



Machine Learning v IoE a IoT

Inteligentne kyberfyzikálne systémy

Praktické cvičenie č. 8. – Machine Learning v IoE a IoT.

Informačný list pre pedagóga**Číslo a názov cvičenia**

8. Machine Learning v IoE a IoT

Cieľ cvičenia

Cvičenie ma poukázať na možnosti využitia strojového učenia (Machine Learning) vo svete IoE a IoT. Z celého komplexu algoritmov strojového učenia si vyberieme algoritmy určené pre predikčnú analýzu dát. Algoritmy strojového učenia si vyberieme z portfólia cloudových služieb Microsoft Azure. Na príklade si ukážeme neurónovú sieť ako jeden z nástrojov predikčnej analýzy. Cvičenie rozdelíme do dvoch častí:

- teoretická časť:
 - základy objavovania znalostí,
 - základy analýzy dát,
 - základy strojového učenia,
- Praktická časť:
 - strojové učenie na cloude,
 - data set pre klasifikáciu a jeho upload a spracovanie na cloude,
 - vytvorenie klasifikačného modelu za použitia neurónových sietí,
 - vygenerovanie API key pre prístup k analýze odkiaľkoľvek (aj z Vášho Excelu) pomocou REST API (get, post, ...).

Odporúčaná forma cvičenia

Vypočutie krátkeho teoretického úvodu. Nasleduje jednotlivá práca v spolupráci s cvičiacim. Skupinové odpovede na otázky.

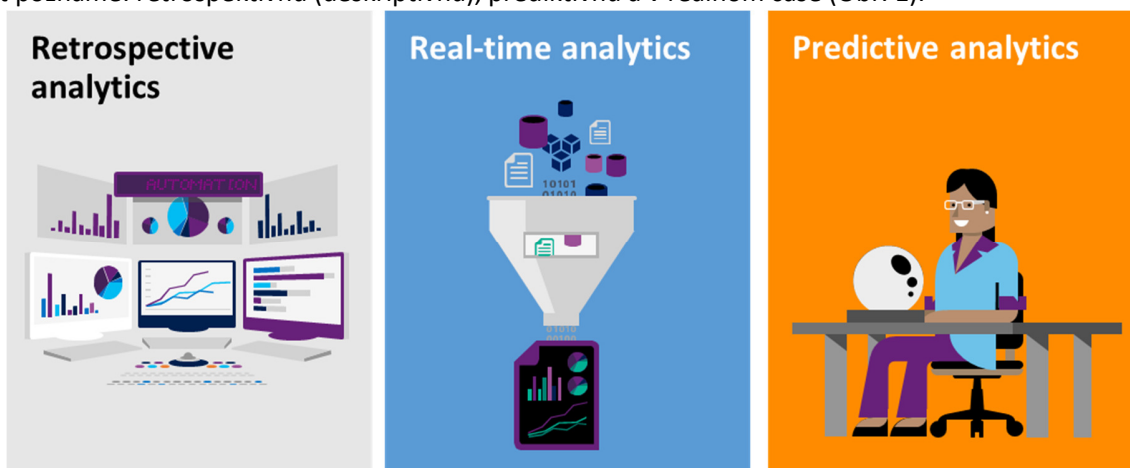
Prílohy

Súbor: pocasie.xlsx

Machine Learning v IoE a IoT

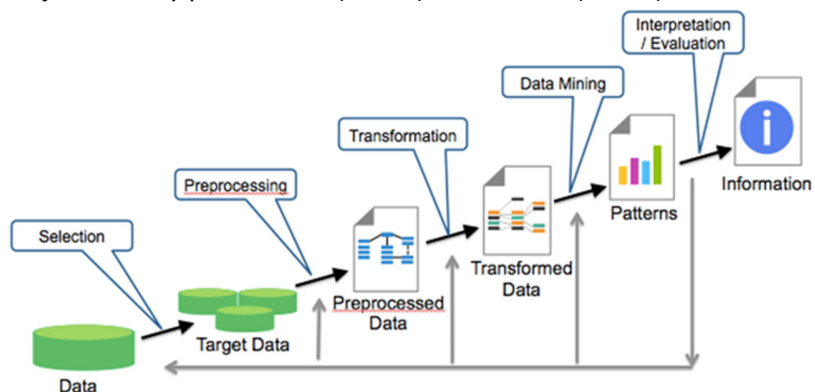
Teoretická časť (20 min)

Analýzu dát poznáme: retrospektívnu (deskriptívnu), prediktívnu a v reálnom čase (Obr. 1):

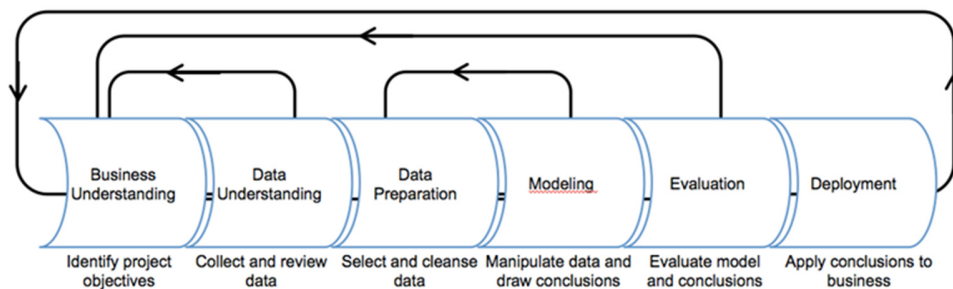


Obr. 1: Typy analýz

Proces objavovania znalostí je iteratívny proces, KDD (Obr. 2) a CRISP-DM (Obr. 3):

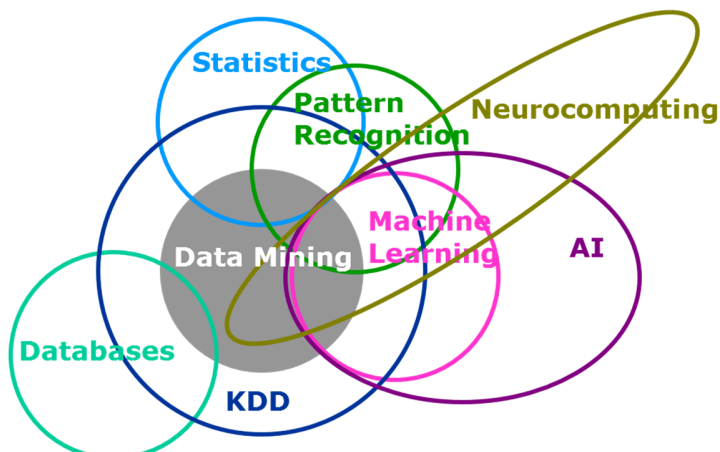


Obr. 2: KDD (Knowledge Discovery in Databases) Process: 1996



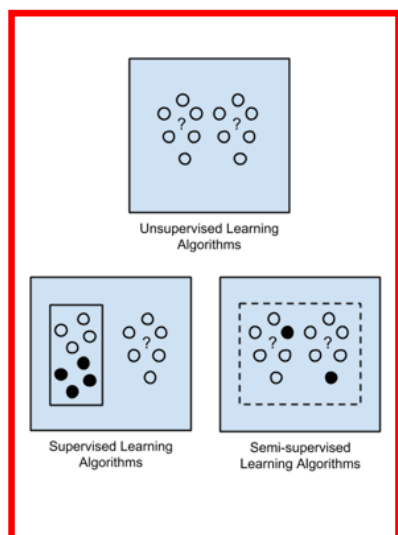
Obr. 3: Cross Industry Standard Process for Data Mining (CRISP-DM): 2000

