
Mesterséges intelligencia oktatása a BGE KKK-n

Budai László¹, Bölcskei Attila², Keresztes Éva Réka³, Talata István⁴

¹egyetemi docens, ²főiskolai tanár, ³adjunktus, ⁴főiskolai tanár

^{2,3,4}Budapesti Gazdasági Egyetem, Külkereskedelmi Kar,
Társadalomtudományi Módszertan Tanszék, ¹Budapesti Gazdasági Egyetem,
Külkereskedelmi Kar, Nemzetközi Kereskedelem és Logisztika Tanszék

E-mail: budai.laszlo@uni-bge.hu, bolcskei.attila@uni-bge.hu,
keresztes.eva@uni-bge.hu, talata.istvan@uni-bge.hu

DOI: [10.29180/978-615-6342-61-4_12](https://doi.org/10.29180/978-615-6342-61-4_12)

Összefoglalás: A mesterséges intelligencia régebbi múltra tekint vissza, mint azt sokan sejtjenék, de napjainkra már elengedhetetlen technológiai lehetőségeket tartogat, az élet minden területére vonatkozóan. A leendő munkavállalók versenyképességének megőrzése az egyetemek és vállalatok közös célja. Ennek érdekében indult útjára a BGE KKK-n egy gyakorlatorientált megközelítésű kurzus. A tanulmány célja, hogy bemutassuk a mesterséges intelligencia oktatásának egy lehetséges felépítését.

Kulcsszavak: mesterséges intelligencia, oktatás, gépi tanulás, szövegbányászat, képanaltika, regresszió, klasszifikáció, klaszterezés

Abstract: Artificial intelligence has a longer history than many might suspect, but nowadays it already holds indispensable technological possibilities for all areas of life. Maintaining the competitiveness of prospective employees is a common goal of universities and companies also. To this end, a course with a practice-oriented approach was launched at the BBS Faculty of International Management and Business. The aim of this study is to present a possible structure of artificial intelligence education.

Keywords: artificial intelligence, education, machine learning, text mining, image analytics, regression, classification, clustering

1. Bevezetés

Napjainkra a mesterséges intelligencia (MI) átszövi egész életünket, mindenki hallott már róla, akár alkalmazza is azt nap, mint nap, esetleg a tudtán kívül is.

Nagyvállalati partnereink tapasztalatai szerint, illetve a közelmúlt eseményei hatására is erőteljesebben beigazolódtott, hogy a 2020-as évek meghatározó technológiái között szerepel a digitális ellátási lánc, üzleti intelligencia, mesterséges intelligencia és a kvantum számítás is.

Mostanra az is beigazolódtott, hogy a mesterséges intelligencia nem csupán egy prioritást élvező technológia, hanem egy szükséges alkalmazási terület.

A jövő munkavállalói számára elengedhetetlen tehát ezen technológiák, módszerek, eszközök rendszerszemléletű ismerete, alkalmazási lehetőségeinek átlátása. Fő cél, hogy egy adott vállalati probléma megoldása céljából a leendő munkavállaló ki tudja választani a legmegfelelőbb eszközt, módszert, technikát.

Többek között a fentebb említett célrendszer miatt indult útjára a Budapesti Gazdasági Egyetem Külkereskedelmi Karán (BGE KKK) egy új, szabadon választható kurzus *Mesterséges intelligencia a gyakorlatban* néven.

Mivel a kurzus erőteljesen gyakorlatorientált, így az elméleti megalapozása a lehető legkompaktabb módon zajlik egy 60 perces prezentáció formájában.

A mesterséges intelligencia olyan eszköz a kezünkben, mely segíti a problémamegoldást, illetve döntéstámogatást nyújt az ember számára.

Az elméleti bevezetőben elhangzik még a csoportosítási lehetősége a mesterséges intelligenciának úgy, mint gyenge-erős AI, illetve a gépi tanulás és mély tanulás kapcsolatáról is szó esik. A gépi tanulás (Machine Learning - ML) matematikai adatmodellekkel tanít be számítógépeket közvetlen felügyelet nélkül. Ez a mesterséges intelligencia egy részhalmaza. A gépi tanulás algoritmusokkal azonosít mintákat az adatokban, amelyekkel ezután adatmodellt készít, és előrejelzéseket végez. A gépi tanulás eredményei az adatok és a tapasztalat mennyiségének növekedésével egyre pontosabbak – ahogyan az emberek is egyre ügyesebbé válnak a gyakorlás által.

