

# Zsámbokiné Ficskovszky Ágnes\* – Zsitnyánszky Lilla\*: Mesterséges intelligencia alkalmazása az áruosztályozás területén - Szkepticizmus vagy generációváltás?

---

## Absztrakt

A tanulmány célja feltérképezni és összefoglalni a vámtarifa- és áruosztályozás területét érintő digitális megoldásokat és aktuális irányzatokat. Alapvető kiindulási pontunk az áruosztályozás komplexitása és az arra adott automatizált válaszok és lehetőségek. Egy rövid áttekintést követően bemutatásra kerül a vámtarifa- és áruosztályozás területét jelenleg is érintő piaci kihívásokkal kapcsolatos digitális trendek. Vizsgáljuk a különböző piaci szereplők áruosztályozással kapcsolatos gyakorlati kihívásait. Munkánkban körbejárjuk az árumegnevezések, mint beviteli adathalmaz kritikus témakörét, illetve, hogy ez hogyan hat a manuális és digitális áruosztályozási megoldásokra, valamint a tömeges árubesorolás problematikáját az idő, pontosság és jog-végrehajtói elvárások függvényében.

**Kulcsszavak:** mesterséges intelligencia, áruosztályozás, vámtarifa, WCO, EU

## Abstract

The aim of the study is to explore and summarize the digital solutions and current trends related to customs tariff and customs classification. Our basic starting point is the complexity of customs classification and the automated responses and possibilities available for it. Following a brief overview, we present digital intentions and trends related to market challenges currently affecting customs tariff and customs classification sector. We examine the practical challenges related to Customs classification faced by various market players. In our work, we cover the critical topic of product descriptions as input data sets and how this affects the manual and digital Customs classification solutions, as well the issue of batch classification, taking into account time, accuracy, and law enforcement expectations.

**Keywords:** artificial intelligence, Customs classification, Customs Tariff, WCO, EU

\*\*\*

## Bevezetés

A globális kereskedelem volumene egyre növekszik, a világ árukereskedelmének értéke 12%-kal 25,3 billió USD-ra emelkedett 2022-ben. (WTO (Ed.), 2023)

A növekedés 2023-ban sem állt le, az árukereskedelem 2023 első negyedévében fellendült, és 2022 negyedik negyedévéhez képest hozzávetőleg 100 milliárd dollárral nőtt. (UNCTAD, 2023)

A kereskedelmi forgalom során az árukat világszerte egységes rendszer alapján kódolják, az ún. Harmonizált Áruleíró és Kódrendszer segítségével. A milliányi áru kódolása azonban nem könnyű feladat, rendkívüli szaktudást igényel, amely nem mindig áll a gazdasági szereplők

---

\* Zsámbokiné Ficskovszky Ágnes, dr., pénzügyőr alezredes, mesteroktató, NKE RTK, Vám- és Pénzügyőri Tanszék; ORCID: <https://orcid.org/0000-0001-5560-0079>, [zsambokine.ficskovszky.agnes@uni-nke.hu](mailto:zsambokine.ficskovszky.agnes@uni-nke.hu)

\* Zsitnyánszky Lilla, Head of Customs, Trade AI GmbH, ORCID: <https://orcid.org/0009-0008-8514-1211>, [lilla@traide.ai](mailto:lilla@traide.ai)

rendelkezésére. A helytelen besorolás azonban igen nagy károkat okozhat, mind a gazdálkodók bevételei, mind az egyes államok költségvetése szempontjából, hiszen a kódszámokhoz, azaz vámtarifaszámokhoz (vtsz.) rendelik hozzá nem csak a fizetendő vámok mértékét, hanem az adókat és egyéb fizetési vagy engedélyeztetési stb. kötelezettségeket is ezekhez kapcsolják. Ebből adódóan a helytelenül megállapított vámtarifaszámok rendkívül sok problémát okozhatnak. Ezért is erősödik napjainkban egyre inkább az az igény, hogy valamilyen módon automatikus folyamattá lehessen tenni az áruk tarifális besorolását, melynek legjobb eszköze a mesterséges intelligencia igénybevétele lehetne.

Bár a mesterséges intelligencia, rövidítve: MI (angolul: artificial intelligence, röviden: AI) fogalma az 1950-es években alakult ki, az emberiség régi vágya, hogy egyes feladatainak elvégzését gépekkel tudja helyettesíteni. Írásos formában először Arisztotelész (Kr. e. 384–324) fogalmazta meg Politika c. művében, hogy elképzelése szerint, az általa „automatonnak” nevezett berendezések vezényszóra képesek lesznek a rabszolgák által végzett tevékenységek elvégzésére. Ez az ötlet persze sohasem valósult meg, de egy másik művében, Organon című írásában lefektette a logika és a szillogizmus alapjait, amelyeket később a matematikában és az MI módozatainak kidolgozásában is alkalmazták. (Búzás, 2021)

Az MI a vámeljárások, illetve különösen a tarifális áruosztályozás során való alkalmazása jelentős előnyökkel járna mind a vámhatóságok, mind a gazdálkodók számára, főként a hibás besorolások miatti hátrányok csökkentése szempontjából. Természetesen az MI alkalmazási módjainak kidolgozása nem könnyű feladat, annál is inkább, mert egy nagyon speciális, más területtel, ahol a mesterséges intelligenciát alkalmazzák nehezen összehasonlítható területről van szó.

Másrésről a hatalmas adathalmaz feldolgozása bonyolult és időigényes, az adatelemzések, az előrejelzések annál pontosabbak, minél több adatot dolgoznak fel, és a nagytömegű adatfeldolgozás kivitelezhetetlen megfelelő szoftver nélkül, amely eredményessé teheti a mesterséges intelligencia alkalmazását az adó- és vámügyi eljárásokban. (Nagy, 2021)

Tanulmányunk a mesterséges intelligencia áruosztályozási folyamatban való alkalmazásának tendenciáit, eredményeit, fő összefüggéseit kívánja bemutatni.

## **Az áruosztályozás folyamata**

Az áruosztályozás az a folyamat, amelynek során a kereskedelmi árucikkek vámtarifaszámainak meghatározása történik. A gazdálkodók az import-export folyamatok során vámárnyilatkozatot nyújtanak be, amely tartalmazza a szabad szöveges formátumú termékleírást, és az ahhoz kapcsolódó vámtarifaszámot.

Az árnyilatkozatot benyújtók azonban gyakran helytelenül határozzák meg a kódokat, ami akaratlanul is bekövetkezhet. A helytelen tarifális besorolás ugyanis sok esetben abból adódik, hogy a vámárnyilatkozatot beadó gazdasági szereplőknek (különösen a kereskedelmi vagy vámügynöki tevékenységet végzőknek) maguknak sincs elegendő információjuk az árurol, így kénytelenek azokra az adatokra támaszkodni, amit a beszállítóktól kapnak. Ez utóbbiak azonban sokszor szintén nem rendelkeznek a helyes besoroláshoz szükséges adatokkal, tekintettel arra, hogy ők sem a termék előállítói, tehát nekik sincs a helyes tarifáláshoz szükséges teljesen megbízható, pontos ismeretük az árurol. Ebből adódóan gyakorta jelentkező probléma a hibás, vagy hiányos információkból eredő téves besorolás.