Strojové učenie a analýza dát (Obr. 4):

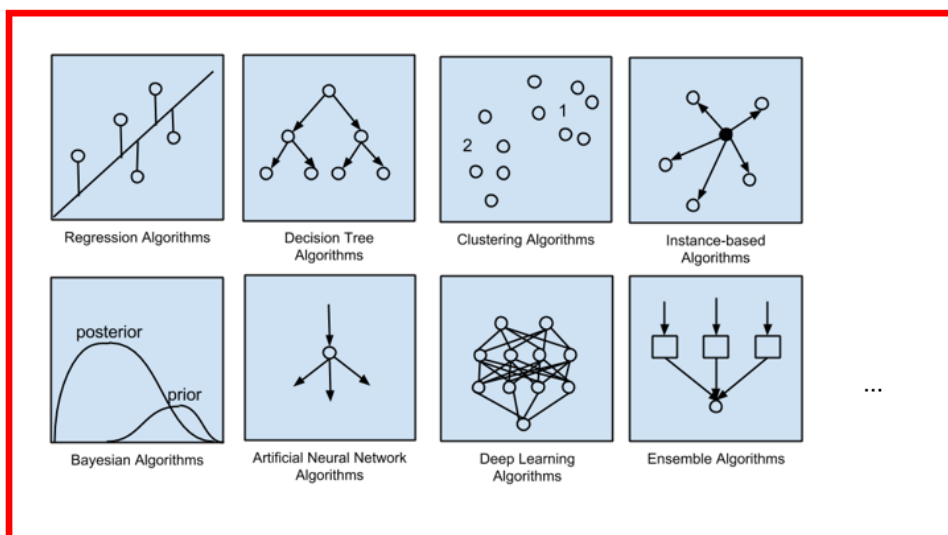


Obr. 4: SAS institute – diagram reprezentujúci postavenie jednotlivých oblastí v analýze dát

Delenie metód strojového učenia: podľa typu učenia (Obr. 5) a podľa typu algoritmu (Obr. 6):



Obr. 5: Podľa typu učenia



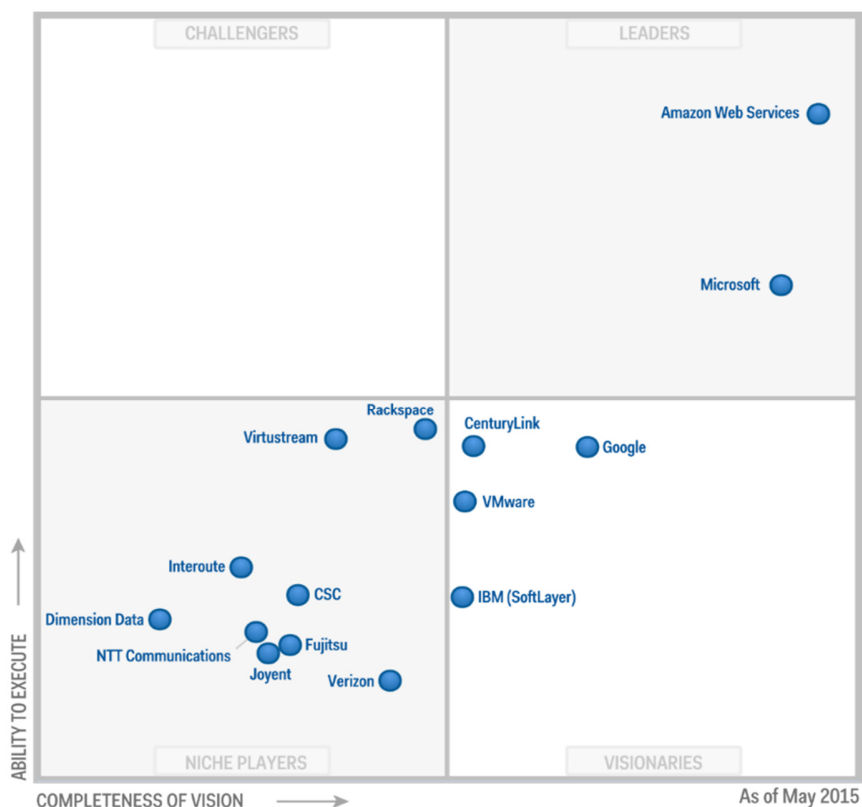
Obr.6: Podľa typu algoritmu

Sprostredkovatelia cloud computingu s možnosťou analýzy dát (Obr. 7):