A gépi tanulás rendkívül könnyen alkalmazkodik különböző helyzetekhez, így ideális olyan forgatókönyvekben, ahol az adatok folyamatosan változnak, a kérés vagy feladat típusa módosulhat, vagy egy megoldás kódolása szinte lehetetlen lenne.

A mély tanulás a gépi tanulás egy speciális formája, amely neurális hálózatokkal szolgáltat válaszokat. A mély tanulás képes egyedül pontosságot mérni, az adatokat pedig az emberi agyhoz hasonlóan osztályozza – így a létező legemberibb mesterséges intelligenciák működését is segíti.

Az MI a legjobb módja annak, hogy munkára bírjuk az adatainkat az adatvezérelt vállalati környezetekben. 10-ből 9 vállalat rálépett, vagy hajlandó lenne rálépni az MI támogatott útra, azonban ehhez számos előfeltételnek kell megfelelni, de leginkább egy tiszta adatökoszisztémának kell jelen lennie a vállalatnál. Sok esetben az MI bevezetésének folyamata ezzel kezdődik, hogy az adatok szintjét rendbe kell rakni. Ezután lehet csak lefektetni, megteremtteni az MI-alapú analitikai alapokat. Célunk ezt követően, hogy egy megbízható MI modellt építsünk és skálázzunk, folyamatosan monitorozva a működését és operacionalizálni. Mindezen tevékenységekre elérhetőek a BGE hallgatói számára teljes IBM, illetve Microsoft portfóliók is.

Az MI tehát segít jobban megérteni az üzleti nyelvet, segít automatizálni a munkafolyamatokat, bizalmat és bizonyos valószínűséggel kiszámíthatóságot ad az adott üzleti folyamatban részt vevők számára.

2. Az MI alkalmazási területei

A kurzuson részt vevő hallgatókkal a következő gépi tanulási problémákra nézünk konkrét esettanulmányokat:

- regresszió

- klasszifikáció
- klaszterezés
- anomália detektálás
- idősor analitika
- ...

Fontos, hogy a hallgatók tisztában legyenek azzal, hogy egy gépi tanulási modell, melyet ők maguk építenek fel, milyen inputra milyen outputot (score) tud szolgáltatni, és ennek az outputnak a megfelelő értelmezése (evaluation) is elengedhetetlen.

Bármelyik fent felsorolt probléma esetében többféle bemeneti adattípusról is beszélhetünk, melyek mindegyikéről szó is lesz a kurzuson, és mindegyik más hozzáállást igényel a modell építése során:

- szöveg → szövegbányászat, szöveganalitika
- hang → hangbányászat, hanganalitika
- kép → képfeldolgozás, képanalitika
- numerikus adattípus

Az alkalmazott technológiákat illetően követjük a vállalati életben is alkalmazott eszközöket, módszertanokat. A kurzus óraszámának korlátosságából kiindulva alapvetően kódolásmentesen építünk gépi tanulási modelleket, de felhívjuk a figyelmet arra is, hogy egy-egy felmerült probléma esetében szükséges lehet a hatékonyság növelése céljából valamilyen programozási környezetet használni (Java, Python...).

Mindezen szempontokat figyelembe véve a következő rendszerek használatára fókuszálunk:

- RapidMiner
- IBM SPSS Modeler
- Microsoft Azure ML Studio
- egyéb kép-, hang- és szöveganalitikai rendszerek (Lobe, <https://www.saanketh.com/>, ...)

Mielőtt megkezdjük saját modelljeink építését, számos IBM és Microsoft esettanulmányt említünk a hallgatóknak az élet szinte minden területét felölelve, a rendszerszemlélet erősítése gyanánt (például hangalapú AI-támogatott chatbot építése).