Másrésről a helytelen kódok alkalmazása szándékosan csalárd tevékenységből is adódhat, amely potenciális kockázatot jelent az állami bevételekre nézve. A határokon átnyúló kereskedelemben elkövetett csalás elleni fellépés az egyik kulcsfontosságú prioritás a vámigazgatási szervek számára a nemzetközi bűnözés és a terrorizmus növekvő kockázata, valamint a vámárnyilatkozatok számának emelkedés miatt, amely főként az e-kereskedelem ugrásszerű

növekedésének következménye. Ezért a vámigazgatási szerveknek világszerte gyorsan növelniük kell azon képességüket, hogy pontosabb adatforrásokat keressenek, hogy jobban fel tudják mérni ezeket a kockázatokat, és növeljék az ellenőrzések hatékonyságát. Ennek fontos részét jelenti annak megállapítása is, hogy a bejelentett HR-kódok megfelelőek-e. (Chen et al., 2021)

A tarifális besorolás alapját a nemzetközileg alkalmazott Harmonizált Áruleíró és Kódrendszer (HR) képezi. A Harmonizált Áruleíró és Kódrendszer a Vámigazgatások Világszervezete (World Customs Organization, WCO) által létrehozott és működtetett olyan többcéltű nemzetközi termékosztályozási rendszer, amelyet több mint 200 ország és gazdasági közösség használ, és amelynek jelentőségét mutatja, hogy a nemzetközi kereskedelemben az áruk tarifális besorolása 98%-ban e nómenklatúra alapján történik. (WCO, n.d.) Az Európai Unió vámtarifája, a Kombinált Nómenklatúra<sup>44</sup> (rövidítve: KN) is ezt a rendszert integrálja magába, ezért az uniós vámeljárások során alkalmazott vámtarifaszámok is ezekre a kódokra épülnek.

A tarifális besorolás nem könnyű feladat, ugyanis a pontos besoroláshoz professzionális tudásra van szükség és számos vonatkozó egyéb információforrást is ismerni kell ahhoz, hogy a több mint tízezer különböző kód közül a helyeset ki tudják választani. (He et al., 2021)

Emiatt a gépi tanulást alkalmazó automatizált eszközökre nagy szükség lenne, egyrészt azért, hogy a hatóságokat támogassa a vámkezelési folyamat optimalizálásában, másrészt segítene megelőzni, hogy a kereskedőket a kódok helytelen besorolása miatt megbírságozzák. (Anggoro et al., 2023)

## **A helytelen besorolás következményei**

A 2019-es, a Harmonizált Rendszer (HS) jövőbeli irányával foglalkozó WCO-konferencia egyik fő témája a HS komplexitása volt. Aggodalmukat fejezték ki a HS besorolás nehézségei miatt, különösen azon kis- és középvállalkozások esetében, amelyek lehetőséget kapnak a világgazdaságban való részvételre, de nem feltétlenül rendelkeznek az ehhez szükséges eszközökkel. A Thomson Reuters/KPMG 2016-os felmérésében a válaszadók 95%-a számolt be arról, hogy nehézségei vannak a HS besorolásával kapcsolatban. Ez a nehézség az osztályozási hibák elfogadhatatlan arányában nyilvánul meg. (Rotchin, 2022)

Például 2017-ben a Kanadai Számvevőszék kiadott egy jelentést a vámeljárásokról, amely feltárta, hogy a Kanadába belépő szállítmányok 20 százaléka rossz vámtarifaszámmal rendelkezik, és ennek eredményeként a kormány évi 21 millió dollár bevételtől esett el. Ugyanakkor az importőrök 136 millió dollár visszatérítést követeltek a 2014–2015-ös pénzügyi évben, miután rájöttek, hogy rossz vámtarifaszámot használtak. (Core Ltd., n.d.)

A vámhatóságokon keresztül biztosított adóbevételek aránya világszerte közel 20%, a nyugat-afrikai országokban pedig meghaladja a 40%-ot. Emellett a vámtételek közvetlenül kapcsolódnak az áruk árához, ami befolyásolja azok globális versenyképességét is. (Lee et al., 2021)

### ***Konkrét tarifális példa: túrabakancs áruosztályozása és kapcsolódó vámtételek***

Áruosztályozási példaként két, különböző vámtarifaszám alá tartozó túrabakancs besorolását és a hozzájuk tartozó harmadikországos vámtételeket vizsgáljuk, tekintettel a lábbeli termékek összetett árubesorolási szempontjaira, illetve a termék-megnevezések hiányosságaira és esetleges pontatlanságára.

<sup>44</sup> A Bizottság (EU) 2022/1998 végrehajtási rendelete (2022. szeptember 20.) a vám- és a statisztikai nómenklatúráról, valamint a Közös Vámtarifáról szóló 2658/87/EGK tanácsi rendelet I. mellékletének módosításáról.

	Kereskedelmi árumegnevezés (fiktív példa)	Tarifális besoroláshoz szükséges árumegnevezés	Vámtarifa szám (KN8)	Szerződéses (harmadik országos) vámtétel
a.	“Freizeit braun” túrabakancs, 40-es; Felsőrész: természetes bőr; talp: kiváló minőségű anyag	Szabadidős túrabakancs bőr felsőrésszel és műanyag talppal, beépített védő fém cipőorr nélkül, unisex, 40-es méretben (talpbélés hossza meghaladja a 24 cm-t), lábszárat nem, csak bokát takaró	6403.91.13	8%
b.	“Freizeit schwarz” túrabakancs, 40-es; Felsőrész: textil; talp: kiváló minőségű anyag	Szabadidős túrabakancs textil felsőrésszel és műanyag talppal	6404.19.90	16,9%

A példákon keresztül látható a tarifális besorolás szerepének fontossága, figyelemmel arra, hogy a helytelen áruosztályozás jelentős vámtételbeli különbséget jelenthet, azaz a harmadik országból az EU-ba irányuló behozatalnak fiskális következményei vannak.

Fontos kiemelni azonban, hogy a helytelen árubesorolás nem csak a vámfizetési kötelezettségre van kihatással, hiszen a nómenklátúra rendszerét használják világszerte adók, illetékek kiszabásához, engedélyezéshez, kereskedelempolitikai intézkedések alkalmazásához, statisztikák készítéséhez stb.