Obr. 7: Sprostredkovatelia cloud computingu s možnosťou analýzy dát

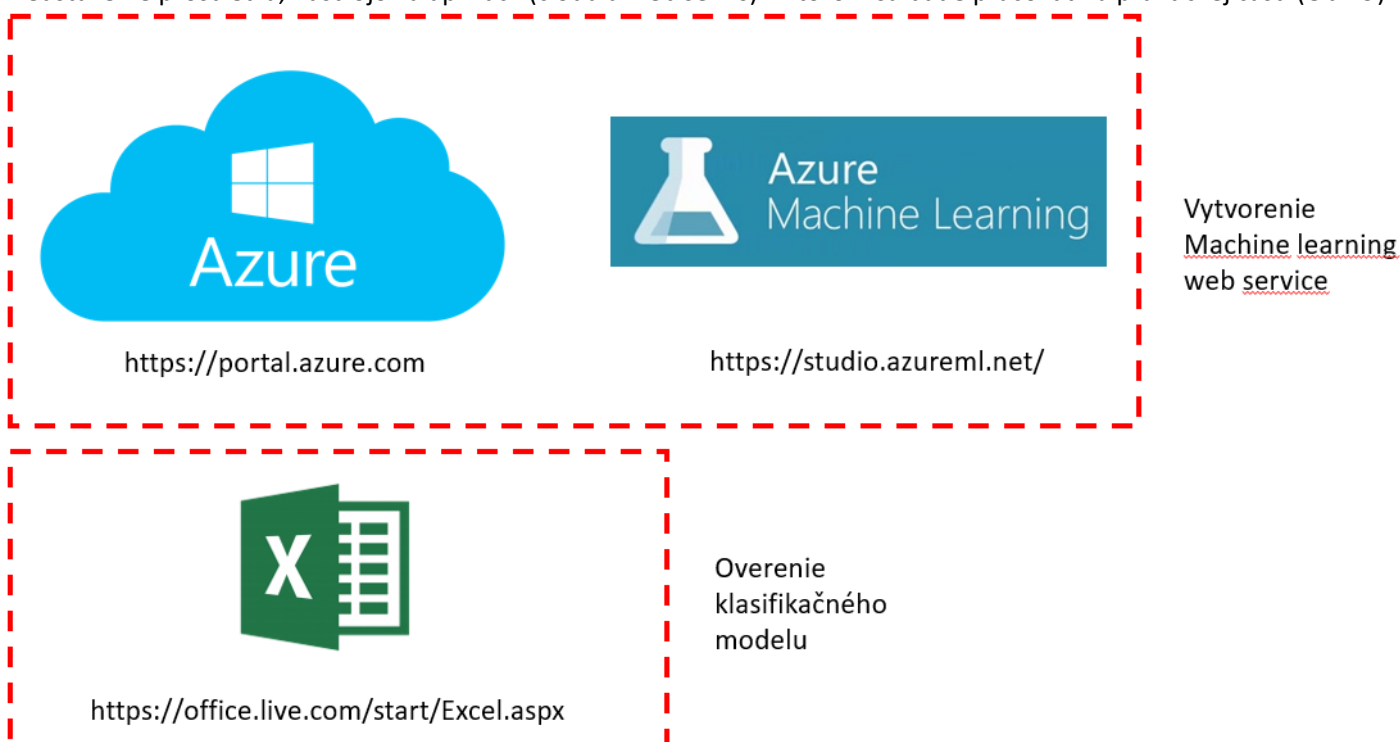
Gartner Magic Quadrants pre IaaS v roku 2015 (Obr. 8):



Obr. 8: Gartner Magic Quadrants pre IaaS v roku 2015

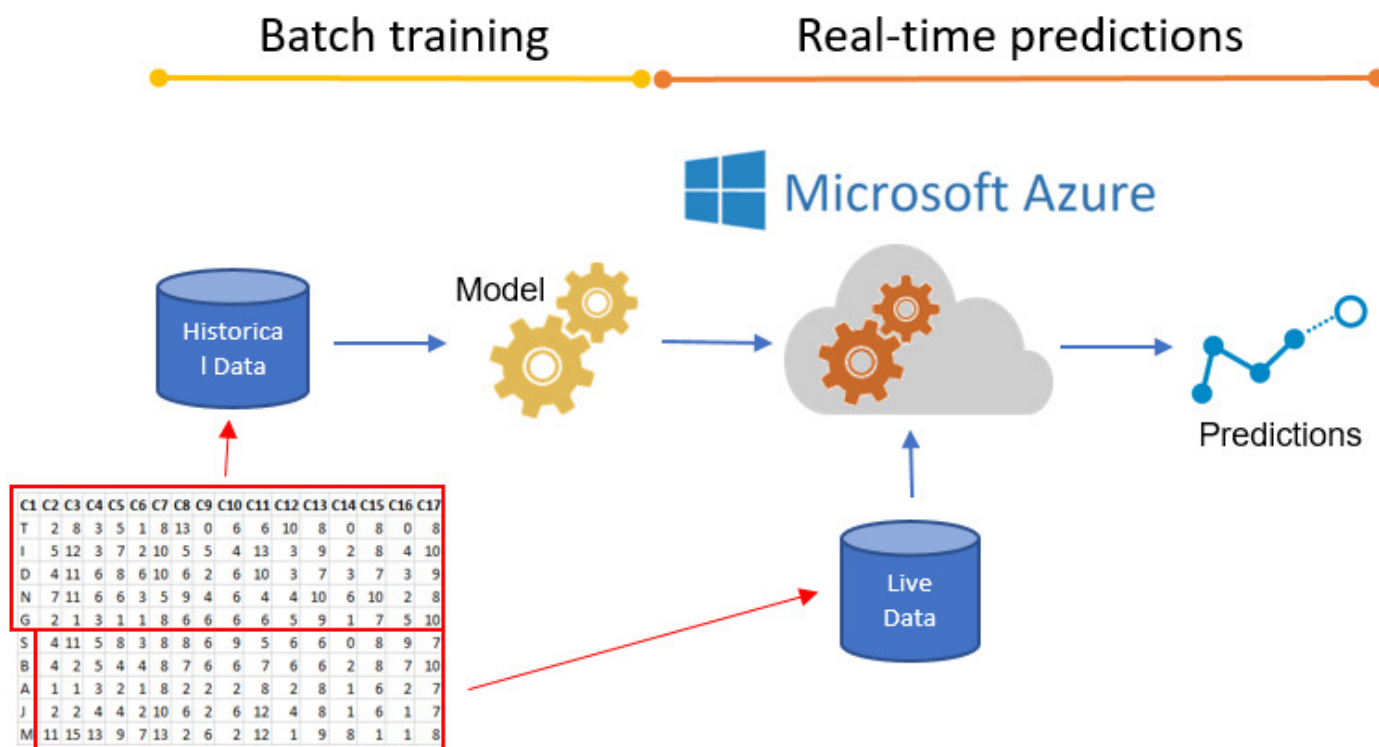
Praktická časť – strojové učenie na cloude (10 min)

Predstavenie prostredia, nástrojov a aplikácií (cloud a web servis) v ktorom sa bude pracovať na praktickej časti (Obr. 9):



Obr. 9: Prostredia, nástroje a aplikácie pomocou ktorých sa vypracuje praktická časť cvičenia

Architektúra real-time predikcie (analýzy), ktorá zahŕňa naučenie modelu na cloud a potrebné servis (Obr. 10):



Obr. 10: Architektúra pre naučenie modelu na cloud a potrebné servis

Príklad data setu určeného pre analýzu (Obr. 11 a Obr. 12):

Col.	Type	Description
1.	Enum	lettr capital letter (26 values, A - Z)
2.	Int	x-box horizontal position of box
3.	Int	y-box vertical position of box
4.	Int	width width of box
5.	Int	high height of box
6.	Int	onpix total # on pixels
7.	Int	x-bar mean x of on pixels in box
8.	Int	y-bar mean y of on pixels in box
9.	Int	x2bar mean x variance
10.	Int	y2bar mean y variance
11.	Int	xybar mean x y correlation
12.	Int	x2ybr mean of x * x * y
13.	Int	xy2br mean of x * y * y
14.	Int	x-egc mean edge count left to right
15.	Int	xegvy correlation of x-egc with y
16.	Int	y-egc mean edge count bottom to top
17.	Int	yegvx correlation of y-egc with x