A gépi tanulási modellt a következőképpen építjük fel:

- hozzáférés az adatmodellhez: lokális betöltés, import, felhő...
- metaadatok megadása
- adattisztítás: szerkezeti épség, hiányzó adatok, anomáliák, outlier-ek, eloszlásvizsgálatok ...
- feature selection: azon adatmezők identifikálása, melyek a legnagyobb prediktori erővel bírnak a célmező előrejelzését (besorolását) illetően
- split: megosztjuk az adathalmazt betanításra és tesztelésre (esetleg validálásra)
- megfelelő modell (algoritmus) kiválasztása

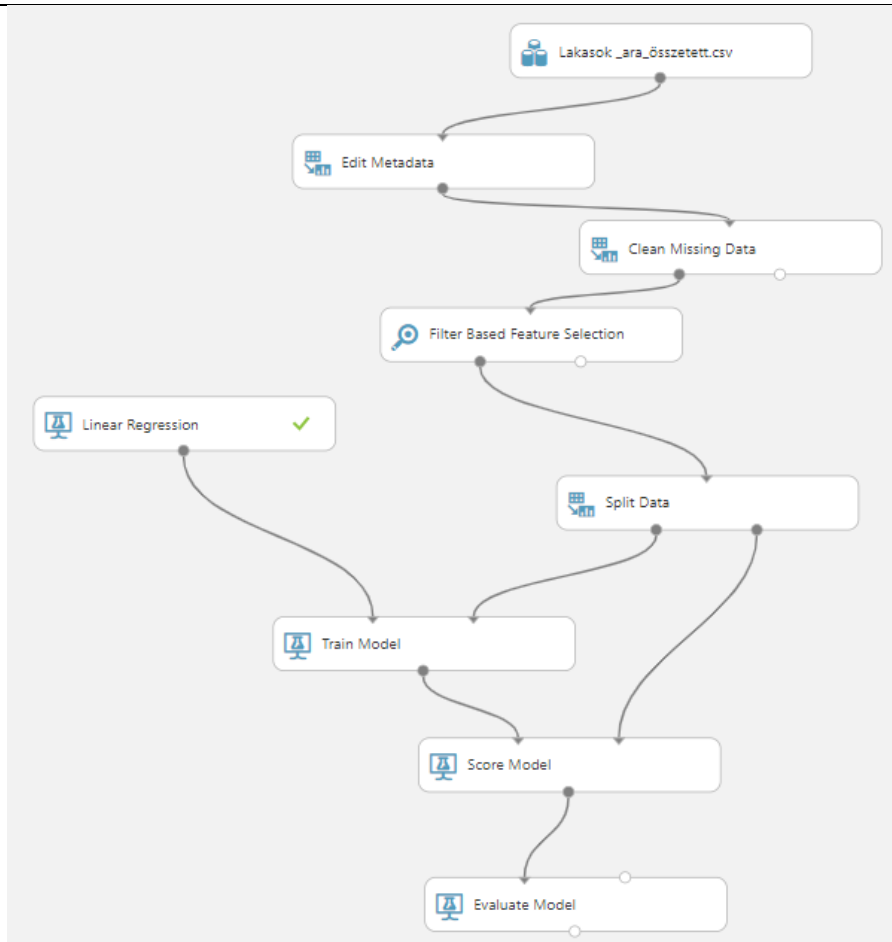
- modell hiperparaméterezése
- modell betanítása
- modell tesztelése
- modell kiértékelése: megfelelő kiértékelési metrikák szerint
- web service építése: a betanított és validált modellt előkészítjük az új szituációkban való alkalmazására
- web service deploy: a modellt 'hadrendbe' állítjuk, aza bevetjük új, éles környezetben
- modell monitorozása: a modell működése bevetett állapotban eltérhet a tesztelt modelltől (például új, előre nem látható ható tényezők), így szükség lehet a modell további fejlesztésére. Itt vizsgáljuk azt is, hogy a modell megbízható és átlátható döntéseket hozzon meg.

A következő fejezetekben néhány példát mutatunk a teljesség igénye nélkül.

3. Regresszió

A regresszió a leggyakrabban használt gépi tanulási technika, amely gyakran áll a tudományos felfedezések, az üzleti tervezés és a tőzsdei piacelemzések háttérében. A regresszió egy egyszerű, gyakori és rendkívül hasznos adatelemzési technika, amelyet köznyelvben "trendvonal illesztésének" nevezünk. A regresszió egy vagy több funkció és egyetlen címke közötti kapcsolat erősségét azonosítja. Egyszerűsége, kiszámítható viselkedése, előrejelzési képessége és magas szintű értelmezhetősége azt jelenti, hogy ezt a technikát a pénzügy, az üzleti, a társadalomtudományok, az epidemiológia és az orvostudomány területén használják [1].