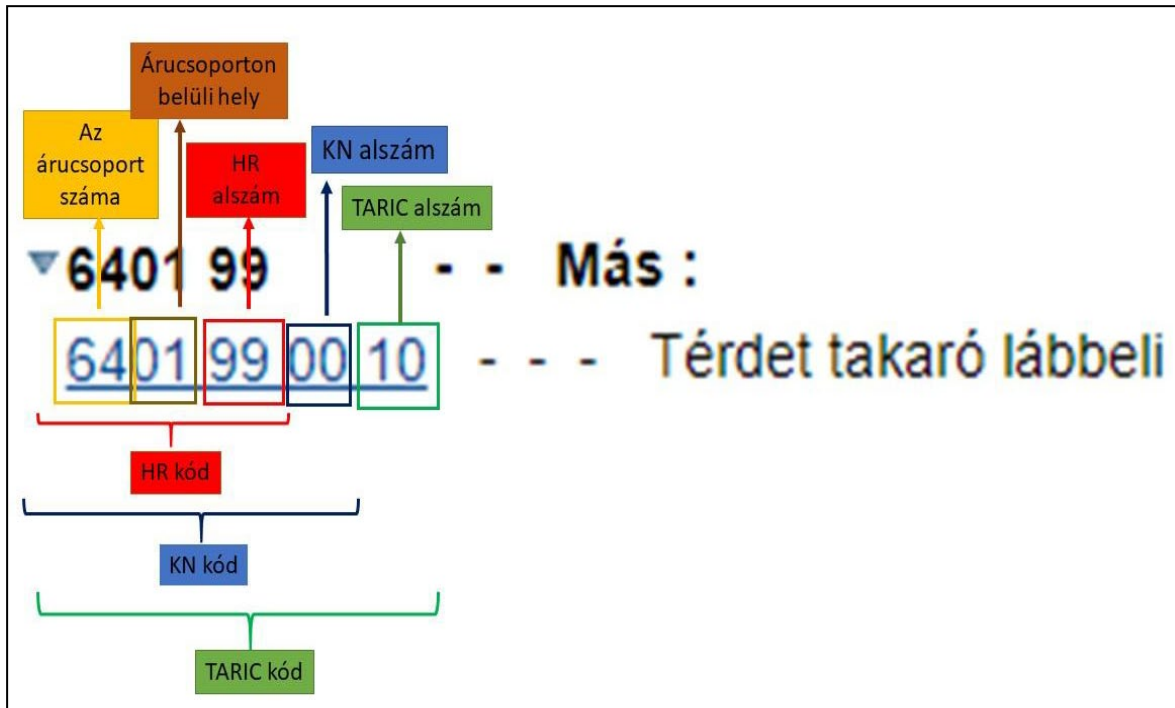
## A tarifális besorolás nehézségei

De miért is olyan nehéz feladat a vámtarifa kódok meghatározása? És miért bonyolult a helyes áruosztályozáshoz szükséges mesterséges intelligenciával való támogatás kifejlesztése? Az áruosztályozási tevékenység egy rendkívül bonyolult folyamat és sokrétű tájékozottságot kíván a tarifális besorolást végzőktől. Több olyan tényező is van, amely jelentősen megnehezíti a megfelelő helyes kód kiválasztását és ennél fogva azt is, hogy az automatizálást, illetve a mesterséges intelligencia igénybevételét megbízhatóan ki lehessen fejleszteni. Ezen tényezők közül a legfontosabbak a következők.

### *A kódok sokszintű tagoltsága*

A HR nómenklatúrában alkalmazott kódok, az ún. HR-kódok alapvetően 6 számjegyből álló vámtarifaszámok, amelyeket további számjegyekkel bővíthetnek igényeiknek megfelelően az azt alkalmazók. Ebből adódóan akár 8-10-12 számjegyből is állhat egy-egy vámtarifaszám. (Ding, Fan, & Chen, 2015). A legtöbb HR-tagállam nyolc számjegyű bontást alkalmaz (HR-6 + 2 = 8), például Kína, Svájc, Norvégia, Ausztrália, Új-Zéland (lásd az 1. táblázatot). Néhány állam nem alkalmaz további felosztást, ami azt jelenti, hogy közvetlenül hat számjegyet (pl. Angola) vagy páratlan számjegyeket használnak (pl. Thaiföldön HR-6 + 1 = 7 vagy a Fülöp-szigeteken HS-6 + 3 = 9; Mauritius külön érdekes, mivel felváltva alkalmaz egy, egy vagy két további számjegyet, amelyeket a 6 HR kódhoz adnak hozzá). (Weerth, 2008) Az Unióban alkalmazott Kombinált Nómenklatúrában a HR 6 számjegyéhez további két számjegy hozzáadásával képzik a KN kódokat, amelyeket az elektronikus vámtarifában, az ún. TARIC rendszerben még további 2 számjeggyel bővítenek. Ez utóbbit TARIC kódnak nevezzük és a vámeljárások során ezeket a kódokat kell alapvetően alkalmazni.

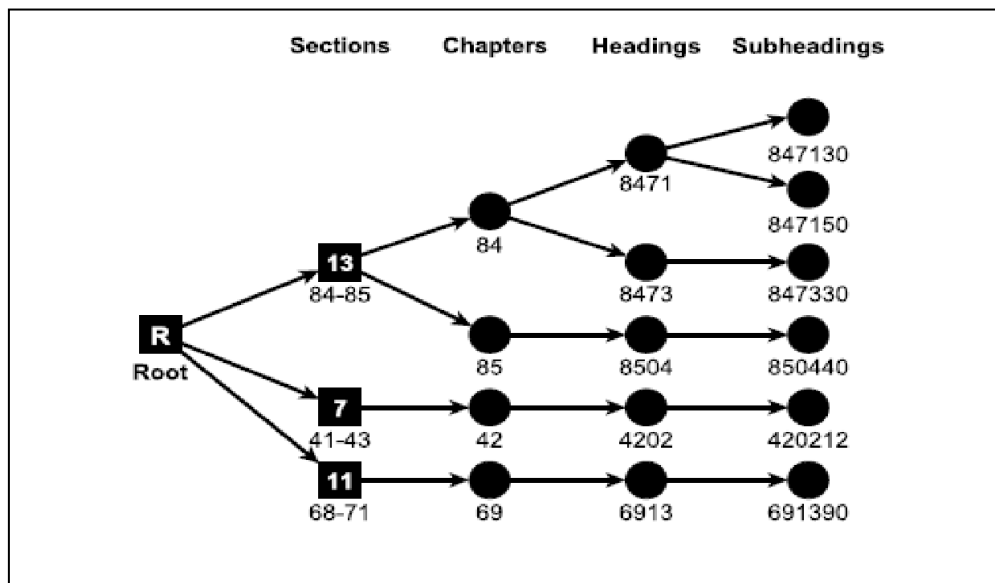
1. ábra: A TARIC kódok felépítése



Forrás: Saját szerkesztés

A HR-nómenklatúra egy olyan ontológiaként ábrázolható, amely csak taxonómiai kapcsolatokat tartalmaz, és egy irányított aciklikus gráfnak (DAG) tekinthető, amelyet az alábbi ábra szemléltet. (Spichakova & Haav, 2020)

2. ábra: A HR-kódok taxonómiája



Forrás: Spichakova & Haav, 2020.

A sok számjegy egyben azt is jelenti, hogy egy-egy árucikkhez időnként rendkívül sok paraméter ismerete szükséges a pontos besoroláshoz.

## ***A nomenklátúra bonyolult struktúrája***

A HR nomenklátúra tehát egy sajátos logika alapján felépülő komplex áruosztályozási rendszer, amely 21 áruosztályba és azon belül 97 árucsoportba tagolva kategorizálja be az árutípusokat, majd ezeken belül vámtarifaszámokra és alszámokra tagozódik tovább. A rendszer lényege, hogy minden létező és a kereskedelemben előforduló árucikket beazonosítható módon kódszámmal lehessen ellátni. A korábban már említett sokrétű alszamos bontásokon túl a besorolást az is nehezíti, hogy az áruosztályozás során önmagából a vámtarifából sokszor nem tudjuk helyesen meghatározni a vámtarifaszámot, mert számos egyéb háttéranyagot is ismernünk kell hozzá.

