SZAKDOLGOZAT

Horváthné dr. Kovács Bernadett

2024

Magyar Agrár- és Élettudományi Egyetem

Szent István Campus

Műszaki Intézet

Adattechnológus-adatelemző szakmérnök szakirányú továbbképzési szak

MINTÁZATFELTÁRÁS NEM FELÜGYELT TANULÁSI MÓDSZERREL ROBOTIZÁLT ÜZEMI TEJTERMELÉSBEN

Belső konzulens:	Tarr Bence Gyula egyetemi tanársegéd
Belső konzulens intézete/tanszéke:	Műszaki Intézet Mérnökinformatika Tanszék
Külső konzulens:	-
Készítette:	Horváthné dr. Kovács Bernadett

Szent István Campus, Gödöllő

Tartalom

1. Bevezetés

Az automatizált fejőrendszerek, az egyedi állatmegfigyelő eszközökkel együtt, hozzájárulnak a tejtermelés menedzsment döntéseinek támogatásához azáltal, hogy számos adatot rögzítenek és továbbítanak a számítógépes rendszer felé. Az ilyen integrált IoT rendszerben, amelyet precíziós technológiaként is ismerünk, nagyon nagy számú, nagyon változatos és nagyon gyors ütemben keletkeznek az adatok (big data), lehetővé téve historikus elemzést, összefüggések vizsgálatát, illetve előrejelzések készítését amellett, hogy valós idejű adatrögzítéssel és feldolgozással gyorsan és hatékonyan képes figyelmeztetést küldeni nem várt adatmintázatok észlelésekor.

A robot fejőrendszerek dinamikus terjedése és használata miatt egyre több nagy adat keletkezik, szaktanácsadók és gazdálkodók együttesen keresik a választ, hogyan lehetne a termelés hatékonyságát (ezzel egyidejűleg a környezet terhelését, vagy az élelmiszerbiztonságot) javítani. A cloud technológia az adattárolási és számításkapacitás ugrásszerű növekedését eredményezte, az adattudomány módszertana elérhetővé teszi nem ismert, akár szokatlan összefüggések, vagy azok mögött álló okozatok feltárását is. A mesterséges intelligencia, ezen belül a gépi tanulási eljárások alkalmazásával a robotfejőgépek adattömege is elemezhető, a modellek támogathatják az üzemi szintű döntéshozást, de akár általánosítható ismereteket is eredményezhet.

Egy összefoglaló tanulmány (Shine és Murphy, 2022) a tejtermelési adatokat gépi tanulás módszerekkel vizsgáló publikációkat 20 évre kiterjedően tekintette át, és csoportosította a kutatás célja (témája) alapján, illetve feltárta az alkalmazott gépi tanulási módszerek megoszlását. Megállapította, hogy jelentősen terjed (2018 óta ötszörösére növekedett) a neurális hálózati algoritmusokat alkalmazó kutatások száma, míg a döntési fa algoritmusokat és a statisztikai regressziós algoritmusokat alkalmazók száma is mintegy háromszorosára nőtt. A főbb tématerületek, amelyek vizsgálatára a két évtizednyi kutatások fókuszáltak: a takarmányozás, állattartás, egészségügy, állatok viselkedése, fejés és erőforrás-gazdálkodás.

1.1. A dolgozat célkitűzései

Üzemi termelési adatok rendelkezésemre álló tejtermelési és szenzoradatai alapján célom, hogy olyan termelési csoportokat azonosítsak, amelyek alapján feltárhatók az egyedek termelés perzisztenciáját, illetve relatív termelési pozícióját leíró mintázatok. A dolgozatban a célkitűzés elérése érdekében 6 kutatási feladatot azonosítottam:

1, A tudomány jelenlegi állásának vizsgálata az automatizált, robot fejőrendszerek big data adatainak gépi tanulási eljárások alkalmazásával történő hasznosítása területén. Célja a tanulmányok áttekintő feltérképezése, tipizálása, a dolgozat témakörébe illeszkedő tanulmányokra szűkítés. **– szakirodalmi áttekintés**