Feladataink között szerepel például autó import regresszió, kereskedelmi termék fogyasztásának (igények) előrejelzése, árak (lakás, autó, termék) előrejelzése.



Forrás: saját készítés

1.ábra Regressziós gépi tanulási modell

Az 1. ábrán látható egy regressziós feladat megoldása, ahol a lakások árát kellett előre jelezni (megbecsülni), számos input paraméter (régión, közbiztonság, alapterület, parkolási lehetőség...) felhasználásával.

A csomópontok (node) sorrendje láthatóan jól követi az előző fejezetben írt építési lépéseket.

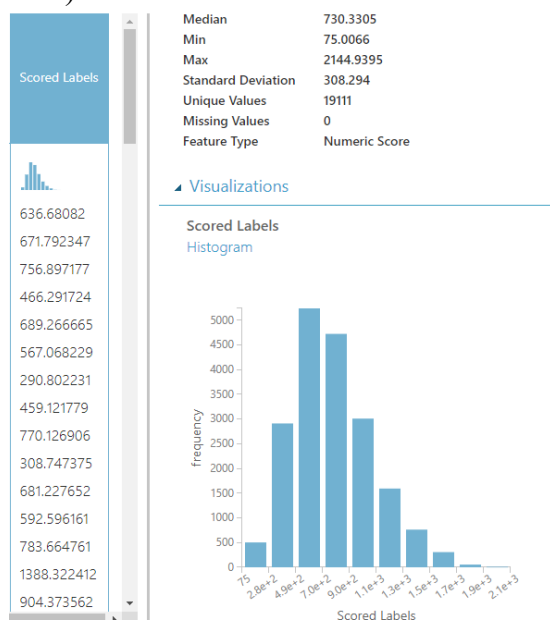
A már betanított modell esetében lehetőségünk van arra, hogy megtekintsük a hozzá tartozó elaszticitás egyenletet is (2. ábra):

Batch Linear Regressor		Feature Weights	
Settings		Feature	Weight
Setting	Value		
Bias	True	Bias	515.511
Regularization	0.001	a piacrakerüléstől az adás- vételig eltelt idő (nap)	4.9292
Allow Unknown Levels	True	közlekedés	-1.96632
Random Number Seed		eldási ár (e Ft/m ²)	1.27057
		társasház állapota	1.04945
		közbiztonság	0.781704
		szórakozóhely	0.581684
		településtípus	0.537748
		régió	-0.51254
		parkolás	0.413609
		zai	0.384878

Forrás: saját készítés

2.ábra Regressziós elaszticitás egyenlet

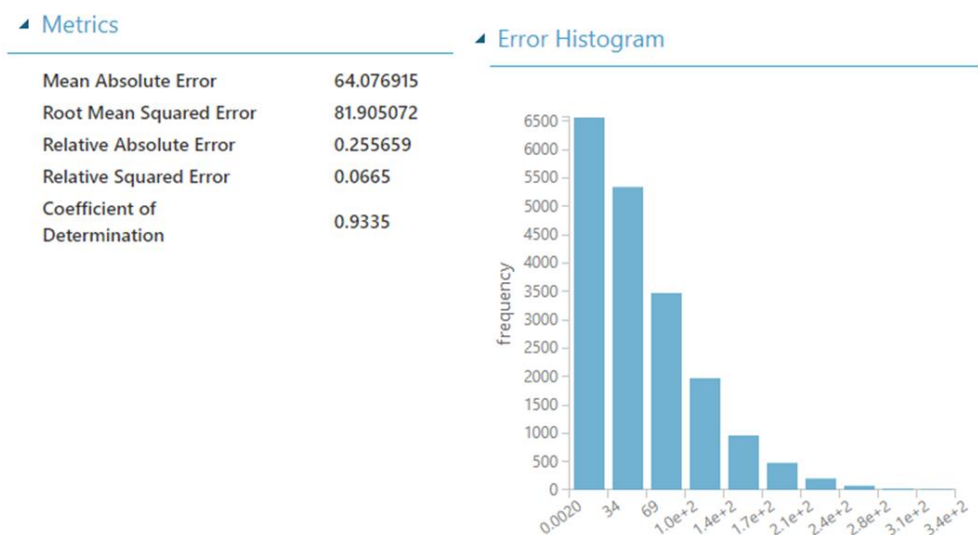
Tesztelés során az adathalmaz jellemzően 20 % - 30 % -át le szoktuk választani, hogy ezen keresztül szimuláljuk a modellnek az új szituációt, melyben helyt kell majd állnia. Ezen eredmény megtekintésére is van lehetőségünk (3. ábra):