Obr. 11: Informácie o atribútoch data setu a grafická reprezentácia data setu

Label

16 primitive numerical attributes

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
T	2	8	3	5	1	8	13	0	6	6	10	8	0	8	0	8
I	5	12	3	7	2	10	5	5	4	13	3	9	2	8	4	10
D	4	11	6	8	6	10	6	2	6	10	3	7	3	7	3	9
N	7	11	6	6	3	5	9	4	6	4	4	10	6	10	2	8
G	2	1	3	1	1	8	6	6	6	6	5	9	1	7	5	10
S	4	11	5	8	3	8	8	6	9	5	6	6	0	8	9	7
B	4	2	5	4	4	8	7	6	6	7	6	6	2	8	7	10
A	1	1	3	2	1	8	2	2	2	8	2	8	1	6	2	7
J	2	2	4	4	2	10	6	2	6	12	4	8	1	6	1	7
M	11	15	13	9	7	13	2	6	2	12	1	9	8	1	1	8

Obr. 12: Data set (zdroj: Letter Recognition Data Set from UCI Machine Learning Repository, Creator: David J. Slate, Odesta Corporation; 1890, Source: <https://archive.ics.uci.edu/ml/datasets/Letter+Recognition>)

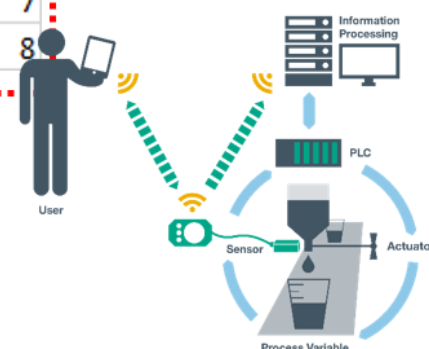
Úlohou je identifikácia 26 písmen latinskej abecedy, súbor obsahuje 20.000 záznamov s 20 rôznymi fontami. Každý záznam má 16 numerických atribútov (ktorý môže nadobúdať hodnotu od 0 do 15) a jeden znak. Tento data set je veľmi podobný data setom v priemysle. Predstavte si, že „T“ je istá porucha, ktorú chceme v systéme identifikovať. Porucha „T“ nastala keď sme za ebou namerali hodnoty: 2, 8, 3, 5, 1, ... Takže máme klasifikačný model pre klasifikáciu jednotlivých typov poruchy: T, I, D, N, G, ... (viď: Obr. 13):

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17
T	2	8	3	5	1	8	13	0	6	6	10	8	0	8	0	8
I	5	12	3	7	2	10	5	5	4	13	3	9	2	8	4	10
D	4	11	6	8	6	10	6	2	6	10	3	7	3	7	3	9
N	7	11	6	6	3	5	9	4	6	4	4	10	6	10	2	8
G	2	1	3	1	1	8	6	6	6	6	5	9	1	7	5	10
S	4	11	5	8	3	8	8	6	9	5	6	6	0	8	9	7
B	4	2	5	4	4	8	7	6	6	7	6	6	2	8	7	10
A	1	1	3	2	1	8	2	2	2	8	2	8	1	6	2	7
J	2	2	4	4	2	10	6	2	6	12	4	8	1	6	1	7
M	11	15	13	9	7	13	2	6	2	12	1	9	8	1	1	8

Hodnoty ktoré boli
namerane keď
porucha „T“
nastala

Typ poruchy

Merania zo senzorov

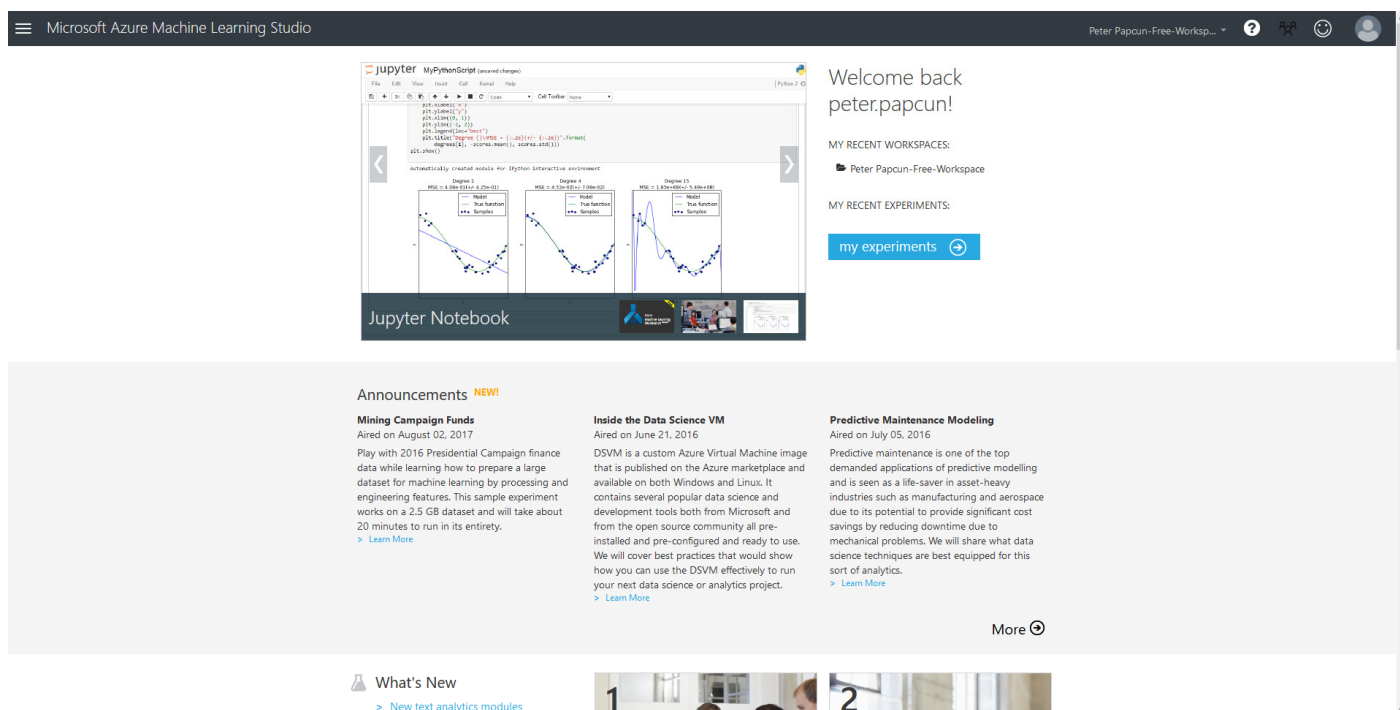


Obr. 13: Podobnosť data setu s data setom porúch (alebo iných príznakov) v priemysle

