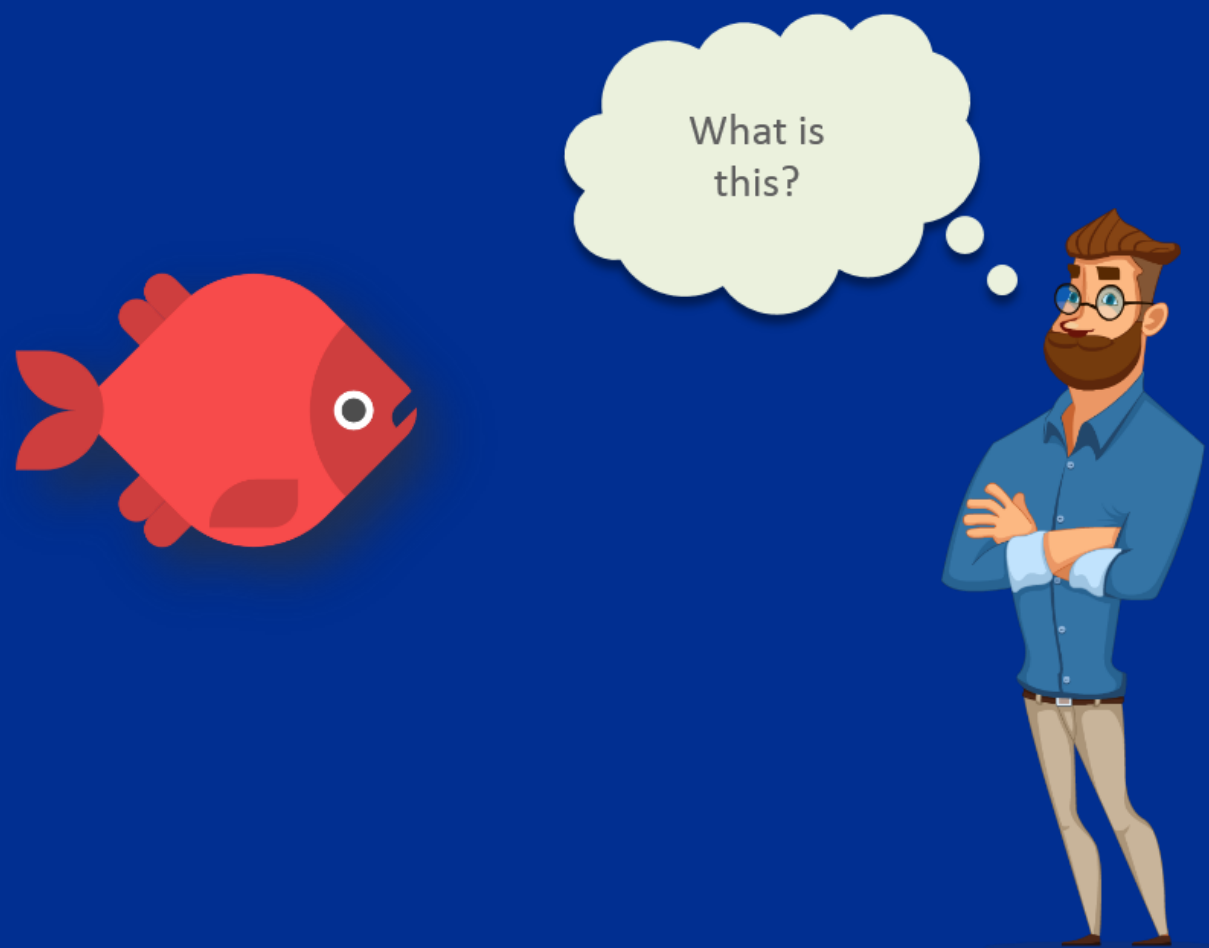
	age	designation	name
0	20	VP	a
1	27	CEO	b
2	35	CFO	c
3	55	VP	d
4	18	VP	e
5	21	CEO	f
6	35	MD	g