Forrás: saját készítés

3.ábra Regresszió score

Végül érdemes időt szánni az adott probléma megoldásához tartozó kiértékelési metrikákra, melyek regressziós feladat esetén leggyakrabban a következők (4. ábra):



Forrás: saját készítés

4.ábra Regressziós gépi tanulási modell kiértékelése, metrikák

Éles vállalati környezetben gyakran előfordul, hogy a regressziós modell jószágát a Mean Absolute Error (MAE) metrikával határozzák meg. Esetünkben például a target mező a lakás árának egy négyzetméterre jutó értéke (Ft/m²). A 4. ábrán látszik, hogy a gépi tanulási modell MAE értéke 64.076915, azaz a modellünk átlagosan 64076,915 Forintot tévedett lakásonként, ami a négyzetméter árát illeti. Mindent egybevetve ez nem is rossz eredmény.

A bemutatott modell a Microsoft Azure ML Studio-ban készült, de hasonlóan elkészíthető az IBM SPSS Modeler, illetve a RapidMiner node-alapú fejlesztői környezetekben is. [2]

4. Klasszifikáció

A másik gyakoribb gépi tanulási feladat az osztályozás (classification). Itt előre adott egy kategória-/osztályhalmaz és a célunk, hogy felügyelt gépi tanulási megközelítésben olyan modellt/döntési szabályokat tanuljunk a tanító adatbázis alapján, ami képes egy ismeretlen példát minél pontosabban valamelyik osztályba (class) sorolni (úgy is mondjuk, hogy 'címkézni' - labeling).

Osztályozási feladat például a jól ismert macska vs. kutya feladat, ahol képeket kell a { macska, kutya } két osztály valamelyikébe besorolni. [3]

Egy másik jól ismert feladat, amikor egy social media poszt szöveges tartalmát kell érzelmi töltet szerint a {pozitív, semleges, negatív} osztályok valamelyikébe vagy téma szerint (több alkalmazás-specifikus osztály) besorolni. [3]

A klasszifikációs problémáknál is számos különböző gépi tanulási modellt építünk a hallgatókkal: túlélési előrejelzés, ügyfél lemorzsolódás, betegség előrejelzése, ügyfél felülértékesítési előrejelzés, twitter szentiment analízis, szöveg besorolási problémák (szöveganalitika), prediktív karbantartás, selejtdetektálás (képanalitika) ...

Klasszifikációs problémák esetében hatékony algoritmusok szoktak lenni a gráfrepresentációs algoritmusok (döntési fák, erdők, dzsungel), de komplexebb probléma esetében neurális háló, illetve Bayes-háló kiépítése is javasolt lehet.

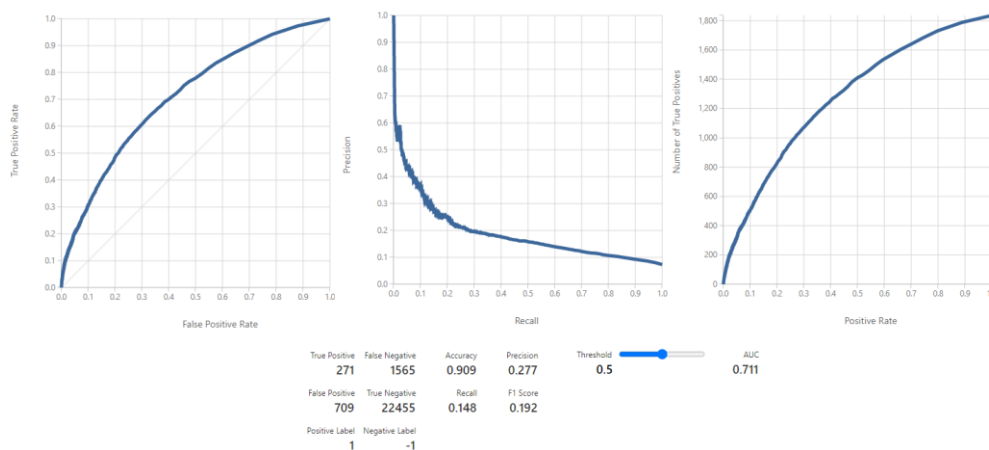
Ezekben az esetekben a leggyakrabban alkalmazott metrikák az ún. konfúziós mátrix elemeiből számíthatók ki (5. ábra):