Például az Európai Unióban az áruosztályozás során minimálisan az alábbi forrásokat kell megvizsgálni egy-egy áru tarifális besorolásával kapcsolatban:

- az alkalmazás, illetve a vámtarifa értelmezésének 6 alapszabálya
- a nomenklatúrában található ún. Megjegyzések
- a Harmonizált Áruleíró és Kódrendszer Magyarázata (HS Explanatory Notes),
- a Kombinált Nomenklatúrához kapcsolódó Magyarázat
- a WCO HR bizottságának áruosztályozási véleményei (Compendium of Classification Opinions)
- az Európai Bizottság áruosztályozással összefüggő rendeletei
- az Európai Bíróságnak a Nomenklátúra, illetve egyes áruk osztályozásának értelmezésével összefüggésben hozott ítéletei
- a Vámkódex Bizottság vám- és statisztikai nomenklátúra szekciójának véleményeit

Ezek egyrészt megkönnyítik egyes áruk besorolását, mert utalnak rá, hogy azt hová kell pontosan besorolni, azonban másrészt nehezítik is, hiszen olyan nagy mennyiségű dokumentumot kell ismernie hozzá az áruosztályozást végzőknek, ami csak magas szintű továbbképzésekkel és éveken át tartó gyakorlással érhető el. E tekintetben is nagy lépés lenne a mesterséges intelligencia alkalmazása, amely ezt a hatalmas mennyiségű információhalmazt is átlátva tudna segíteni a helyes vámtarifaszámok megtalálásában.

Ezen kívül az is gondot okoz, hogy a vámtarifa állandó fejlődésben, változásban van. A 6 számjegyű HR-kódokat ötévente felülvizsgálják. A nemzeti HR-kódok pedig gyakrabban, néha évente többször is változnak. (Ding at al., 2015) Ezért ilyen tekintetben is kihívást jelent egy olyan informatikai háttér kifejlesztése, amely alkalmazkodni tud a folyamatosan változó vámtarifához.

## ***A szöveges meghatározások összetettsége***

A pontos áruleírások alkalmazásának fontossága különösen hangsúlyos a helyes áruosztályozás során. Ahhoz, hogy a mesterséges intelligencia igénybevételével végezzünk áruosztályozást, az árucikkek minden fontos paraméterét tartalmaznia kell a megadott árumegnevezésnek. Természetesen ez nem csak az MI használata esetén áll fenn, mert a tarifális besorolás elsődleges forrása a pontos megnevezés alkalmazása, de a gépi tarifálási módszereknél kulcsfontosságú. Ahhoz tehát, hogy a helyes vámtarifaszámot meg tudjuk határozni minden olyan információt be kell írunk az árumegnevezés megadásakor, ami a vámtarifaszámok és alszámok kiválasztása során döntő lehet.

Ezen kívül az árunyilatkozatokban, fuvarlevelekben vagy egyéb okmányokban feltüntetett szöveges meghatározásokból kell az MI-nek megtalálnia a pontos kódokat, amelynek nehézségei több tényezőtől fakadnak:

- A szöveges leírások sokszor túl rövidek, szűk teret adva ezzel, hogy a pontos besoroláshoz szükséges minden jellemző rendelkezésre álljon.

- A leírások nem követik a természetes nyelv szintaxisát és gyakran hiányoznak egyes mondattani szerkezeti elemek.
- A szövegleírás lehetséges címkéinek száma meglehetősen nagy, akár több mint 5000 is lehet a legmagasabb részletezési szinten.
- Jellemzők olyan kulcsszavak alkalmazása, amelyek a vámtarifában nem csak egy árucikk besorolásánál jelennek meg. Például a „dízel” kifejezés nem csak a 27. árucsoportba tartozó üzemanyagoknál jelenik meg, hanem többek között a 84. árucsoportba tartozó motoroknál, vagy a XVI. áruosztályba tartozó különféle közlekedési módú járművek vámtarifaszámainak szövegében is.
- Nem minden termékmegnevezés olyan gyakori, mint a többi, ami megnehezíti a szélsőségek kiszűrését.
- Gyakorik az elgépelések és a nyelvtani hibák, amelyek értelmezése nehéz a pontosabb szöveggörnyezet elemzése nélkül.
- A szöveges leírások gyakran tartalmazzak speciális, a szállításra jellemző, domain-specifikus szavakat, például az STC (“said to contain”) vagy az FCL (“full-container load”).
- Gyakori a Hapax legomenon, azaz olyan egyedi szavak vagy kifejezések használata, amelyek nem illenek bele az általános szóhasználatba, egyetlen egyszer fordulnak csak elő. (Luppes, 2019)

Itt fontos még megjegyezni, hogy például az eltérő nyelvek (Paramartha at al., 2021) és az eltérő szakmai terminológiák használata (Ding at al., 2015) is problémás lehet, különösen például olyan esetekben, amikor szakszavak meghatározásával igyekeznek pontosítani egy-egy terméket.

Míndezek ahhoz a problémához vezetnek, hogy az egyszerű szöveges keresés, amelyet jelenleg számos HR-kód adatbázis és keresőrendszer használ, nem tud segíteni a kereskedőknek a megfelelő HR-kódok meghatározásában. (Spichakova & Haav, 2020). A HR-kód ellenőrzése végső soron a narratív adatok vagy dokumentumok humán szakértő általi felülvizsgálatát igényli. (Kappler, 2011)

## **Manuális áruosztályozás és mesterséges intelligencia**

Amint azt az előzőekben kifejtettük, az Európai Unióban számtalan az áruosztályozáshoz szükséges forrást kell megvizsgálni és figyelembe venni egy-egy áru tarifális besorolásával kapcsolatban, nem elegendő önmagában a Nomenklátúra szövege és Megjegyzés pontjai.<sup>45</sup> (C-72/21. sz. ügy)

Míndamellet, hogy egy-egy termék manuális áruosztályozása rendkívül időigényes feladat, a piaci szereplők számára ezek a források, tarifális döntések nem mindig elérhetők. Fontos továbbá kiemelni, hogy a besorolandó termékek mennyisége napi szinten több száz, ezer is lehet, ami rendkívüli humánerőforrást, kapacitást kíván meg, köt le. A digitalizáció megjelenésével és fejlődésével különböző automatizációs megoldások születtek az áruosztályozási tevékenység felgyorsításához és megkönnyítéséhez, most ezeket vizsgáljuk a fenti konkrét példán keresztül.

### ***Manuális árubesorolás és az azt támogató források***

Az áruosztályozási folyamat alapvetően a termék-megnevezéssel és termékazonosítással kezdődik. A tarifális árubesorolás egyik gyakorlati nehézsége a tarifális- és kereskedelmi árumeznevezések közötti terminológiai és szemantikai különbség. (Spichakova & Haav, 2020). Annak érdekében, hogy egy termék tarifális besorolása pontos és sikeres legyen, szükség van a termék tarifális szempontból vizsgálandó jellemzőire. Ezen információk sok esetben internetes

<sup>45</sup>Lásd: 2022. április 28-i, C-72/21 sz. ügy, EU:C:2022:312, 29. pont, valamint az ott hivatkozott ítélkezési gyakorlat.

(vásárlói) oldalakon, fotókon, műszaki ábrákon, biztonsági adatlapon, technikai rajzokon, zárt termékinformációs rendszerekben, gyártói megkereséssel stb. érhetőek el.