Machine Learning

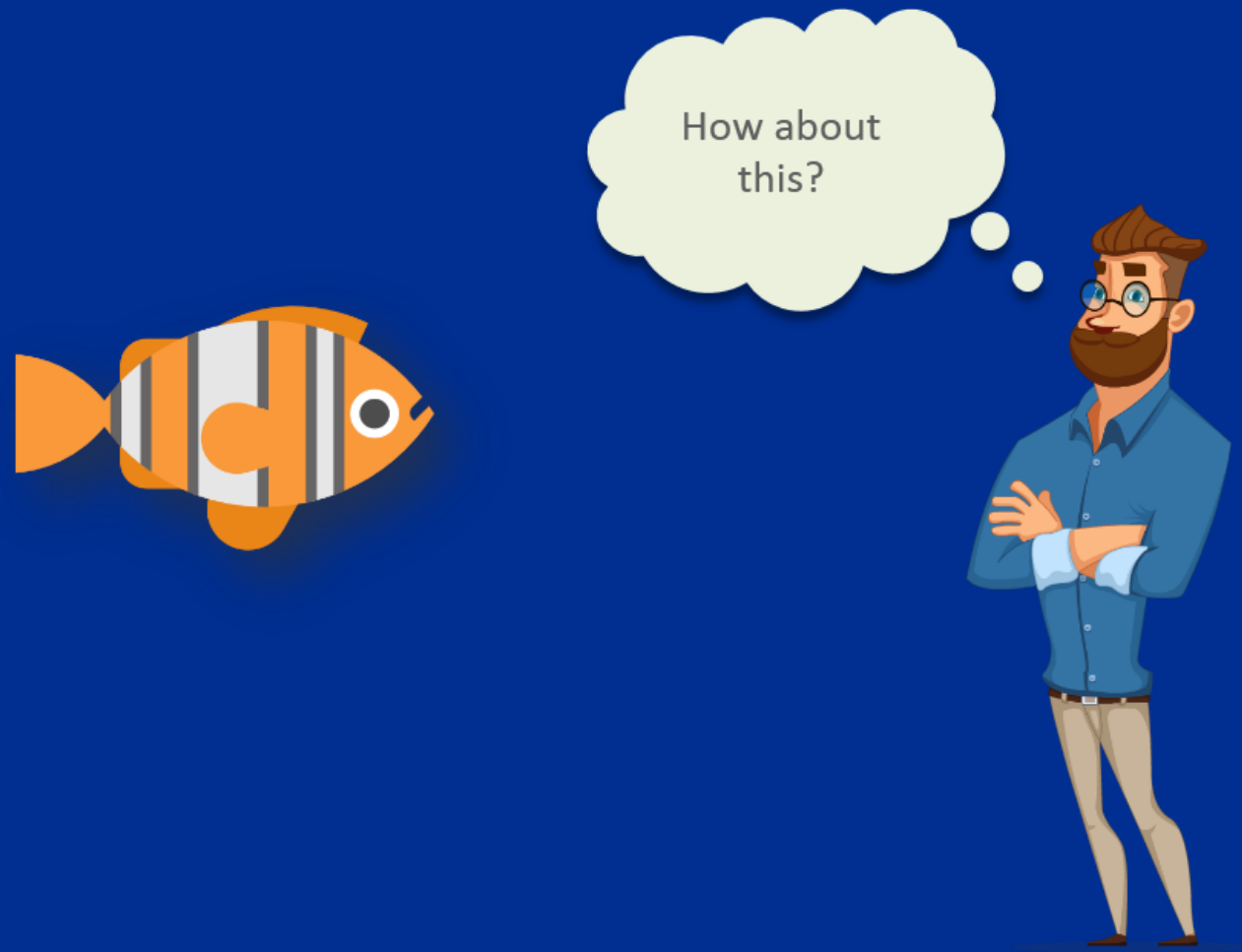
- Machine learning is getting computers to program themselves
- If programming is automation, then machine learning is automating the process of automation