True Positive	False Negative
271	1565
False Positive	True Negative
709	22455

Forrás: saját készítés

5.ábra Bináris klasszifikációhoz tartozó konfúziós mátrix

Ekkor a különböző arányok kiszámításával, illetve karakterisztikus görbék alkalmazásával a következő fontosabb metrikákhoz juthatunk (6. ábra):



Forrás: saját készítés

6.ábra Bináris klasszifikációhoz tartozó metrikák





A megfelelő kiértékelési metrika kulcsfontosságú egy gépi tanulási alkalmazás fejlesztésénél. Gondoljunk csak bele, ha olyan képosztályozót

fejlesztünk, ami röntgen képek alapján próbálja megmondani, hogy adott betegségben szenved-e a páciens (bináris osztályozási feladat), akkor nem mindegy, hogy a preferált rendszer a betegség osztály pontosságát (akire betegséget predikálunk az tényleg legyen az) vagy fedését (ne veszítsünk el beteg eseteket, nem baj, ha tévesen betegnek predikáltak száma nagyobb) tekinti fontosabb szempontnak.

5. Klaszterezés

Találjunk olyan csoportokat objektumok, adatok egy halmazában, hogy az egy csoportban lévő objektumok egymáshoz hasonlóak, míg a más csoportokban lévők pedig különbözőek.

A klaszterezés nem felügyelt gépi tanulásnak minősül, mert itt nem ismerjük az egyes egyedekre jellemző labelt.

Assignments	DistancesToClusterCenter no.0	DistancesToClusterCenter no.1	DistancesToClusterCenter no.2
			
1	1.000061	0.964622	1.027728
1	1.191315	0.503935	1.015065
2	1.031485	1.18886	0.520114
2	1.017962	1.049648	0.872991
0	0.927989	1.043121	1.003872
1	1.005788	0.983493	1.002046

Forrás: saját készítés

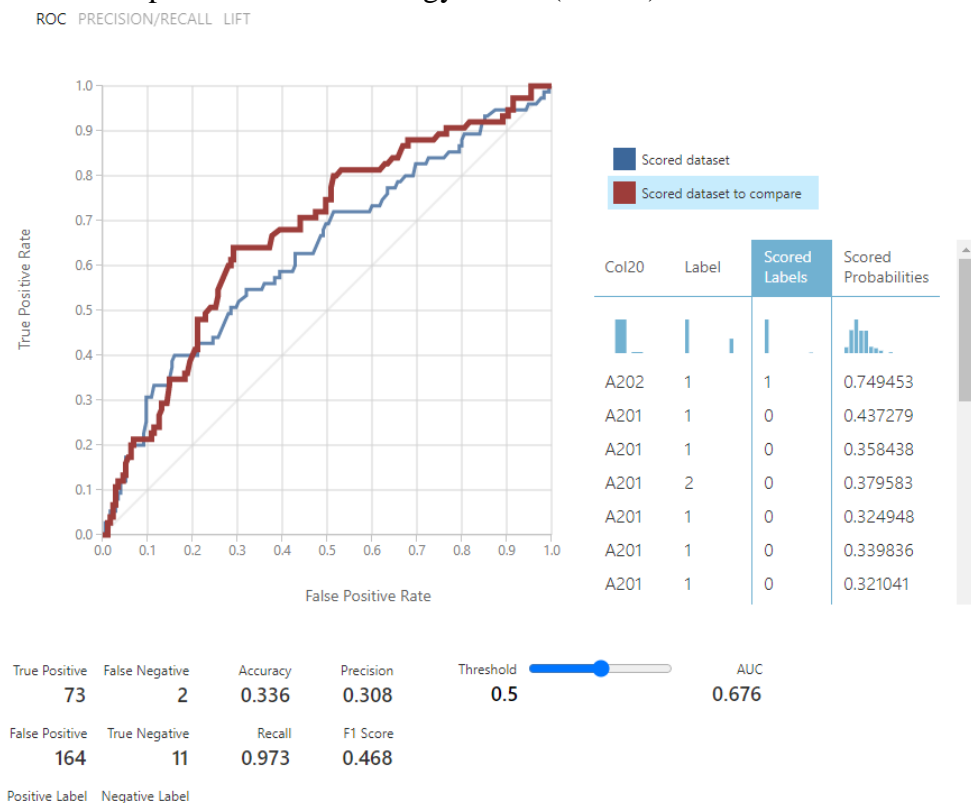
7.ábra Klaszterezési gépi tanulási modell eredménye

