Using Azure Machine Learning in Data Science

Muaz Hrustanbegović
MsC

Microsoft

Using Azure Machine Learning in Data Science

Muaz Hrustanbegović

platinum sponzor

LOGOSÓFT

telekomunikacijski sponzor

oficijelni brend konferencije

HT ERONET

Lenovo.™

gold sponzori

Addiko Bank

AUTHORITY PARTNERS

bh

COMTRADE DISTRIBUTION

EPSON EXCEED YOUR VISION

KimTec

LANACO INFORMACIONE TEHNOLOGIJE

nsoft

PROINTER IT SOLUTIONS AND SERVICES

SEMOS we can give you everything you need

sys company

TW teamwork solution provider

veeAM

silver sponzori

VW Volkswagen

PORSCHE SARAJEVO

Audi

APP IMPACT impacting your business

INFODOM

mistral because it matters

PHILIPS Televizori

prijatelji konferencije

zvanično craft pivo

communis tele communication

FANFAN

GALAKTIKA

PROEVENT

SARAJEVO BUSINESS FORUM '18

UNIQA

WA WeAreDevelopers

BREW

medijski sponzori

akta

ALJAZEERA BALKANS

20a atvbl.com

Banke & Biznis

Banjaluka.com prvi banjalučki portal

BHRT

hicsak.info br. Internet magazin

business

Dnevni list vaša dnevna novina

FAKTOR

FBL

FENA FEDERAL NEWS AGENCY

hayat.ba

HAYAT | HD

INFO www.info.ba

Centar za razvoj i afirmaciju kulture i obrazovanja Litteratus

poduzetnik.ba

racunalo www.racunalo.com

REUNION

RSG RADIO

STUDENT

STUDOMAT

TERMINAL

About Me

MUAZ HRUSTANBEGOVIĆ

SENIOR DEVELOPER

Experienced BI Developer (MS Stack)
Data Engineer/Data science
Machine Learning
Multiple MSNetwork speaker

Using Azure Machine Learning in Data Science

@ Galerija | 45 min

MAKE IT cLoud 📣 NetWork 乙 mıstral

Faculty of Electrical Engineering Sarajevo
Originally, BI developer and data geek
Working with Data Science and ML recently

Goals

Introduction to ML and Data Science

Learn how to create and consume Azure ML experiment

Try to solve real world problem using Data Science approach

Agenda

Introduction and history of ML

Data Science - Lifecycle

Where to use ML?

Azure ML

Fancy Demo

Create and consume ML experiment

Conclusion

What is Machine Learning?

-**Machine learning** is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. **Machine learning** focuses on the development of computer programs that can access data and use it learn for themselves.

-**Machine learning** is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

History of Machine Learning

- Blaise Pascal made an "arithmetic machine" - 1642
- Alan Turing conceives his "Universal Machine„ - 1936
- Neural Network - 1943
- Lots of Sci-Fi
- Deep Blue beat Garry Kasparov - 1997
- Reboot – 2006
- Google's AlphaGo - 2016

Data Science

Data science, also known as data-driven science, is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms, either structured or unstructured, similar to datamining.

Data Science includes following roles:

- Data Scientist
- Data Analyst
- Data Architect
- Data Engineer
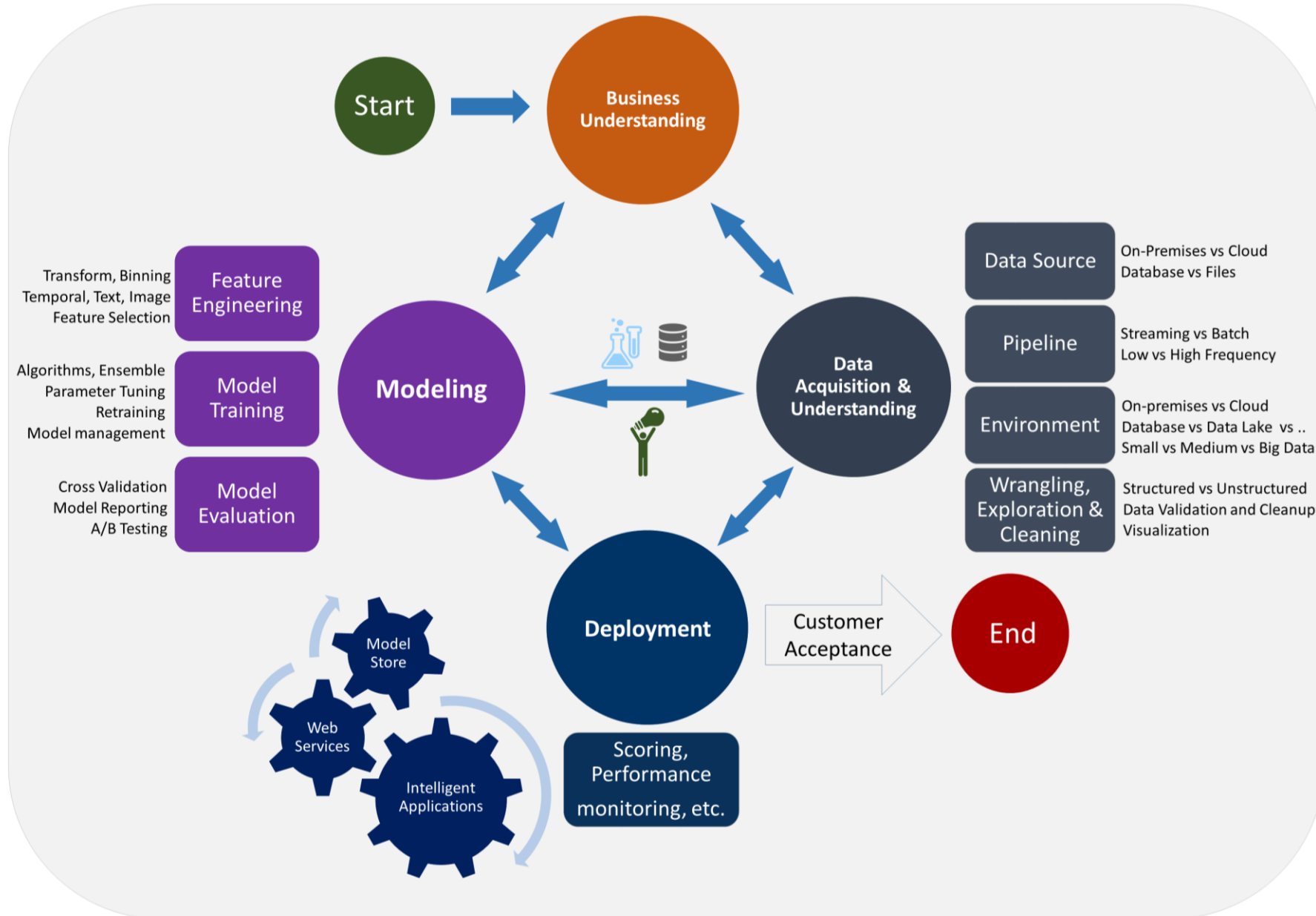- Business Analyst
- Data and Analytics Manager

Data Science Workflow

- Hypothesis
- Data Selection and Preprocessing
- Data Splitting
- Feature Selection and Feature Engineering
- Model Selection
- Train, evaluate and score model
- Again?

Data Science Workflow

Machine Learning – Learning types

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs

Unsupervised learning is the machine learning task of inferring a function to describe hidden structure from "unlabeled" data.

Reinforcement learning is an area of machine learning inspired by behaviourist psychology[citation needed], concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

Machine Learning - Alghoritms

- Regression - Forecast the future by estimating the relationship between variables.

- Anomaly Detection - Identify and predict rare or unusual data points.

- Clustering - Separate similar data points into intuitive groups.

- Classification - Identify what category new information belongs in

- Recomender – Associations between cases

Where to use ML?