A példánkban használt két túrabakancs kereskedelmi megnevezése alapján, szükség van további információkra annak érdekében, hogy a pontos tarifális besorolás elvégezhető legyen:

- a) “Freizeit braun” túrabakancs, 40-es; Felsőrész: természetes bőr; talp: kiváló minőségű anyag
- b) “Freizeit schwarz” túrabakancs, 40-es; Felsőrész: textil; talp: kiváló minőségű anyag

A lábbelik áruosztályozása során, figyelembe véve a 64. Árucsoporthoz tartozó megjegyzéseket és a vámtarifaszámok szövegeiben (2022/1998/EU végrehajtási rendelet) foglaltakat, többek között az alábbi szempontokat kell figyelembe venni (Horváthné Zsilka & Zsámbokiné Ficskovszky, 2013)

- Lábbeli anyaga (bőr, textil, műanyag, gumi, fa stb.)
- A talp és felsőrész anyaga (különös tekintettel a két vagy több különböző anyagból álló felsőrész meghatározására)
- Felhasználási mód (sportcipő, sportolásra szánt lábbeli, ortopéd cipők elhatárolása)
- Méret (talpbélés hossza)
- A lábbeli szárának hosszúsága (bokát-, lábszárat takaró vagy egyéb)
- Gyártás módja (gyártástechnológia figyelembevétele)
- Mi minősülhet “résznek” (talp, sarok stb.)

Jelen esetben alkalmazva az 1. és 6. általános értelmezési szabályokat, illetve a listában szereplő kritériumok mellett szükséges még figyelembe venni a 64. Árucsoporthoz tartozó HR és KN Magyarázatot, illetve a 6403 és 6404 vámtarifaszámhoz kapcsolódó egyéb tarifális döntéseket és állásfoglalásokat. Ezen információk nagyrésze Magyarországon a TARIC-WEB<sup>46</sup> oldalon érhetőek el magyar nyelven, az EU Vámkódex Bizottság vám- és statisztikai nomenklátúra szekciójának állásfoglalásai pedig a komitológiai nyilvántartásban<sup>47</sup> illetve a CLASS konzultációs oldalon<sup>48</sup> angol nyelven.

Fentiek függvényében a nevezett lábbeli példák manuális tarifális besorolása a következő kiegészített és módosított áruleírás alapján lehetséges.

- a) Szabadidős túrabakancs bőr felsőrésszel és műanyag talppal, unisex, 40-es méretben (talpbélés hossza meghaladja a 24 cm-t), lábszárat nem, csak bokát takaró: 6403.91.13
- b) Szabadidős túrabakancs textil felsőrésszel és műanyag talppal: 6404.19.90

## Mesterséges intelligencia (AI), gépi tanulás (ML) és mélytanulás (DL)

Az automatizált áruosztályozás és vámtarifaszám meghatározás területén számos innováció és kutatás zajlott és zajlik napjainkban is. 2015-ben például a Singapore-i Vámhatóság publikálta eredményeit az ún. “Background Net” alapú gépi tanulás témában, ahol az árunityilatkozatokban megadott szöveges áruleírásokat alapul véve tettek kísérletet az automatizált árubesorolásra. (Cuaya-Simbro at al., 2022) (Ding at al., 2015)

Egy másik kutatásban a vámtarifaszám javaslatokra több gépi tanulási modellt is vizsgáltak és alkalmaztak meghatározott szöveges áruleírás-adategyütteseket alapul véve. A tesztben alkalmazott gépi tanulási és osztályozási modellek a következők voltak: Naïve Bayes algoritmus, K-Nearest Neighbor algoritmus, Decision Tree (döntéshozatali fa), Random Forest (véletlen

<sup>46</sup> <https://kkk.nav.gov.hu/eles/1/taricweb/>

<sup>47</sup> <https://ec.europa.eu/transparency/comitology-register/screen/documents?lang=hu>

<sup>48</sup> <https://webgate.ec.europa.eu/class-public-ui-web/#/search>



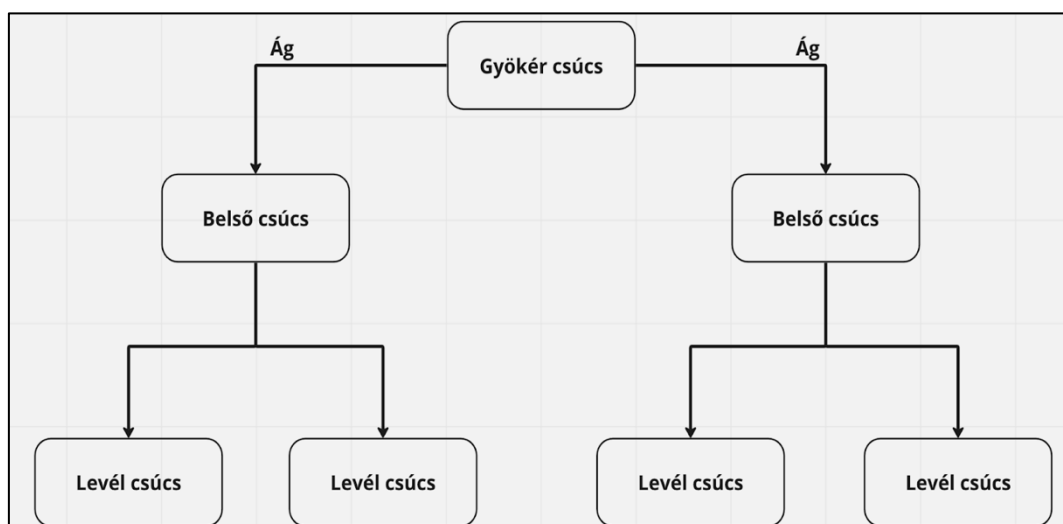
erdők módszer), Linear Support Vector Machine (lineáris támogató vektoros gépi modell) és Adaboost algoritmus. (Altaheri & Shaalan, 2020.)

A legújabb gépi tanulási modellek az ún. mélytanulási modellek, mint például az ún. Neural Machine Translator (NMT: neurális gépi fordítás), Long Short Term Memory (LSTM: hosszú rövidtávú memória), Language Modell (nyelvi modell). (Chen at al., 2021), Sentence Retrieval (mondatlekérdezéses modell) (Lee at al., 2021) avagy képi információkból nyert adatok elemzésére is képes Artificial Neural Network (ANN: mesterséges neurális hálózat) és convolutional neural network (CNN: konvolúciós neurális hálózat) modellek: SqueezeNet, GoogLeNet, ResNet-50, and AlexNet. (Hendrawan, at al., 2019)

A mesterséges intelligencia eszközeinek és modelljeinek alkalmazásával végzett tesztek és kutatások jelentős előrelépést és fejlődést mutatnak az automatizált vámtarifaszám javaslatok tekintetében. Ezen modellek közül hármát mutatunk be röviden: döntéshozatali fa, egyezéshasználati modell és mélytanulási modell.

### *Döntéshozatali fa (decision tree model)*

*3. ábra: A döntéshozatali fa felépítése*



*Forrás: Tan at al., 2011., és Barbosa, 2021. alapján saját szerkesztés*

A döntéshozatali fákat széles körben alkalmazzák osztályozási módszerként az adattudomány területén és közösségeiben, mivel a modell képes összetett problémák megoldására azáltal, hogy az adott problémát vizualizálja, és így lehetővé teszi az adott kérdéskör egyszerűbb értelmezését és elemzését (jelen esetben a tarifális besorolást). A döntéshozatali fa algoritmus, ahogyan a neve is mutatja egy fa struktúrából ún. node-okból (csomópont/csúcs) épül fel. A node-oknak alapvetően három fajtájára kerül az algoritmusban alkalmazásra: root-node (“gyökér csúcs”), internal nodes (“belső csúcs”) és leaf nodes (“levél vagy terminális csúcs”). (Tan at al., 2011) (Barbosa, 2021)