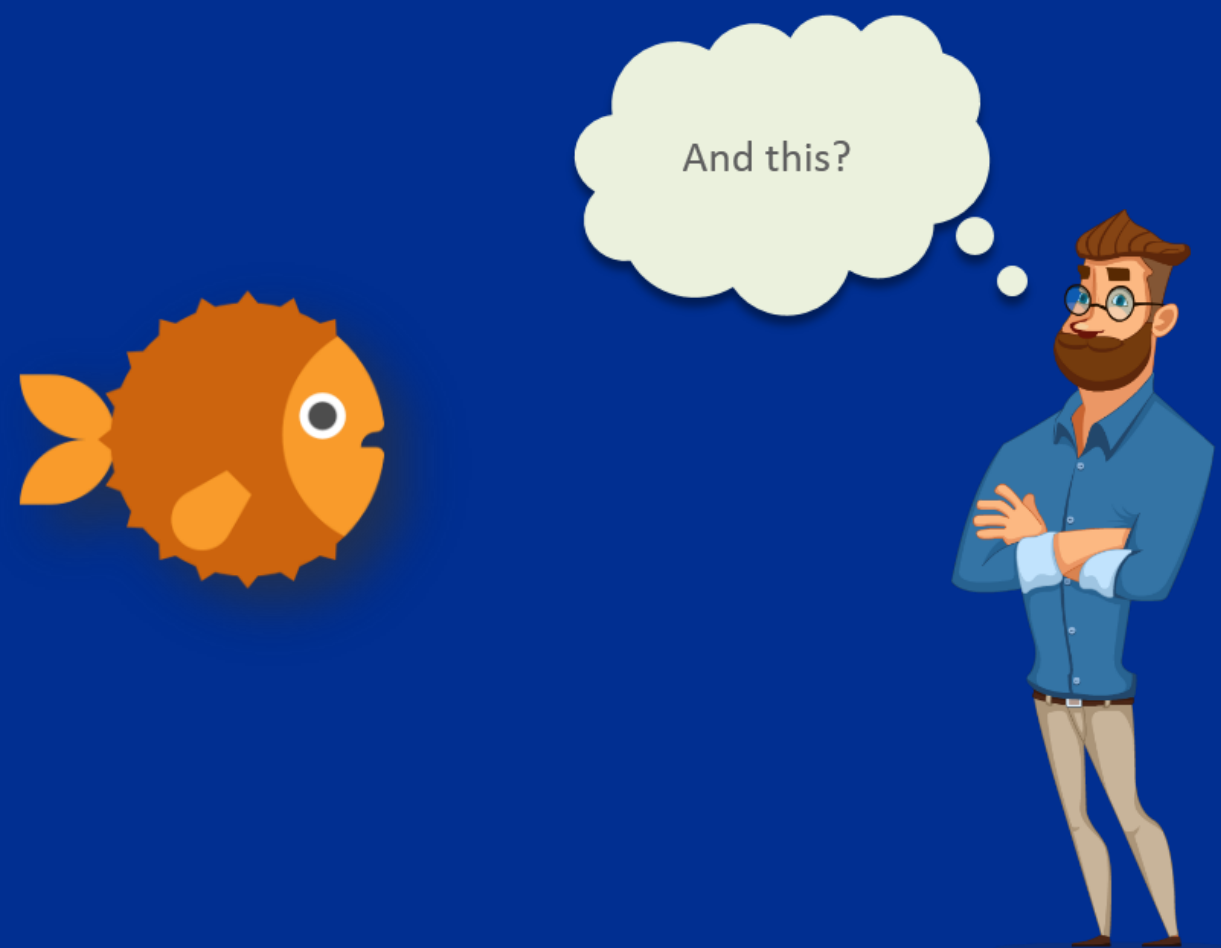
Machine Learning



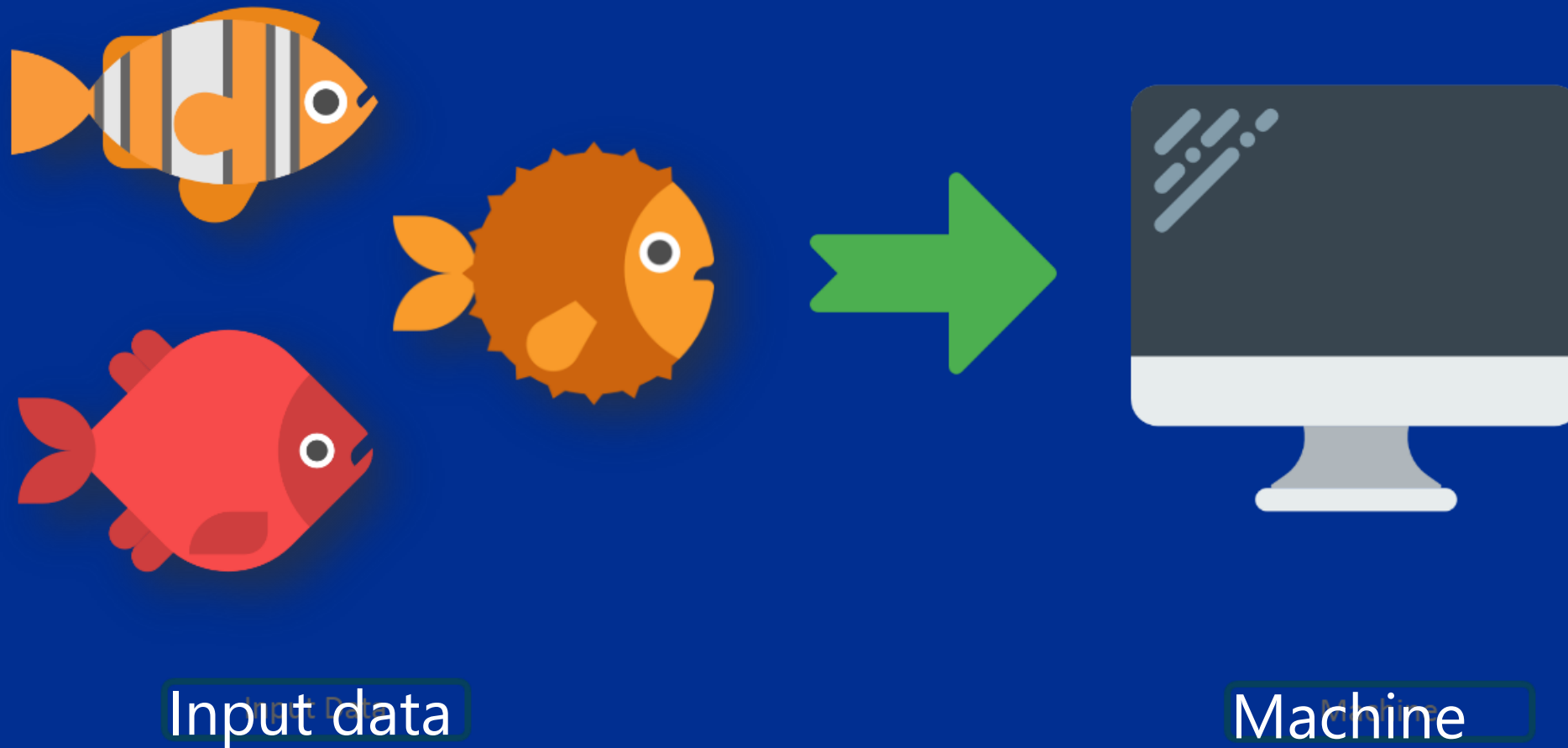
Machine Learning



Machine Learning



Machine Learning



Machine Learning



New Data

New Data

Machine

Machine

Result

Result



Pre-Requisites of ML

- A Pattern should exist
- Mathematical model/algorithm is unknown
- Abundant data

ML Frameworks



Why .NET ?