- Health

- Finance

- Insurance

- Social Media

- Marketing
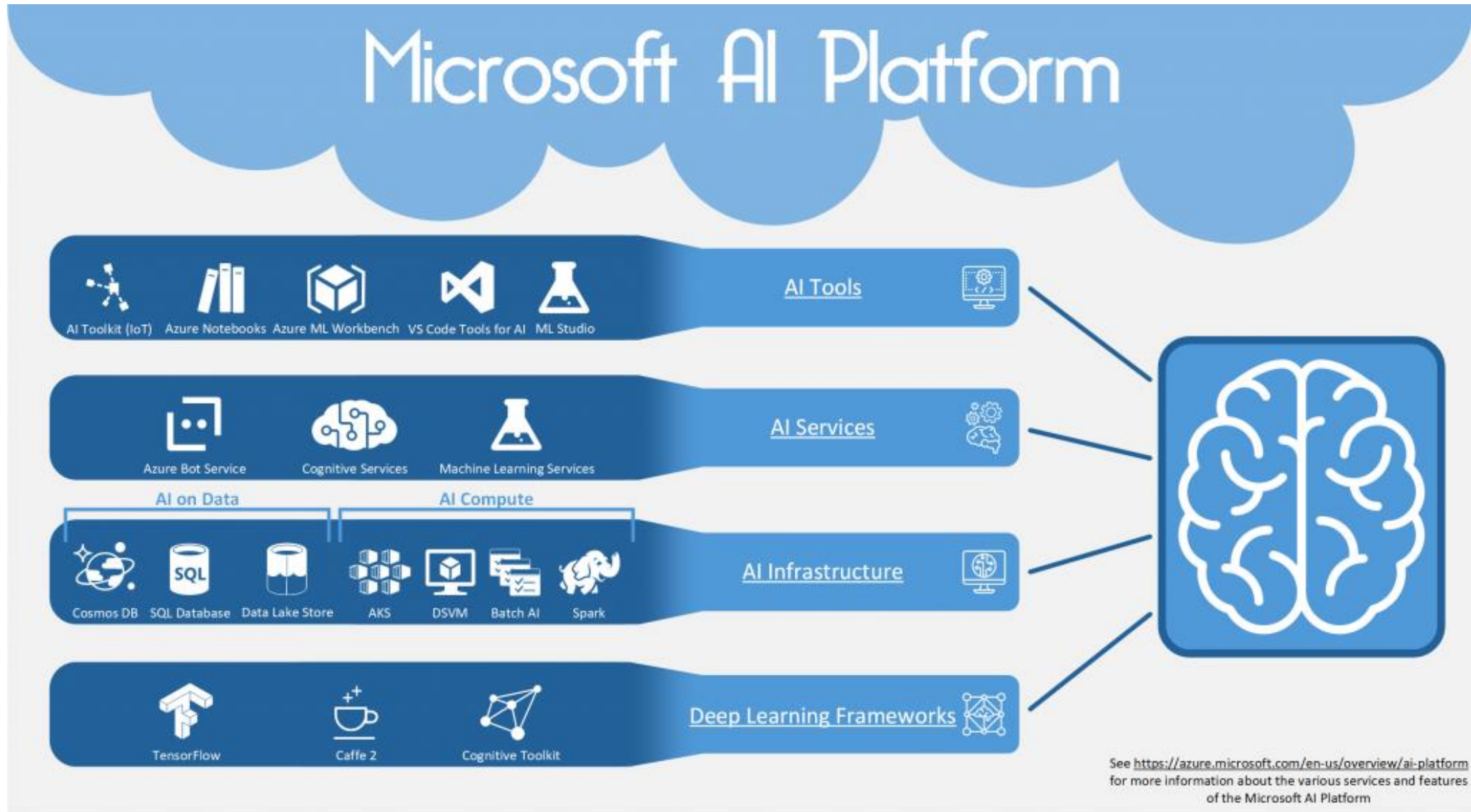
- Where we used it?

Common issues in Data Science

- Where is the data?

- Overfitting/Underfitting

- High level of abstraction

- Err... How to use this?

- Project failures

- Lack of Data Science education.

- It is all about Culture

What Microsoft offers?

What Microsoft offers?

Microsoft AI Platform

AI Tools
AI Toolkit (IoT), Azure Notebooks, Azure ML Workbench, VS Code Tools for AI, ML Studio

AI Services
Azure Bot Service, Cognitive Services, Machine Learning Services

AI on Data — **AI Compute**

AI Infrastructure
Cosmos DB, SQL Database, Data Lake Store, AKS, DSVM, Batch AI, Spark

Deep Learning Frameworks
TensorFlow, Caffe 2, Cognitive Toolkit

See https://azure.microsoft.com/en-us/overview/ai-platform
for more information about the various services and features
of the Microsoft AI Platform

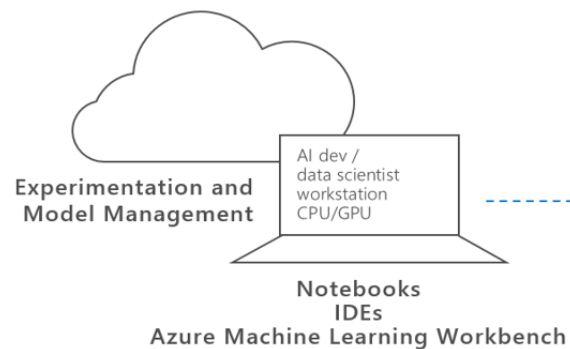# What Microsoft offers - Azure?