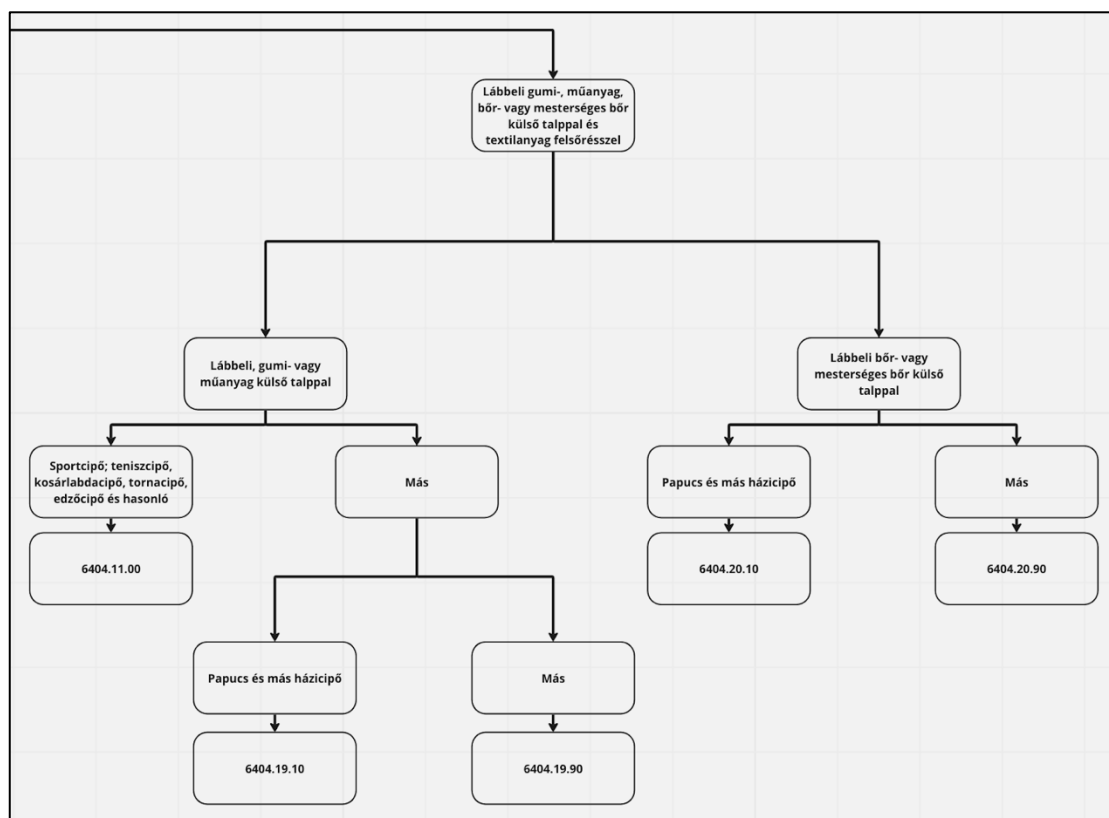
A döntéshozatali fa, mint tarifális osztályozási módszer bizonyos "if-then" előre meghatározott szabályokat alkalmaz, amely szabályok kölcsönösen kizárják egymást. Ezek a szabályok az adatokat az elemekhez kapcsolódó jellemzőkkel kapcsolatos kérdések sorozatának feltevésével osztályozzák, és a tanuló adathalmaz felhasználásával egyenként, szekvenciálisan tanulhatóak.

A döntési fák az ún. “top-down”, felülről-lefelé irányuló hierarchiát alkalmazzák, amely minden esetben egy gyökércsúcsból indul, és ahol az adatokat hasonló értékekkel rendelkező attribútumokat tartalmazó részhalmazokra osztják fel. (Yang, 2019) A döntéshozatali fában a struktúra felépítése befejeződik egy levél csúcsnál, amely egy konkrét döntést jelent vagy

továbbhaladhat egy belső csúcsig. Egy belső csúcs minden esetben egy kérdést vagy attribútumot képvisel, melynek ágai számos további gyerekcsúcsra vezetnek, a lehetséges válaszok számának függvényében. (Kingsford & L. Salzberg, 2008) (Song & Lu, 2015) A csúcsok közötti ágak (“branches”) a fa azon részei, amelyek a csúcsok közötti kapcsolatot teremtik meg.

A példaként nevesített lábbelik áruosztályozására - figyelembe véve a lábbelikre vonatkozó tarifális kritériumokat - a következő döntéshozatali fa definiálható (tekintettel a 6403 és 6404 vámtarifaszámokhoz tartozó részletes attribútumokra, a tanulmányban a döntéshozatali fa egy részlete kerül csak szemléltetésre).

4. ábra: A döntéshozatali fa felépítése 6404 vtsz., részlet



Forrás: Saját szerkesztés

Fontos megemlíteni, hogy a definiált “általános” tarifális attribútumok automatizálása önmagában nem mindig elegendő, figyelemmel a beviteli adathalmazra (pl. kereskedelmi árumegevezés) illetve arra, hogy az attribútumok további magyarázatra szorulhatnak. Így például, hogy tarifális értelemben mit jelent a “sportcipő” kifejezés a 6404.11.00 vtsz.-nál, szükség van a 64. Árucsoportokhoz tartozó Alszámos megjegyzés, valamint a vámtarifaszámhoz tartozó KN Magyarázat rendelkezéseit is figyelembe venni, adott esetben előre definiálni és ennek fényében a belső- vagy levél csúcsokat a fa struktúrában meghatározni.

A piaci szereplők és vállalatok sok esetben alkalmazzák tarifális gyakorlatukban a döntéshozatali fa modellt figyelembe véve a besorolandó termék sajátos tulajdonságait összehangolva a vámtarifa követelmény-rendszerével. (Barbosa, 2021) Hátránya a modellnek azonban, hogy az előre meghatározott szabályok minden esetben konkrét termékekre, adott esetben termékkörre alkalmazandók, és amennyiben egy tarifális döntés, vámtarifaszám változás következtében vagy egy új besorolandó termék megjelenése miatt a struktúra módosítására vagy új struktúra

elkészítésére van szükség, az igen időigényes feladat, illetve nehézkes a régi modell naprakészen tartása.

### ***Egyezéshasználati vagy részleges egyezés modell (Fuzzy Matching/Approximate String Matching)***

Az áruosztályozás területén alkalmazott egyik legelterjedtebb automatizációs törekvés az ún. egyezéshasználati vagy részleges egyezési modell (a köznyelvben gyakran fuzzy matching-nek is nevezik) piaci bevezetése volt, amely során nyelvi algoritmusokat használtak/használnak a szöveges árumegnevezések (illetve a modell alkalmazása kiterjeszhető pl. egyedi árucikk azonosító számokra is) egyezőségének vizsgálatára. Ebben az esetben a modell által vizsgálandó információ lehet pl. az árumegnevezés és/vagy vállalat specifikus árucikk azonosító. Fontos, hogy ezek az algoritmusok nemcsak teljes (100%) nyelvi/karakter egyezés esetén tesznek javaslatot, hanem részleges egyezés (fuzzy match) esetén is ajánlhatnak fel vámtarifaszámot, alapul véve egy bizonyos százalékos egyezést. A részleges egyezés mértékét a modellek százalékban mérik. Jellemző, hogy a százalékos küszöbérték előre beállítható, amely fölötti egyezésekre tesz csak javaslatot az adott modell. (Proford, n.d.) Ahhoz, hogy egy ilyen megoldás az áruosztályozási tevékenységben segítséget nyújtson, szükséges a gépi tanulás, azaz a modellt előzetesen trenírozni kell olyan korrekt adatokkal és információkkal (árumegnevezés/cikkszám és a hozzájuk tartozó vámtarifaszám ún. ground truth adatok), amelyeket alapul véve és felhasználva a modellnek viszonyítási alapja lesz a vámtarifaszám javaslatokhoz.