.NET ecosystem momentum

>1.5M
.NET Core developers
in Visual Studio

#1 Most Loved
Framework (2019 & 2020)

.NET Core



Top 30
Highest velocity OSS projects

github.com/dotnet
github.com/aspnet



Top 5
Language on GitHub
C#



7x
Faster than Node.js

ASP.NET Core



40%
New to .NET are students

[dot.net download survey](#)

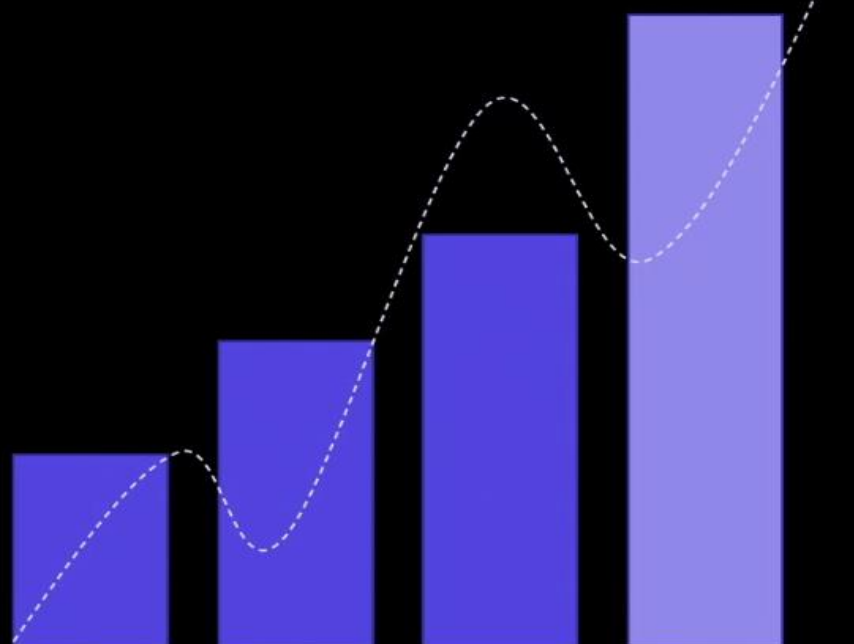
Why .NET ?



.NET adoption

5 million .NET devs

Monthly active developers in Visual Studio Family



Visual Studio

+1 million new

Monthly active .NET developers in the last year

.NET Core

+600K new

Monthly active .NET Core developers in the last year

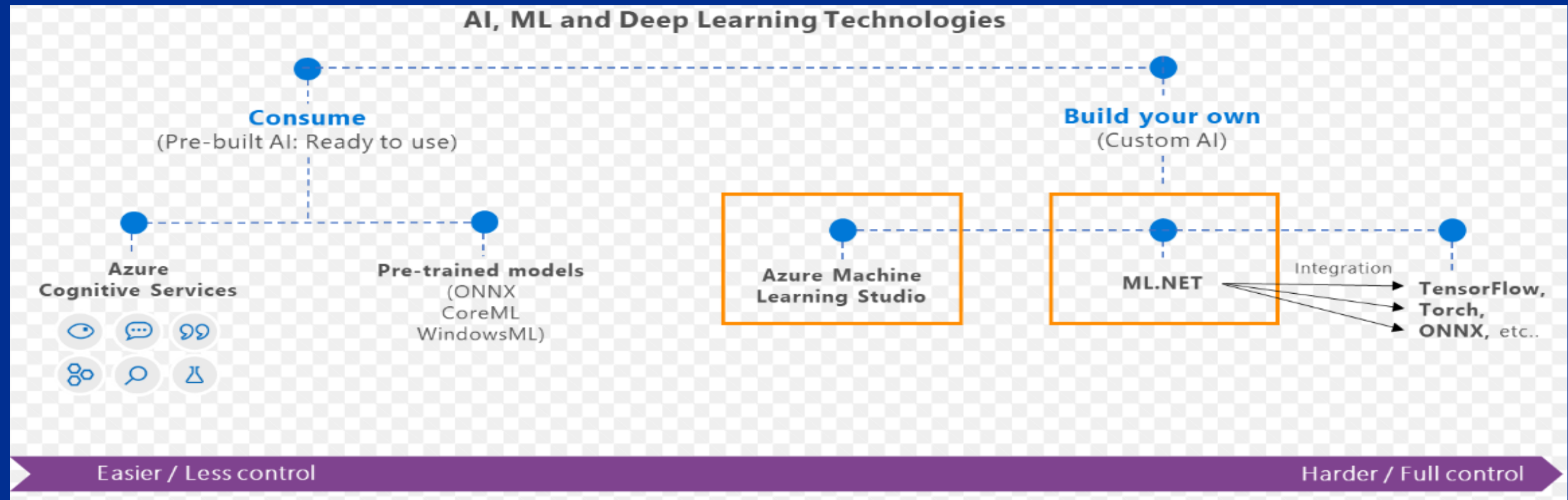
.NET on Linux

2 million

Publishes to Linux from Visual Studio

What is ML.Net ?

- ML Framework from Microsoft for developing custom AI/ML applications.
- Originated in 2002 as part of Microsoft Research Project



What is ML.Net ?

ML.NET

Machine Learning framework made for .NET developers



Build-your-own

Build your own custom models by writing C# or F# code



Developer focused

ML.NET provides just the right amount of productivity and control



Extensible

Tap into other machine learning toolkits with the rich extensibility model like TensorFlow



Proven

ML.NET has been used internally in products like Office and Bing for years






Open source and Cross-platform

Runs on Windows, macOS and Linux and developed in the open on GitHub

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

ML.Net – Proven at Scale, Enterprise ready

A list of five applications using ML.NET, each with an icon in a circle on the left and a grey arrow pointing right containing the application name and its use case. The applications are: Bing Ads (Ad Predictions), Excel (Chart Recommendations), Power Point (Design Ideas), Windows 10 (Windows Defender), and Azure Stream Analytics (Anomaly Detection). A "+ more" link is at the bottom right.