Praktická časť – vytvorenie služby pre strojové učenie v prostredí Microsoft Azure (60 min)

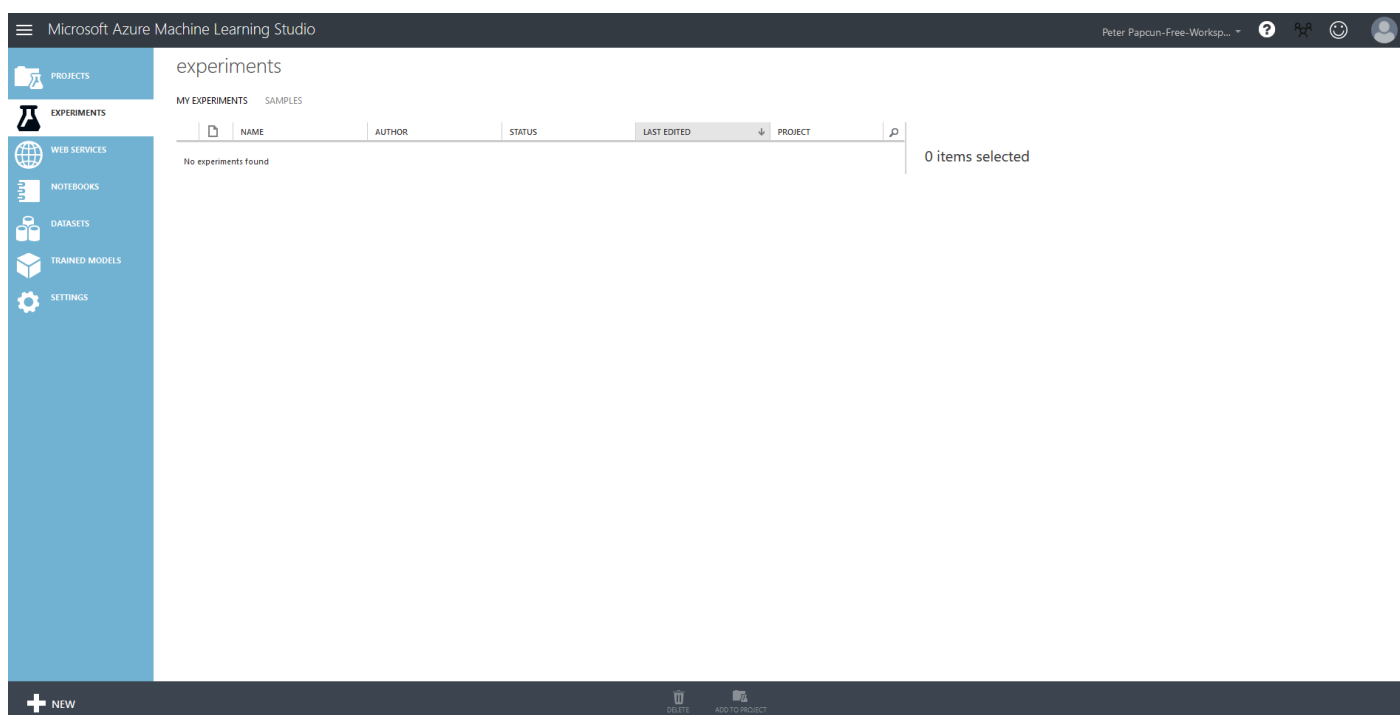
Navštívte stránku: <https://studio.azureml.net/>

Prihláste sa pomocou účtu, ktorý Ste si vytvorili pre Microsoft Azure na 5. cvičení (Obr. 14):



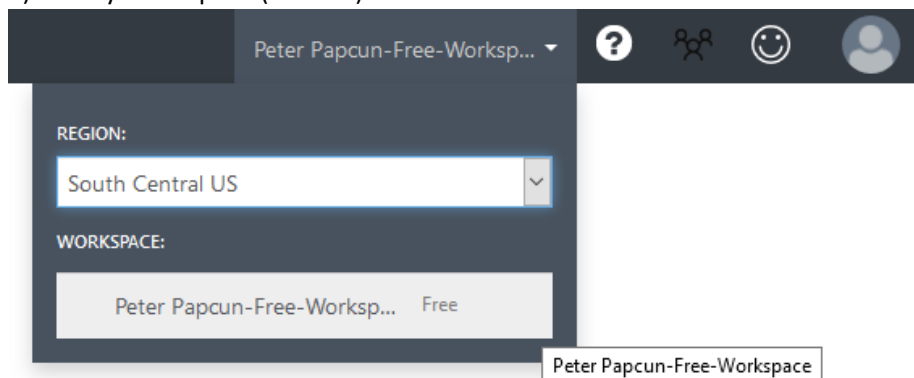
Obr. 14: Microsoft Azure Machine Learning Studio

Po prihlásení, poprípade po kliknutí na tlačidlo „my experiments“ sa Vám otvorí toto okno (Obr. 15):



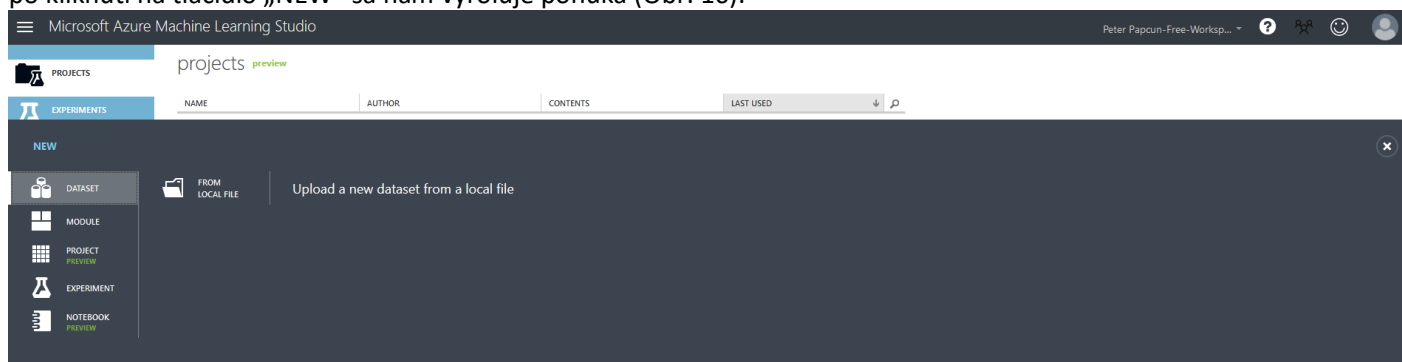
Obr. 14: My experiments

Vytvorte (skontrolujte) si nový Workspace (Obr. 15):