Konkrét példánkon keresztül szemléltetve a modell alkalmazása a következőképpen néz ki.

Ground truth adatnak a következő információ került megadásra:

- a. “Freizeit braun” túrabakancs, 40-es; Felsőrész: természetes bőr; talp: kiváló minőségű anyag: 6403.91.13

Kereskedelmi árumegnevezés (fiktív példa)	ML által javasolt vámtarifaszám	Egyezés a ground truth adat alapján
“Freizeit braun” túrabakancs, 40-es; Felsőrész: természetes bőr; talp: kiváló minőségű anyag	6403.91.13	100%
“Freizeit braun” túrabakancs, 40-es	6403.91.13	80%
“Freizeit braun” túrabk., 40	6403.91.13	80%
“Freizeit braun”	6403.91.13	73%

- b. “Freizeit schwarz” túrabakancs, 40-es; Felsőrész: textil; talp: kiváló minőségű anyag: 6404.19.90

Kereskedelmi árumegnevezés (fiktív példa)	ML által javasolt vámtarifaszám	Egyezés a ground truth adat alapján

“Freizeit schwarz” túrabakancs, 40-es; Felsőrész: textil; talp: kiváló minőségű anyag	6404.19.90	100%
“Freizeit schwarz” túrabakancs, 40-es	6404.19.90	78%
“Freizeit schwarz” túrabk., 40	6404.19.90	82%
“Freizeit schwarz”	6404.19.90	73%

Fontos tehát, hogy a modell itt nem vizsgál áruosztályozási összefüggéseket, csupán az adott szöveg- vagy adathalmazzal, azaz a megadott árumeagnevezésekkel és/vagy árucikk azonosítóval és azok különböző verzióival dolgozik. Ez a modell olyan esetekben alkalmazható sikeresen, ahol az adott árukörnél a tarifális szempontokat elegendő egyszer megvizsgálni és ground truth adatként megadni, illetve ahol az árumeagnevezések repetitív módon jelennek meg, tehát kizárható annak a lehetősége, hogy az adott árumeagnevezés kapcsán egyéb vámtarifaszám alkalmazás is szóba jöhet.

### ***Mélytanulási modell***

A mélytanulási modell a legújabb az áruosztályozási területen alkalmazott technológia. A legalapvetőbb különbség a gépi tanulás és mélytanulási technológia között, hogy míg a gépi tanulás ún. historikus adatokkal dolgozik, nem vizsgál mélyebb rétegekben összefüggéseket, addig a mélytanulás ún. neurális hálót alkalmaz, azaz összefüggéseket vizsgál a döntés és javaslatételhez. (HWSW, 2022) Tekintettel arra, hogy maga az áruosztályozási tevékenység összefüggések vizsgálatából és döntések sorozatából áll, az ezen a technológián alapuló modellek jelentős minőségbeli előrelépést mutatnak. Annak érdekében, hogy egy áruosztályozásra specializálódott mélytanulási modell alkalmazása sikeres legyen, itt is szükség van a modell adatokkal való trenírozására. De ebben az esetben a gép már vizsgálja a különböző összefüggéseket/mintákat a megadott ground truth információk - azaz vámtarifaszám és az ahhoz rendelt árumeagnevezés - alapján, így a gép az általános értelmezési szabályokat és tarifális döntéseket is figyelembe véve tud javaslatot tenni a vámtarifaszámra.

Úttörő kutatás és a technológia legfrisebb innovációja, hogy ezzel a modellel már képek, fotók alapján is kinyerhetők adatok és információk. Ezzel az előrelépéssel az egyik legkritikusabb probléma oldható meg a jövőben, mivel az áruosztályozáshoz szükséges információk jelentős hányada képi információ forrásként hamarabb elérhető és alkalmasabb adatokat tartalmaz, mint a szöveges és sokszor hiányos leírások (Cuaya-Simbroski et al., 2022)

### **Összefoglalás**

Tanulmányunk célja az áruosztályozás mint komplex területen alkalmazott aktuális digitalizációs folyamatok, innovációk, kutatások és modellek bemutatása volt. Vizsgáltuk, hogy milyen különböző az áruosztályozási tevékenység felgyorsítására, hatékonyságnövelésre bevezetett kísérletek és kutatások voltak/vannak folyamatban. Megállapítható, hogy a piaci szereplők és gazdálkodók számára továbbra is kihívás a besorolandó árukhoz kapcsolódó tarifális

információk megszerzése, sok esetben csupán az áru kereskedelmi megnevezése áll rendelkezésre az áruosztályozás során. Fontos tény és nehezítő tényező az árubesorolásra rendelkezésre álló szoros időkorlát a besorolandó áruk mennyiségének függvényében.

A mesterséges intelligencia, illetve az azon alapuló informatikai modellek alkalmazásával jelentős előrelépés történt a folyamatok megkönnyítésére és felgyorsítására, fontos azonban megjegyezni, hogy ezek a támogatások/modellek és rendszerek, emberi erőforrás igénybevétele nélkül nem működnek. Egy bármilyen modell által javasolt vámtarifaszám és/vagy ahhoz kapcsolódó árubesorolási háttér minden esetben megerősítésre kell kerülnön a felhasználó által.

A modellek azonban jelentős segítséget nyújtanak az áruosztályozási folyamat felgyorsítására, hibák kiszűrésére.

A jelenlegi mélytanulási modellek fejlesztése biztató és izgalmas eredményeket ígérnek.

## Irodalomjegyzék

- [1.] A Bizottság (EU) 2022/1998 végrehajtási rendelete (2022. szeptember 20.) a vám- és a statisztikai nomenklatúráról, valamint a Közös Vámtarifáról szóló 2658/87/EGK tanácsi rendelet I. mellékletének módosításáról. [http://data.europa.eu/eli/reg\\_impl/2022/1998/oj](http://data.europa.eu/eli/reg_impl/2022/1998/oj)
- [2.] Altaheri, F., & Shaalan, K. (2020). Exploring Machine Learning Models to Predict Harmonized System Code. In M. Themistocleous & M. Papadaki (Eds.), *Information Systems* (pp. 291–303). Springer International Publishing. [https://doi.org/10.1007/978-3-030-44322-1\\_22](https://doi.org/10.1007/978-3-030-44322-1_22)
- [3.] Anggoro, A., Corcoran, P., De Widt, D., & Li, Y. (2023). Using DistilBERT to Assign HS Codes to International Trading Transactions. In *World Conference on Information Systems and Technologies*. Pisa, Italy. <https://orca.cardiff.ac.uk/id/eprint/155999/>
- [4.] Barbosa, I. (2021). *Using Machine Learning to classify HS codes*. Master's Dissertation. Universidade do Porto, Porto. <https://repositorio-aberto.up.pt/bitstream/10216/134987/2/484345.pdf>
- [5.] Búzás Gy. M. (2021). A mesterséges intelligencia története. *Central European Journal of Gastroenterology and Hepatology*, 7(3), 121–127. <https://doi.org/10.33570/CEUJGH.7.2.121>
- [6.] C-72/21. sz. ügy: A Bíróság (nyolcadik tanács) ítélete 2022. április 28. Augstākā tiesa (Senāts) (a Legfelsőbb Bíróság, Lettország) által 2021. február 4-én benyújtott előzetes döntéshozatal iránti kérelem - SIA "PRODEX" kontra Valsts ieņēmumu dienests (Állami Adóhatóság) <https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=CELEX:62021CJ0072&from=EN>
- [7.] Chen, H., van Rijnsoever, B., Molenhuis, M., van Dijk, D., Tan, Y.-h., & Rukanova, B. (2021). The use of machine learning to identify the correctness of HS Code for the customs. *IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)*, 1–8. <https://doi.org/10.1109/DSAA53316.2021.9564203>
- [8.] Chen, X., Bromuri, S., & Van Eekelen, M. (2021). Neural Machine Translation for Harmonized System Codes prediction. *2021 6th International Conference on Machine Learning Technologies*, 158–163. <https://doi.org/10.1145/3468891.3468915>
- [9.] Core Ltd. (n.d.) The high price of using incorrect tariff codes. <https://www.coresupplychains.com/blog/the-high-price-of-using-incorrect-tariff-codes/>
- [10.] Cuaya-Simbro, G., Hernandez-Vera, I., Ruiz, E., & Gutierrez-Fragoso, K. (2022). Automatic Tariff Classification System using Deep Learning. *International Journal of Advanced*