-  **Bing Ads** (Ad Predictions)
-  **Excel** (Chart Recommendations)
-  **Power Point** (Design Ideas)
-  **Windows 10** (Windows Defender)
-  **Azure Stream Analytics** (Anomaly Detection) [+ more](#)

ML.Net – Possibilities



Sentiment Analysis



Forecasting



Issue Classification



Predictive maintenance



Image classification



Recommendations



Object detection

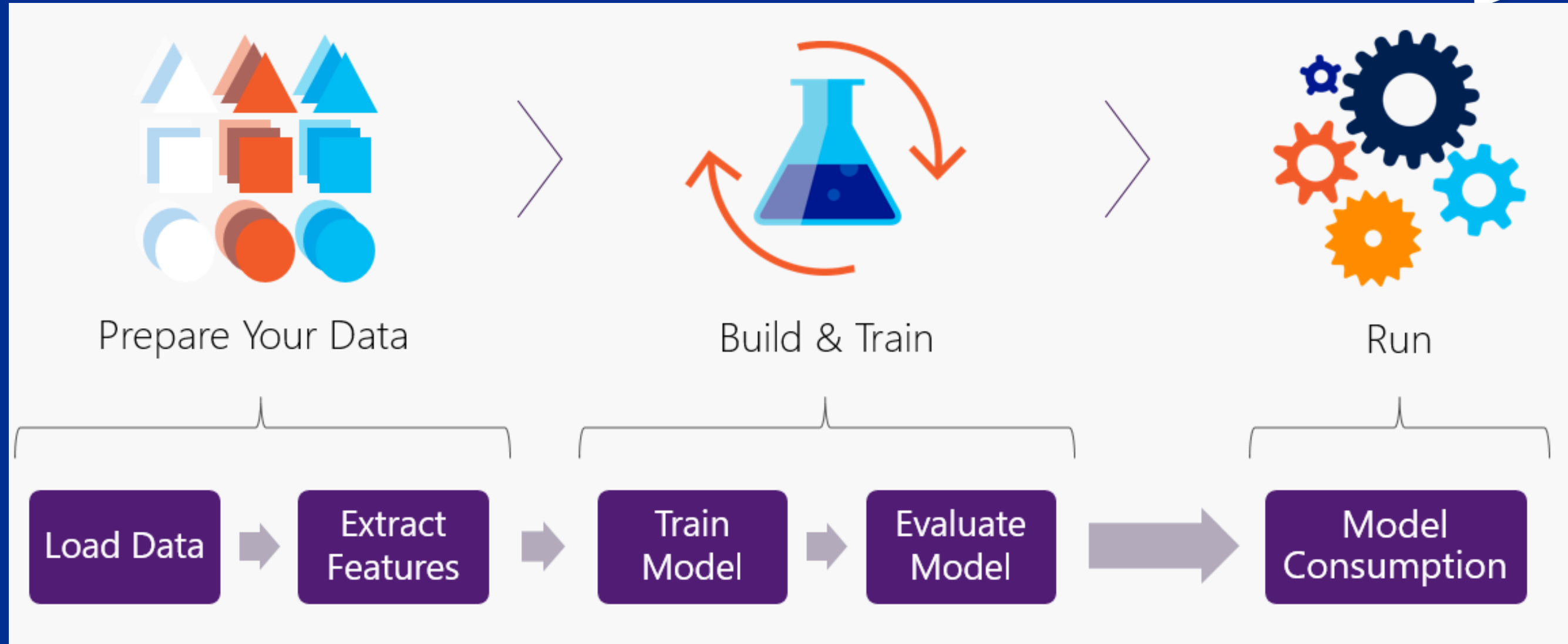


Customer segmentation

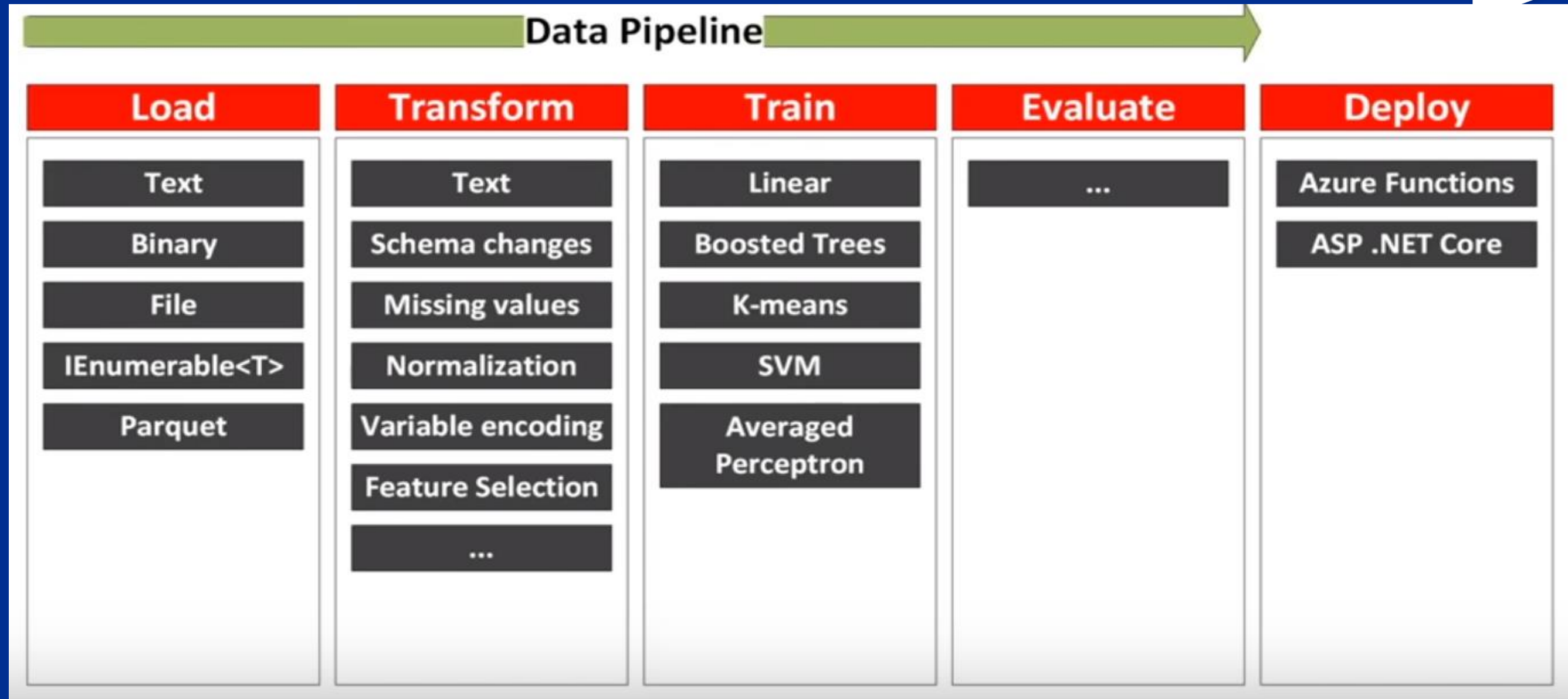


And more! Samples @ <https://github.com/dotnet/machinelearning-samples>

ML.Net – Workflow



ML.Net – Data Pipeline



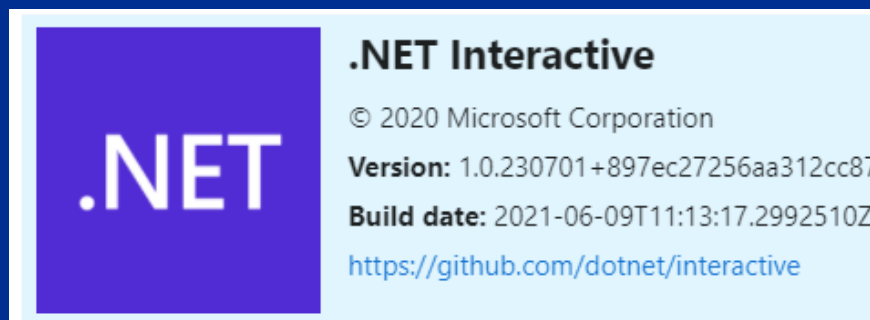
Cricket – EDA and Prediction

Statistics(2019-2020)



- Played by 104 Nations
- IPL Valuation: \$6.7 billion
- IPL Viewers : 370 Million

Tools and Frameworks



Cricket : Problem Statement

- Perform analysis on cricket dataset
- Predict score on a specific ball within 6 overs



Dataset

T20 Match - Men

- Duration: 2017 – 2021
- Matches : 1010
- Teams: 56
- Columns: 22
- Records: 231 K
- Mix of number and strings

- match_id
- season
- start_date
- venue
- innings
- ball
- batting_team
- bowling_team
- striker
- non_striker
- bowler
- runs_off_bat
- extras
- wides
- noballs
- byes
- legbyes
- penalty
- wicket_type
- player_dismissed
- other_wicket_type
- other_player_dismissed



Data Cleaning

- Filter dataset to include records till 6 over.
Low memory and fast execution
- Check for Null Values
- Aggregation : Score per ball →
runs_off_bat + extras
- Cumulative sum : Total Score per ball
- Remove features/columns