ML only

Azure ML Services - preview

Integrated, end-to-end data science and advanced analytics solution. It enables data scientists to prepare data, develop experiments, and deploy models at cloud scale.

AZURE MACHINE LEARNING

AZURE MACHINE LEARNING SERVICES

Experimentation and Model Management

AI dev / data scientist workstation CPU/GPU

Notebooks
IDEs
Azure Machine Learning Workbench

TRAIN & DEPLOY OPTIONS

AZURE

Spark
SQL Server
Virtual machines
GPUs
Container services

ON-PREMISES

SQL Server
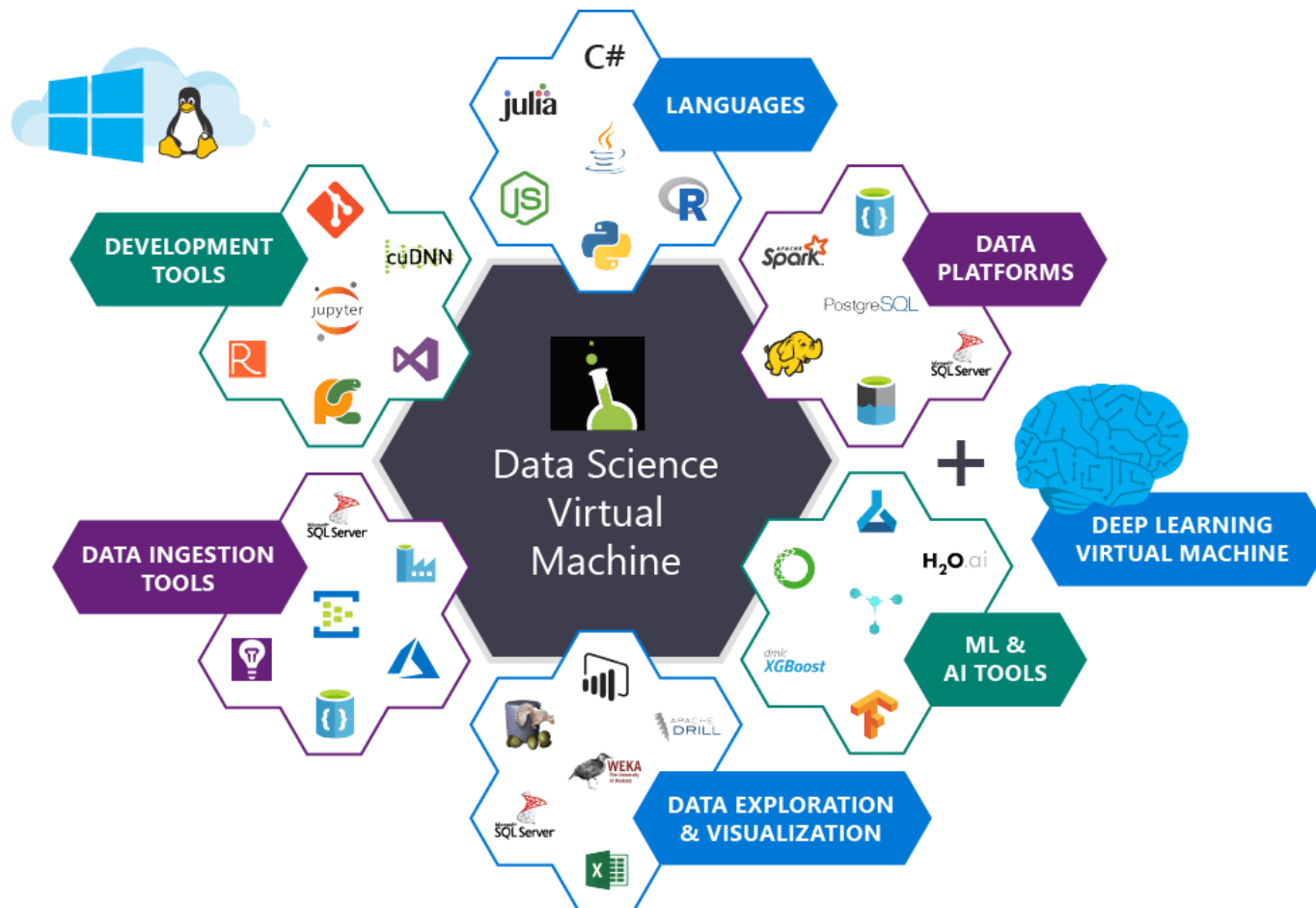Machine Learning Server

EDGE COMPUTING

Azure IoT Edge

Azure Data Science VM

Customized VM images on Azure, loaded with data science tools used to build intelligent applications for advanced analytics.

Windows and Linux editions

Azure ML Studio

Azure Machine Learning Studio is a collaborative, drag-and-drop tool you can use to build, test, and deploy predictive analytics solutions on your data.

- PROJECTS - Collections of experiments, datasets, notebooks, and other resources representing a single project
- EXPERIMENTS - Experiments that you have created and run or saved as drafts
- WEB SERVICES - Web services that you have deployed from your experiments
- NOTEBOOKS - Jupyter notebooks that you have created
- DATASETS - Datasets that you have uploaded into Studio
- TRAINED MODELS - Models that you have trained in experiments and saved in Studio
- SETTINGS - A collection of settings that you can use to configure your account and resources.

Azure ML – Sources and Data Formats 🧪

- Azure Blob Storage
- Azure SQL DB
- Azure SQL DW*
- Azure Table
- Desktop Direct Upload
- Hadoop Hive Query
- Manual Data Entry
- OData Feed
- On-prem SQL Server*