A 7. ábrán láthatjuk egy feladat végeredményét, ahol vállalatokat kellett klaszterekbe sorolni, melyek esetében szöveges leírás volt megadva a tevékenységükről.

6. Anomália detektálás

Az adatbányászatban az anomáliák felderítése (angolul: anomáliák felderítése vagy outlier detektálás) azon elemek, események vagy néhány megfigyelés azonosítása, amelyek gyanút keltenek, és jelentősen eltérnek a többség többi adattól.

Példánkban egy bank hitelkockázati elemzését végeztük el, melyben azonosítani próbáltuk az outlier ügyfeleket (8. ábra):



Forrás: saját készítés






8. ábra Outlier detektálás eredményeinek összehasonlítása

A 8. ábrán mutatjuk a két alkalmazott gépi tanulási algoritmus (one-class SVM, PCA-based) hatékonyságának összehasonlítását, illetve az outlier-ként megjelölt ügyfeleket is ('Scored labels' oszlopban az első ügyfél például 75 %-os valószínűséggel annak tekinthető).

7. Idősor analitika

Az idősorok meghatározott időközönként egyenként mért adatok sorozatos mintái. Az idősor-elemzés célja, hogy feltárja ezekben az adatokban meghatározott mintázatokat, hogy előre jelezze a jövőbeli értékeket a korábban megfigyelt alapján.

Egy kereskedelmi vállalat eladási statisztikái alapján néztünk meg néhány modell összehasonlító elemzését (9. ábra):

Method	ME	RMSE	MAE	MPE	MAPE	MASE	sMAPE
	-20.022193	327.281259	277.071659	-2.304445	10.32845	0.916385	4.989664
seasonal arima	-20.022193	327.281259	277.071659	-2.304445	10.32845	0.916385	4.989664
	-372.174512	553.416508	454.515075	-15.792259	18.203335	1.50326	7.990454
non seasonal arima	-372.174512	553.416508	454.515075	-15.792259	18.203335	1.50326	7.990454
	39.634227	287.938078	243.495637	0.284865	8.932013	0.805336	4.420576
average seasonal arima & ets	39.634227	287.938078	243.495637	0.284865	8.932013	0.805336	4.420576
	99.290647	303.149926	242.013166	2.874174	8.814185	0.800433	4.521298
seasonal ets	99.290647	303.149926	242.013166	2.874174	8.814185	0.800433	4.521298
	-343.269082	532.710593	437.872678	-14.723818	17.501862	1.448217	7.731356
non seasonal ets	-343.269082	532.710593	437.872678	-14.723818	17.501862	1.448217	7.731356

Forrás: saját készítés

9.ábra Idősor analitikai ML modellek metrikáinak összehasonlítása

A megfelelő gépi tanulási algoritmus kiválasztását követően lehetőség van a modell alkalmazására előrejelzés, becslés céljából.

8. Összegzés

A tanulmányban bemutattuk a mesterséges intelligencia oktatási lehetőségének egy irányvonalát.

A hallgatók a kurzus során megismerkedhetnek azokkal a problémátípusokkal, melyekre hatékony választ tud nyújtani az MI. Tisztában vannak a gépi tanulási modellek belső működésével, hiperparaméterezési lehetőségeivel.

A hallgatók hatékonyan tudnak adott probléma esetén megfelelő eszközt választani, illetve a kurzus megfelelő alapot nyújt a téma további elmélyítésére, az érdeklődés felkeltésére.

Irodalomjegyzék

- [1] <https://www.inf.u-szeged.hu/~rfarkas/ML20/regression.html>
- [2] <https://learn.microsoft.com/hu-hu/training/modules/understand-regression-machine-learning/1-introduction>
- [3] <https://www.inf.u-szeged.hu/~rfarkas/ML20/osztalyozas.html>