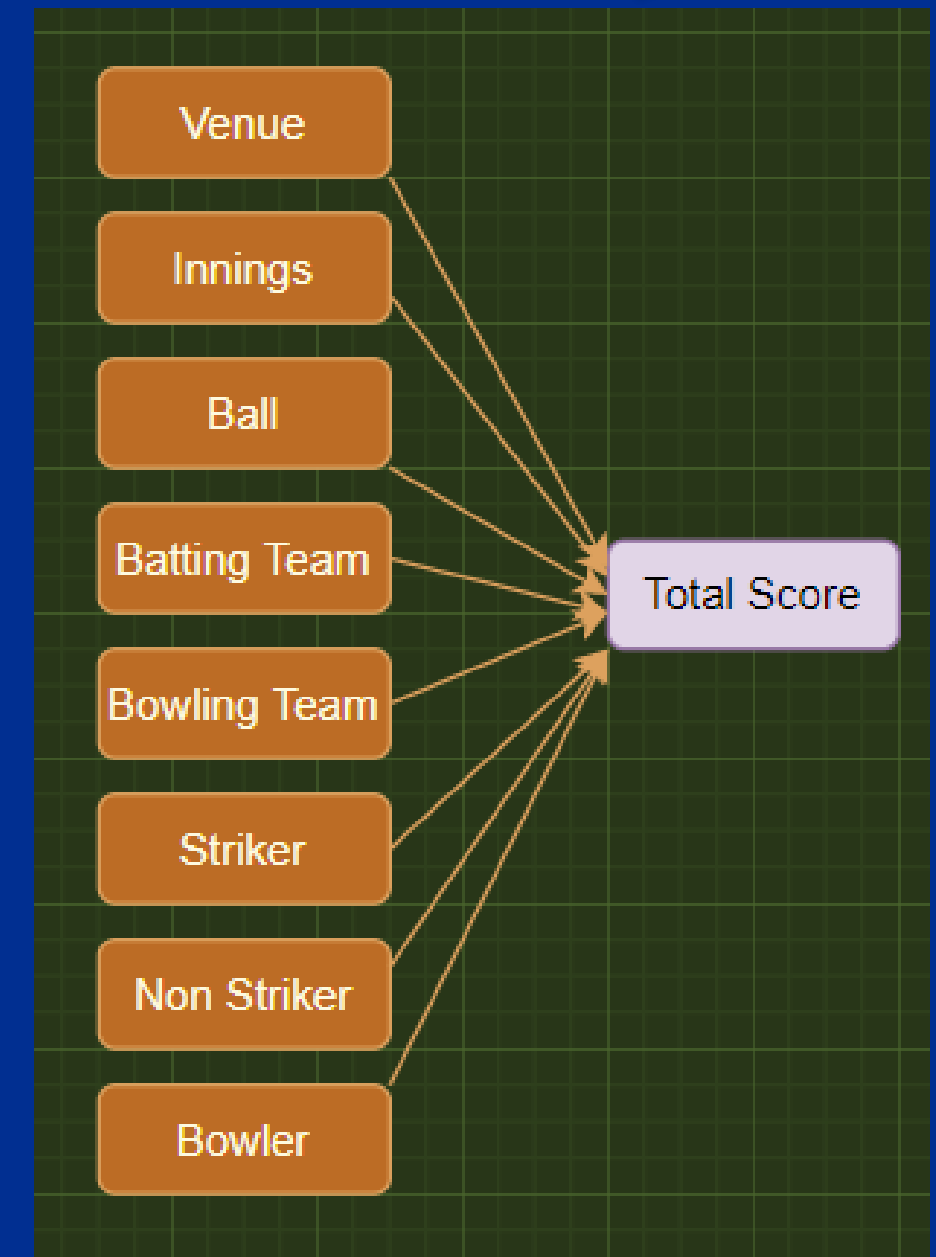
- match_id
- season
- start_date
- venue
- innings
- ball
- batting_team
- bowling_team
- striker
- non_striker
- bowler
- runs_off_bat
- extras
- wides
- noballs
- byes
- legbyes
- penalty
- wicket_type
- player_dismissed
- other_wicket_type
- other_player_dismissed



Prediction

Regression using ML.Net

- Define Classes : Match, MatchScorePrediction
- Load Dataset
- Split Dataset: Train/Test : 80/20
- One Hot Encoding
- Model Algorithm : FastTree
- Train → Evaluate → Predict the Model



ML.Net – API (Jupyter Notebook)

```
*****
*      Model quality metrics evaluation
*      -----
*      RSquared Score:      0.82
*      Root Mean Squared Error: 6.73

***** Predict...
Match Info:

Venue: Vidarbha Cricket Association Stadium_ Jamtha
Batting Team: India
Bowling Team: New Zealand
Inning: 1
Ball: 3.4
Striker: V Kohli
Non-Striker: Yuvraj Singh
Bowler: CJ Anderson
^^^^^ Prediction: 24.15512

*****
Predicted score: 24.1551, actual score: 20
*****
```

ML.Net – AutoML (Model Builder)



Top 3 models explored							
	Trainer	RSquared	Absolute-loss	Squared-loss	RMS-loss	Duration	#Iteration
1	LightGbmRegression	0.9012	3.72	25.48	5.05	6.7	1
2	SdcaRegression	0.8234	5.09	45.56	6.75	4.0	2
3	FastTreeRegression	0.8213	5.08	46.09	6.79	4.7	3

ML.Net – Web App



AI Endeavour Home Cricket Prediction



Cricket Score Prediction

Select Venue ▼

Rajiv Gandhi International Stadium

Select Ball ▼

2

Select Non Striker ▼

MS Dhoni



Predict Score

15.08

Select Inning ▼

2

Select Batting Team ▼

Kolkata Knight Riders

Select Bowling Team ▲

Rajasthan Royals

Select Over ▼

5

Select Striker ▼

JDP Oram

Select Bowler ▼

SK Raina



Reset

Made with ❤ in India. © Copyright 2022



Demo

#AzConfDev

ML.Net : Customer Stories



Asgard Systems

Asgard Systems uses demand forecasting in grocery stores to reduce food waste and gas house emissions.

[Learn more >](#)



Scancam

Scancam uses ML.NET to detect vehicles at fuel station pumps and provides alerts for known offenders who previously drove off without paying for their fuel.

[Learn more >](#)



SigParser

SigParser converts e-mail signatures to contacts and eliminates manual data entry; it uses ML.NET to predict if an e-mail sender is human or an automated system.

[Learn more >](#)



endjin

endjin uses ML.NET with AutoML to improve the process of classifying articles for their Azure newsletter and to revolutionize simple, everyday tasks.

[Learn more >](#)



Microsoft Real Estate & Security

Microsoft Real Estate & Security uses ML.NET to detect and classify HVAC system faults on Microsoft's campus and convert them to work orders.