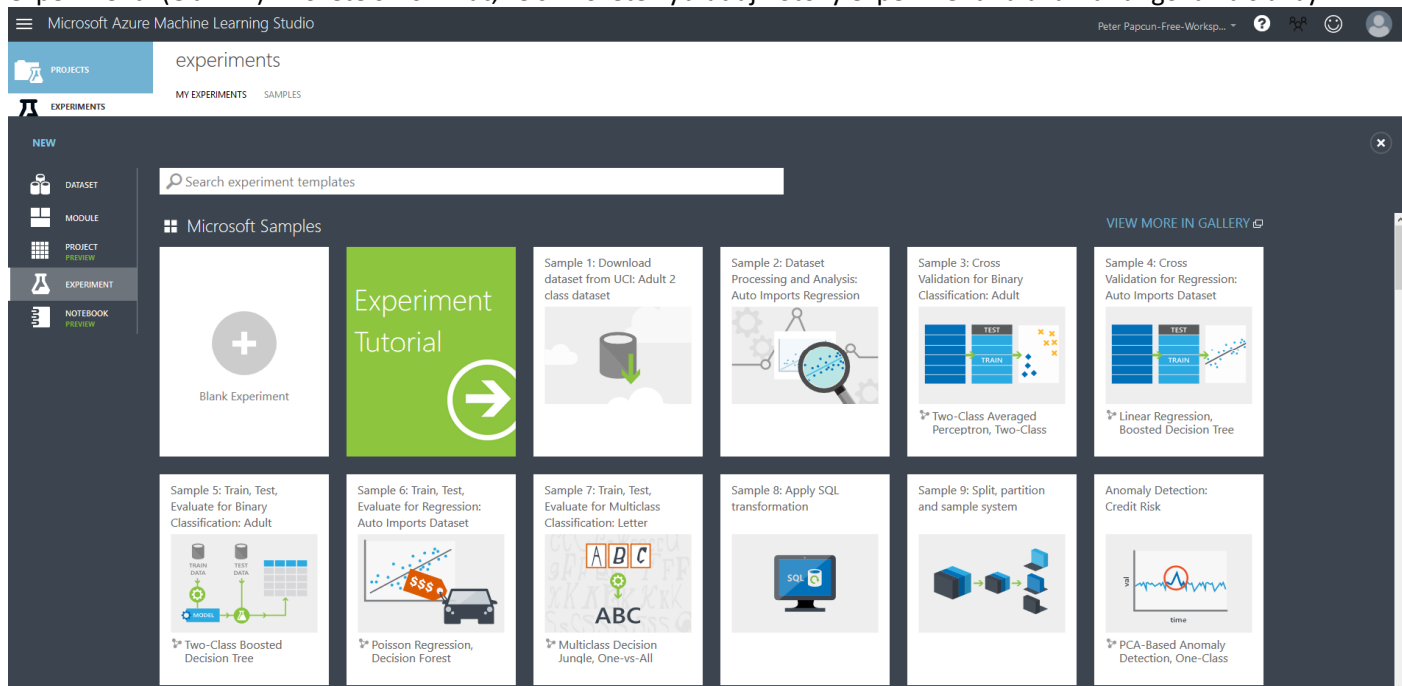
Obr. 15: New Workspace

Ak je to potrebné, nový data set pridáme pomocou kliknutia na tlačidlo „NEW“ v spodnej lište aplikácie (Obr. 14, dole), po kliknutí na tlačidlo „NEW“ sa nám vyroluje ponuka (Obr. 16):



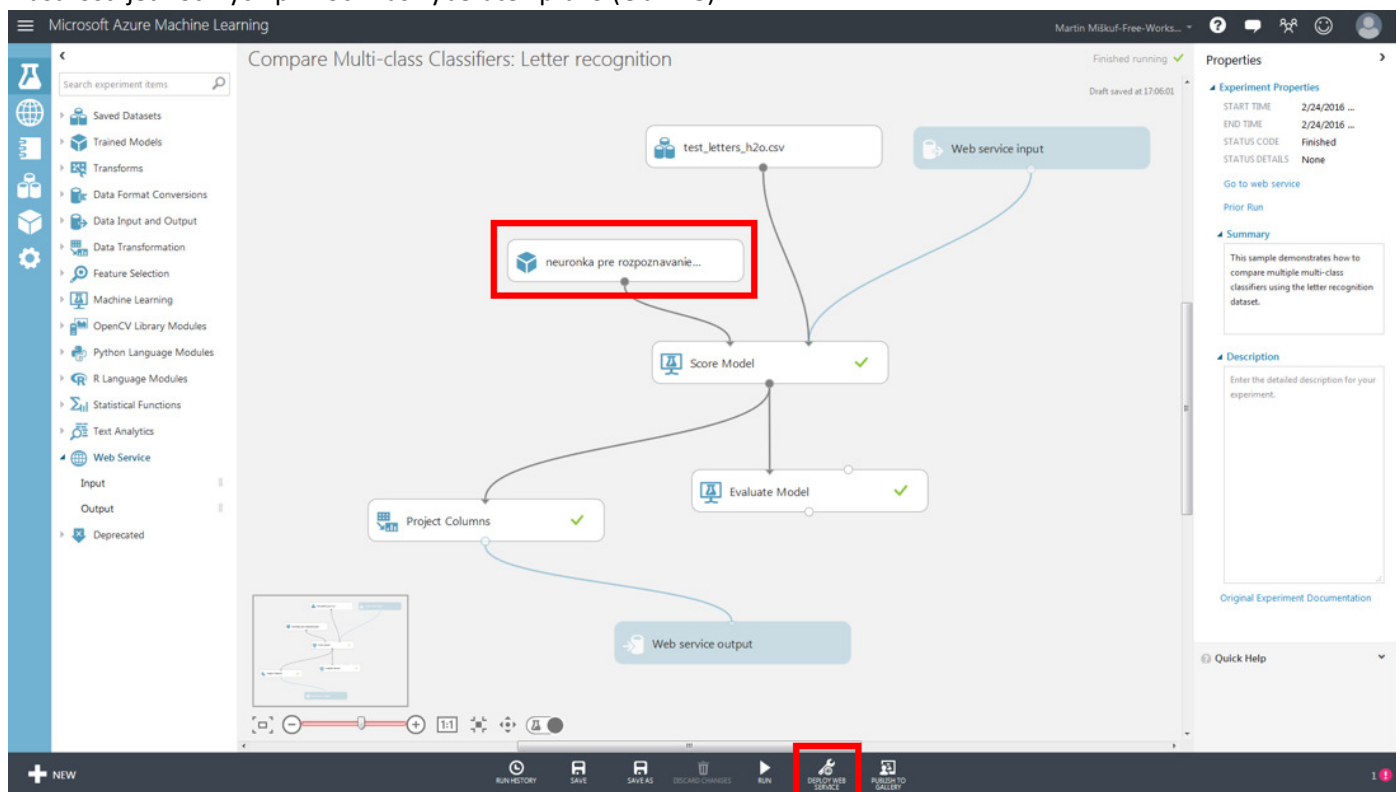
Obr. 16: Nový dataset

Pre ukážku fungovania aplikácie to nie je potrebné, pretože *Machine Learning Studio* disponuje mnohými data setmi a tak môžeme vytvoriť nový experiment. V záložke „NEW“ klikneme na tlačidlo „EXPERIMENT“, kde vyberieme „Blank experiment“ (Obr. 17). Môžete si všimnúť, že si môžete vybrať aj hotový experiment na ukážku fungovania služby.