- Web URL (HTTP)

- ARFF
- CSV
- SVMLight
- TSV
- Excel
- ZIP

Azure ML – Experiment and Modules

Experiment are datasets and analytical modules, which you connect together to construct a predictive analysis model.

Create an experiment from scratch, or you can use an existing sample experiment

https://gallery.azure.ai/

A module is an algorithm that you can perform on your data.

Azure ML – Python Module

Azure Machine Learning Studio uses the Anaconda distribution.

Upload a ZIP file containing Python code to your workspace

Read, load, and manipulate data from sources not supported by the [Import Data](#) module

Limitations:
- All files saved locally are isolated and deleted once the module finishes.
- Can only access current directory and its sub-directories
- It is not possible to return trained models directly back to the Azure Machine Learning runtime

Azure ML – R Module

Extend your experiment with custom R script

Accepts multiple input datasets and yields a single dataset as output

Unsupported packages:
- The package has a Java dependency.
- The package binaries are not compatible with the sandboxed Azure environment.
- The package requires direct Internet access, or network access.

List of supported packages:
https://docs.microsoft.com/en-us/azure/
machine-learning/studio-module-reference/
r-packages-supported-by-azure-machine-learning

Azure ML – Web Service

Deploy an Azure Machine Learning predictive model as a Web service.

Two types of services:

- Request-Response Service (RRS)
- Batch Execution Service (BES)

Call web services directly from Excel

Predicting BTC prices

Or how to make millions by investing in cryptocurrencies... Before reality kicks in.