[Learn more >](#)

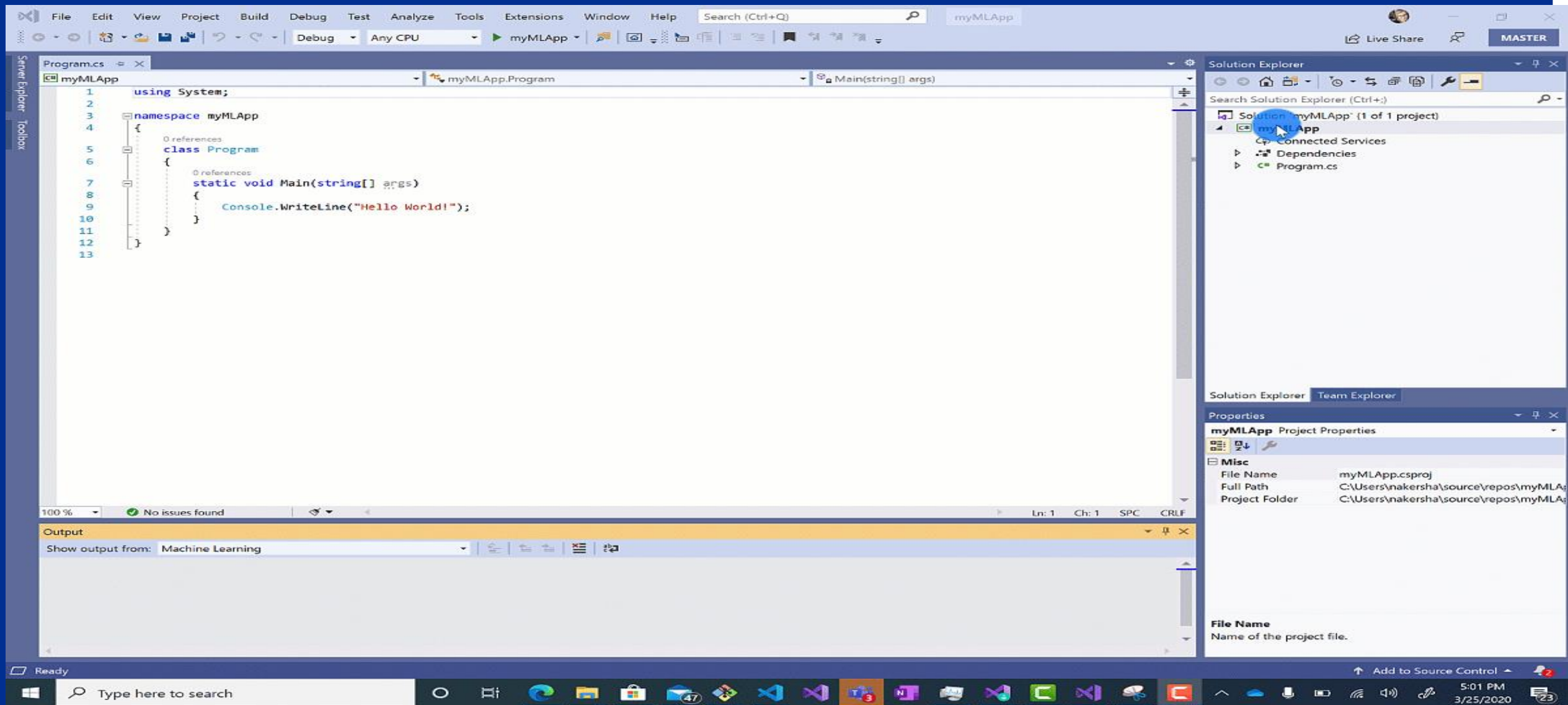


Power BI

Power BI uses ML.NET to help users identify key influencers and customer segments so that they can understand the factors that drive their business metrics.

[Learn more >](#)

ML.Net : Model Builder(AutoML)







References

- ML.Net : <https://dotnet.microsoft.com/apps/machinelearning-ai/ml-dotnet>
- ML.Net: <https://devblogs.microsoft.com/cesardelatorre/what-is-ml-net-1-0-machine-learning-for-net/>
- ONNX : <https://onnx.ai/>
- Photo-Search (ONNX) : <https://github.com/Tak-Au/Photo-Search>
- Music Repair : <https://www.youtube.com/watch?v=nnV-1q-z9uE>
- ML Cookbook : <https://github.com/dotnet/machinelearning/blob/master/docs/code/MlNetCookBook.md>
- Deploy to Azure functions : <http://luisquintanilla.me/2018/08/21/serverless-machine-learning-mlnet-azure-functions/>
- <https://rubikscore.net/2019/02/18/ultimate-guide-to-machine-learning-with-ml-net/>
- <https://www.youtube.com/watch?v=dojO4zEL9sg>
- <https://www.youtube.com/watch?v=zy7Y9CHji2k>

Resource



- Github: <https://github.com/praveenraghuvanshi/tech-sessions/tree/master/21102022-AI-ML-using-ml-dotnet-azconf-2022>
- Dev.to : Cricket Analysis and Prediction using ML.Net(C#) - DEV Community  
- Short Url: <https://bit.ly/3MH34hv>



Platinum Partner



Gold Partner



Silver Partner



Q & A



Applying Data Science and Machine Learning using Microsoft .Net

Praveen Raghuvanshi

@praveenraghuvan