Obr. 17: Nový experiment

Po vytvorení nového experimentu si vyskladáme nový klasifikačný model pomocou „grab and drop“ z ponuky vľavo, vlastnosti jednotlivých prvkoch zas vyberáte vpravo (Obr. 18):

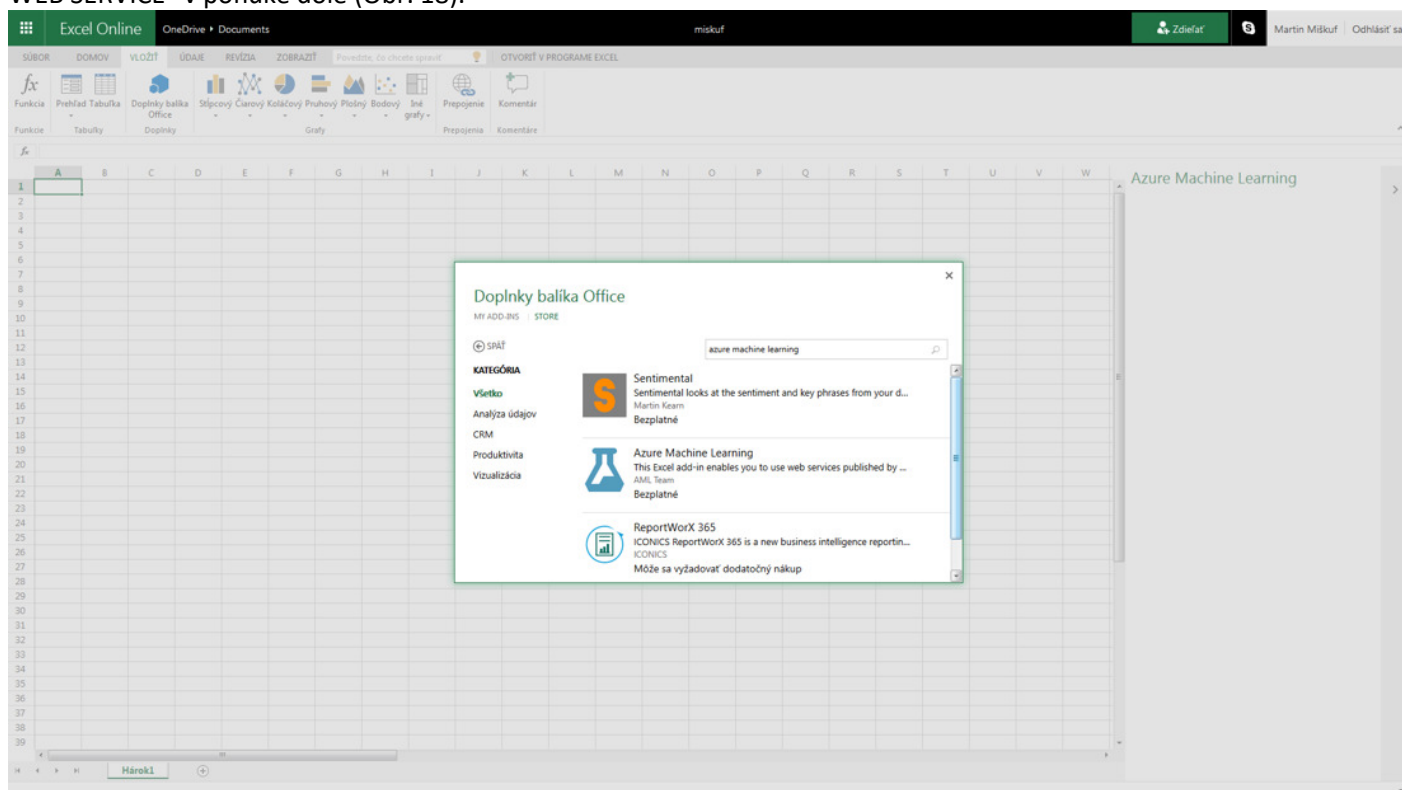


Obr. 18: Klasifikačný model

Tento model si môžete otestovať pomocou tlačidla „RUN“, alebo môžete službu deploy-núť.