- Computer Science and Applications (IJACSA)*, 13(7), 904–911.  
<https://doi.org/10.14569/IJACSA.2022.01307105>
- [11.] Ding, L., Fan, Z., & Chen, D. (2015). Auto-Categorization of HS Code Using Background Net Approach. *Procedia Computer Science*, 60, 1462–1471.  
<https://doi.org/10.1016/j.procs.2015.08.224>
- [12.] He, M., Wang, X., Zou, C., Dai, B., & Jin, L. (2021). A Commodity Classification Framework Based on Machine Learning for Analysis of Trade Declaration. *Symmetry*, 13(6), 964. <https://doi.org/10.3390/sym13060964>
- [13.] Hendrawan, Y., Widyaningtyas, S., & Sucipto, S. (2019). Computer vision for purity, phenol, and pH detection of Luwak Coffee Green Bean. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 17(6), 3073.  
<http://doi.org/10.12928/telkomnika.v17i6.12689>
- [14.] Horváthné Zsilka, J., & Zsámbokiné Ficskovszky, Á. (2013). *A Kombinált Nomenklátúra II. Egyetemi jegyzet*. Nemzeti Közzolgálati Egyetem.
- [15.] HWSW. (2022. október 10.). Mi a különbség a deep learning és a klasszikus machine learning között? <https://www.hwsz.hu/hirek/65256/deep-learning-online-kepzes-tanfolyam-gyakorlat.html>
- [16.] Kappler, H. (2011). Reversing the trend: low cost and low risk methods for assuring proper duty payments. *World Customs Journal*, 5(2), 109–122. (Letöltés dátuma: 2023.06.27.)
- [17.] Kingsford, C., & L. Salzberg, S. L. (2008). What are decision trees? *Nature Biotechnology*, 26(9), 1011–1013. <https://doi.org/10.1038/nbt0908-1011>
- [18.] Lee, E., Kim, S., Kim, S., Park, S., Cha, M., Jung, S., Yang, S., Choi, Y., Ji, S., Song, M. & Kim, H. (2021). Classification of Goods Using Text Descriptions With Sentences Retrieval. <https://doi.org/10.48550/ARXIV.2111.01663>
- [19.] Luppés, J. (2019). Classifying Short Text for the Harmonized System with Convolutional Neural Networks [Master Thesis]. Radboud University.  
[https://www.ru.nl/publish/pages/769526/luppés\\_-\\_classifying\\_short\\_text\\_for\\_the\\_harmonized\\_system\\_with\\_convolutional\\_neural\\_networks.pdf](https://www.ru.nl/publish/pages/769526/luppés_-_classifying_short_text_for_the_harmonized_system_with_convolutional_neural_networks.pdf)
- [20.] Nagy, Z. (2021). Mesterséges intelligencia lehetőségei az adó- és vámügyi eljárásokban. In Czene-Polgár, V., Csaba, Z., Zsámbokiné Ficskovszky, Á. & Szabó, A. (Eds.), „*Tradíció, tudomány, minőség*” 30 éves a Vám- és Pénzügyőri Tanszék - *Tanulmánykötet* (pp. 226–233). Magyar Rendészettudományi Társaság Vám- és Pénzügyőri Tagozata.  
<https://10.37372/mrttvpt.2021.2.15>
- [21.] Paramartha, I. G. Y., Ardiyanto, I., & Hidayat, R. (2021). Developing Machine Learning Framework to Classify Harmonized System Code. Case Study: Indonesian Customs. 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), 254–259.  
<https://doi.org/10.1109/EIConCIT50028.2021.9431888>
- [22.] Proford. (n.d). A technológiával segített fordítás alapvetései. Professzionális Fordításszolgáltatók Egyesülete. <http://proford.hu/wp-content/uploads/tudasbazis/A%20technol%C3%B3gia%C3%A1val%20seg%C3%ADtet%20ford%C3%ADt%C3%A1s%20alapvet%C3%A9sei.pdf>
- [23.] Rotchin, R. (2022). What impact is technology having on efforts to improve HS classification efficiency and accuracy? *WCO News* 97(1), 53–57.  
[https://mag.wcoomd.org/uploads/2022/03/WCO\\_News\\_97\\_2022\\_1.pdf](https://mag.wcoomd.org/uploads/2022/03/WCO_News_97_2022_1.pdf)

- [24.] Song, Y.-Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*, 27(2), 130–135. <https://doi.org/10.11919/j.issn.1002-0829.215044>
- [25.] Spichakova, M., & Haav, H.-M. (2020. December). Application of Machine Learning for Assessment. *Baltic J. Modern Computing*, 8(4), 698–718. <https://doi.org/10.22364/bjmc.2020.8.4.13>
- [26.] Tan, P.-N., Steinbach, M., & Kumar, V. (2011). Osztályozás: Alapfogalmak, döntési fák és modellek kiértékelése (Ispány, M., Ford.). In Bevezetés az adatbányászatba. Panem Könyvkiadó. (Eredeti kiadás: 2005). <https://gyires.inf.unideb.hu/KMITT/a04/ch04s03.html>
- [27.] UNCTAD. (2023). Global trade update June 2023. Geneva: United Nations Commission on Trade and Development. [https://unctad.org/system/files/official-document/ditcinf2023d2\\_en.pdf](https://unctad.org/system/files/official-document/ditcinf2023d2_en.pdf)
- [28.] Weerth, C. (2008). Structure of Customs Tariffs Worldwide and in the European Community. *Global Trade and Customs Journal*, 3(6), 221–225. <http://hdl.handle.net/10419/144712>
- [29.] World Customs Organization. *What is the Harmonized System (HS)?* <https://www.wcoomd.org/en/topics/nomenclature/overview/what-is-the-harmonized-system.aspx>
- [30.] World Trade Organization (Ed.). (2023). *Global trade outlook and statistics*. World Trade Organization. [https://www.wto.org/english/res\\_e/publications\\_e/trade\\_outlook23\\_e.htm](https://www.wto.org/english/res_e/publications_e/trade_outlook23_e.htm)
- [31.] Yang, X.-S. (2019). *Introduction to Algorithms for Data Mining and Machine Learning*. Academic Press. <https://doi.org/10.1016/C2018-0-02034-4>