Demo 1
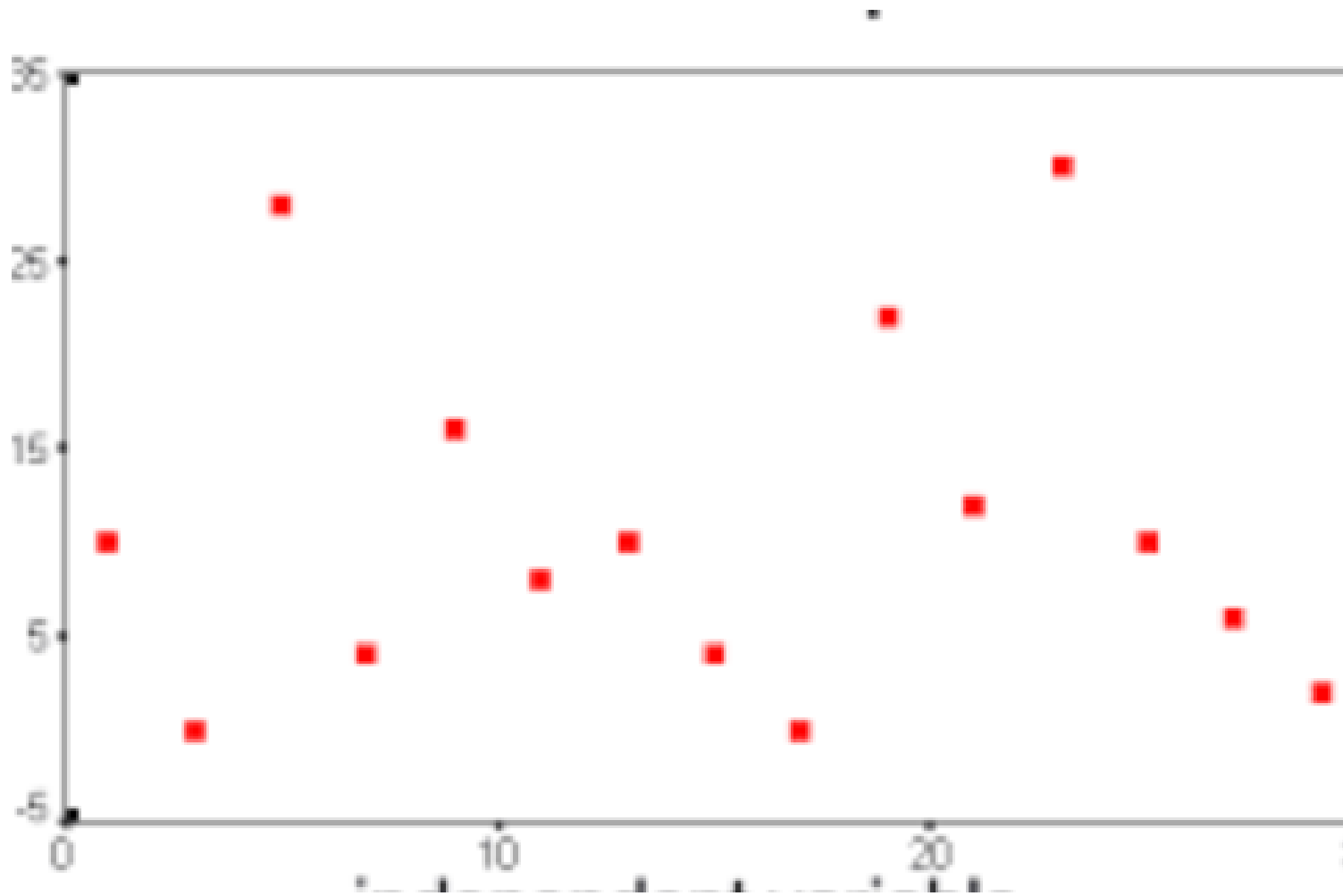Predictin BTC prices using Linear Regression