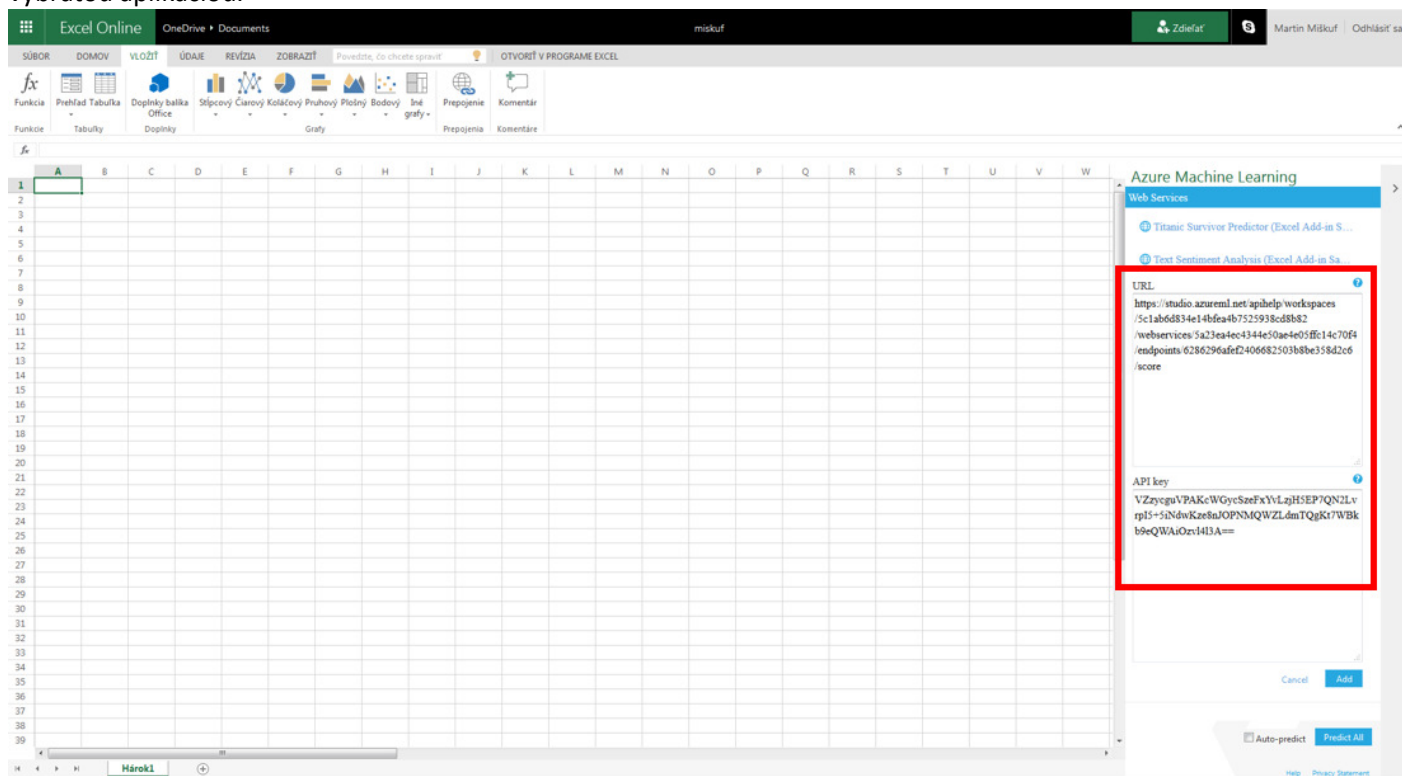
Obr. 19: Vytvorenie API

Deploy-nutie slúži nato, aby služba bola prístupná ako API cez Internet. Službu deploy-nete pomocou tlačidla „DEPLOY WEB SERVICE“ v ponuke dole (Obr. 18).



Obr. 20: Excel Online

Po kliknutí na spomenuté tlačidlo sa Vám otvorí okno (Obr. 19), kde si viete pozrieť „API HELP PAGE“ pre spojenie s Vami vybranou aplikáciou.



Obr. 21: Azure Machine Learning v Excel Online

Veľmi potrebné je skopírovať si „API key“, ktorý slúži na pripojenie aplikácie k Vašej vytvorenej službe. Otvorte si Excel, poprípade na tomto cvičení bude postačovať Excel Online. V aplikácii otvorte „Doplnky balíka Office“ (Obr. 20).

Obr. 22: Tabuľka vytvorená Vašou deploy-nutou službou

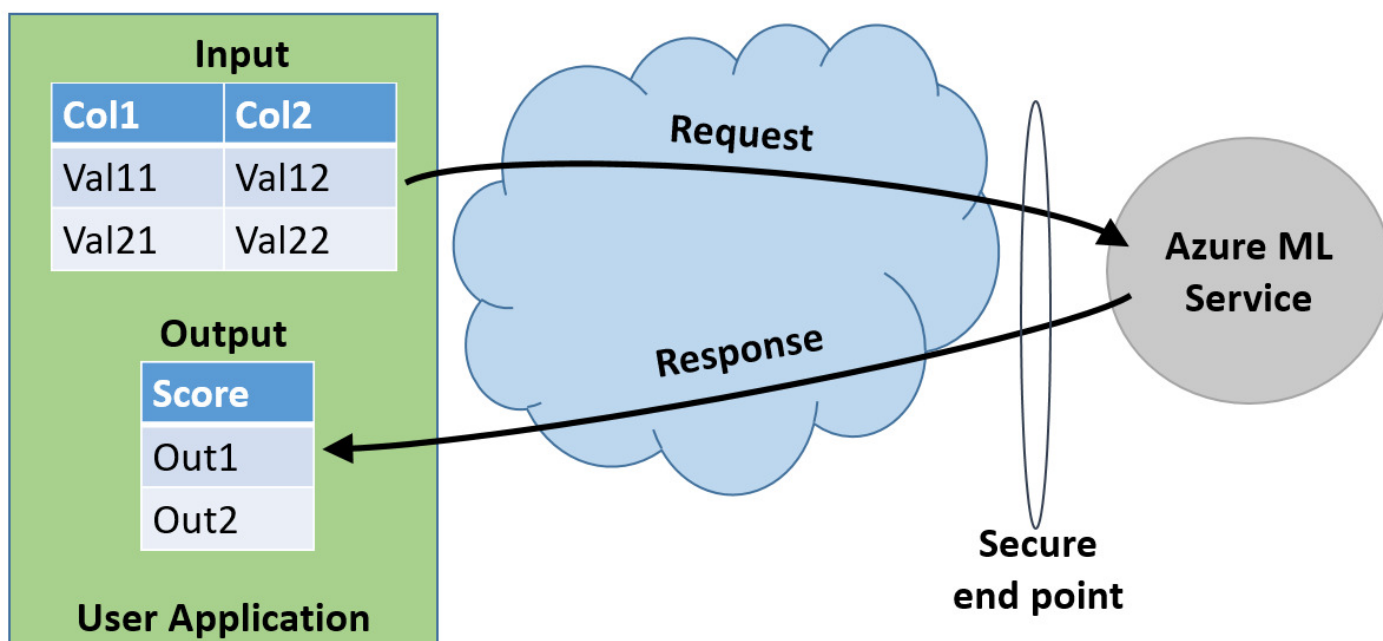
Z balíka si vyberieme „Azure Machine Learning“, kde v záložke „Web Services“ vyplníme URL a API key Vášho vytvoreného servisu (Obr. 21).

Obr. 23: Tabuľka po vložení hodnôt

Keď vložíte URL a API key kliknite na tlačidlo „Add“ (Obr. 21), tak po kliknutí sa Vám vytvorí tabuľka (Obr. 22). Teraz môžete do tabuľky vložiť Vami zvolené hodnoty kliknúť na „Predict“ (Obr. 23). Po kliknutí na tlačidlo „Predict“ Vám služba vyplní predikovanú hodnotu (posledný stĺpec), v tomto predikovanú písmeno z data-setu spomínanom vyššie (Obr. 24):

Obr. 24: Predikovaná hodnota v poslednom stĺpci

Ma pozadí toho celého sa udial HTTP request, ktorý je bližšie vysvetlený scémou na Obr. 25:



Obr. 25: Request a response

Tento postup si môžete zopakovať na priložených dátach o počasi (pocasioe.xlsx).