Linear Regression

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$
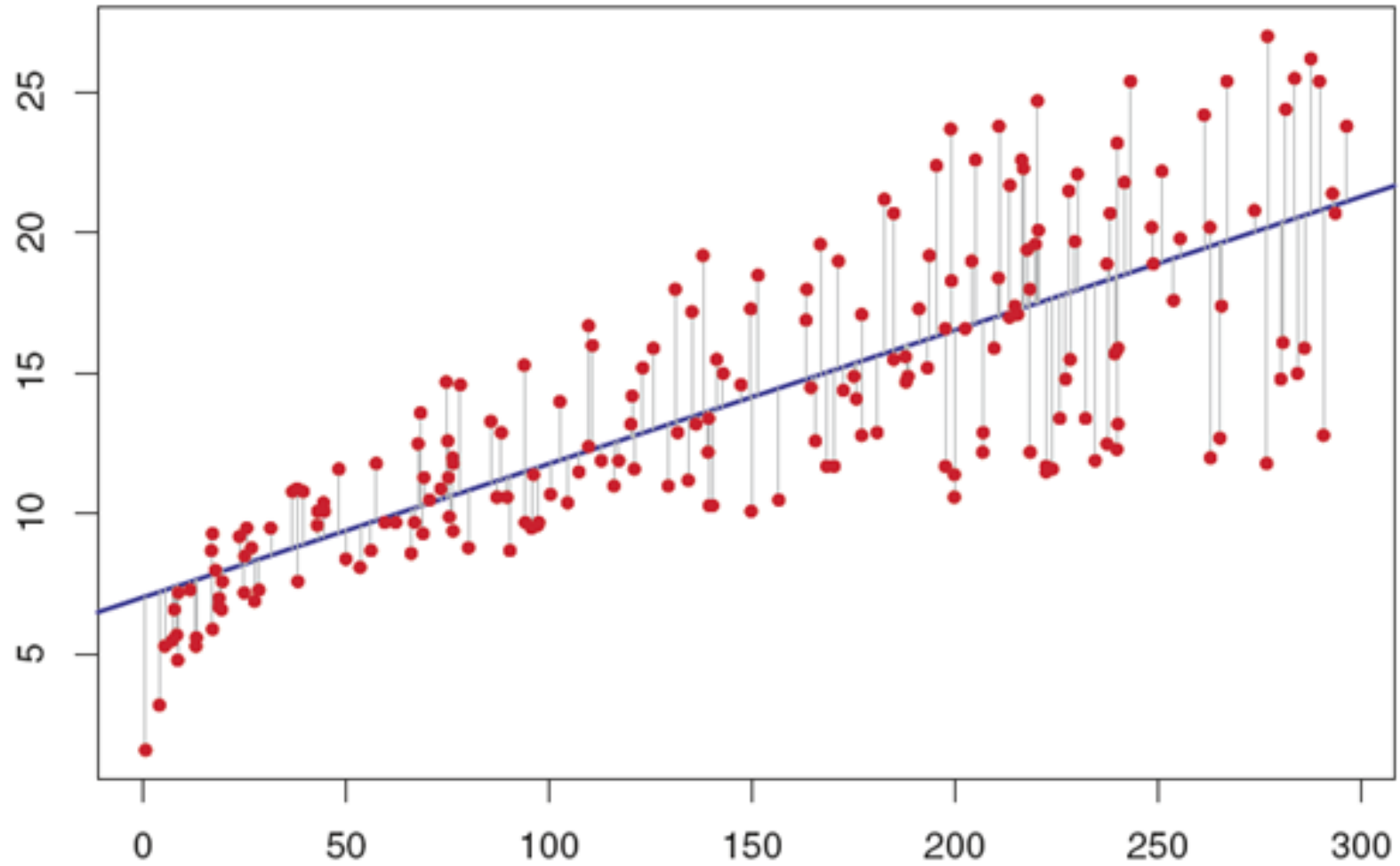

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

$$X = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Linear Regression

Linear Regression

Demo 2

Consume prediction via Excel

Further?

Sentiment Analysis of #BTC twitter data

Pricing - Studio

	FREE	STANDARD
Price	Free	$9.99 per seat per month $1 per studio experimentation hour
Azure subscription	Not required	Required
Max number of modules per experiment	100	Unlimited
Max experiment duration	1 hour per experiment	Up to 7 days per experiment with a maximum of 24 hours per module
Max storage space	10 GB	Unlimited - BYO
Read data from On-Premises SQL [Preview]	No	Yes
Execution/performance	Single node	Multiple nodes
Production Web API	No	Yes
SLA	No	Yes

Pricing - Production Web API

	DEV/TEST*	STANDARD S1	STANDARD S2	STANDARD S3
Tier Price per month	$0	$100.13	$1,000.06	$9,999.98
Features				
Included transactions (per month)	1,000	100,000	2,000,000	50,000,000
Included compute hours (per month)	2	25	500	12,500
Total number of web services[1]	2	10	100	500
Overage rates	N/A	$0.50 per 1,000 transactions $2 per API compute hour	$0.25 per 1,000 transactions $1.50 per API compute hour	$0.10 per 1,000 transactions $1 per API compute hour

* API usage restrictions apply on the testing tier—Limited to two concurrent RRS calls.

[1] Number of web services customer can associate with a plan at any given point of time.

Will this make me a
Data Sciencitst?

Why use it?

Easy to learn

Easy to use

Easy to share

Degree of customization

Mature product

Web Service

Again, Web Service

Q&A?

Panel diskusija: Azure Cloud & DevOps by Mistral Technologies