2, A gépi tanulási eljárások áttekintése a vizsgálati célnak megfelelően. **– szakirodalom/módszertan**

3, A témában megjelent eredeti kutatási eredményeket közlő és áttekintő tanulmányok eredményei alapján az adatelemzés megtervezése, vizsgálati célok, paraméterek, modellek kijelölése. **- módszertan**

4, Benchmark tanulmány alapján az adatgyűjtés megtervezése, az adatbázis megszervezése, a kutatás modelljének felvázolása. **- módszertan**

5, Adattisztítás és az elemzések elvégzéséhez szükséges program elkészítése, a feltáró elemzések futtatása, az eredmények tárolása. **- eredmények**

6, Eredmények kommunikálása, a feltárt összefüggések gyakorlati hasznosíthatóságának vizsgálata, validálása, a kutatás korlátai és kitekintés. **– eredmények, következtetések**

A dolgozat felépítése a fenti témaköröket követi.

2. Szakirodalmi áttekintés

A szakirodalmi áttekintés - részben - a vonatkozó a tanulmányok áttekintő feltérképezése, tipizálása, a dolgozat témakörébe illeszkedő tanulmányokra szűkítés érdekében valósult meg; másrészt pedig a gépi tanulási eljárásokat tekintettem át a vizsgálati célnak megfelelően.

2.1. Az automatizált, robot fejőrendszerek big data adatainak gépi tanulási eljárások alkalmazásával történő hasznosítása

A tudomány jelenlegi állásának vizsgálata az automatizált, robot fejőrendszerek big data adatainak gépi tanulási eljárások alkalmazásával történő hasznosítása területén képezte a dolgozat kiindulási alapját. Ennek érdekében szisztematikus irodalomelemzést végeztem, amelynek módszertana az alábbi részfeladatokat jelentette:

a, tudományos publikációk, tanulmányok összegyűjtése a témában,

b, a találatok alapján szövegelemzést végeztem az automatizált fejőrendszerek és a gépi tanulási eljárások kapcsolati hálójának feltárására,

c, a kulcsszavak kapcsolati hálója segítségével azonosítottam a központi témakört, illetve leszűrtem azokat a tanulmányokat, amelyek a dolgozat témaköréhez szorosan kapcsolódtak,

d, a szűkebb találati lista segítségével léptem a későbbiekben tovább a második kutatási feladatra.

2.1.1. Az irodalomkutatás módszertana

Az Elsevier kiadó tudományos köteteiben kereső Web of Science keresőmotorral, a cikkek absztraktjában, címében és kulcsszavaiban (TS mező) előre meghatározott kulcsszavak megfelelő kombinációjával (1. keresőalgoritmus) végeztem el a keresést.

1. keresőalgoritmus

TS=((((("robot*" OR "automated") AND "milking") OR ("milking" AND "machine")) AND (("ML" OR ("machine" AND "learning") OR ("classification" OR "regression" OR "cluster*" OR ("neural" AND "network") OR "fuzzy")) OR (("artificial" AND "intelligence") OR "AI")))

A találati lista közel 500 tételt tartalmazott, amelyet bibTeX txt formátumban mentettem el későbbi szoftveres feldolgozás érdekében.

A duplikátumok eltávolítását követően, illetve szükséges technikai átalakítások után a találatokat szövegelemző szoftverben (VOSviewer) dolgoztam fel. A szövegelemzés során a teljes találati listában meghatározott kulcsszavak együttes előfordulásának hálózati térképén (co-occurence, treshold=4) kirajzolódó klaszterek jelentették a találatok további szűkítéseinek

alapját. A túl általános, illetve több alakban megjelenő kifejezések (78 db) eltávolítása érdekében ún. thesaurus állományt készítettem és használtam fel az együttes előfordulások hálózati térképéhez. A szerzői közösségek hálózata és a közös hivatkozások hálózata további adalékokkal szolgált a témakörhöz szorosan kapcsolható tanulmányok azonosításában. A vizuális áttekintés során a célom az volt, hogy azonosítsam azokat a tanulmányokat, amelyek a robotfejőgépekből származó adatok gépi tanulási eljárásokkal történő feldolgozásával foglalkoztak.

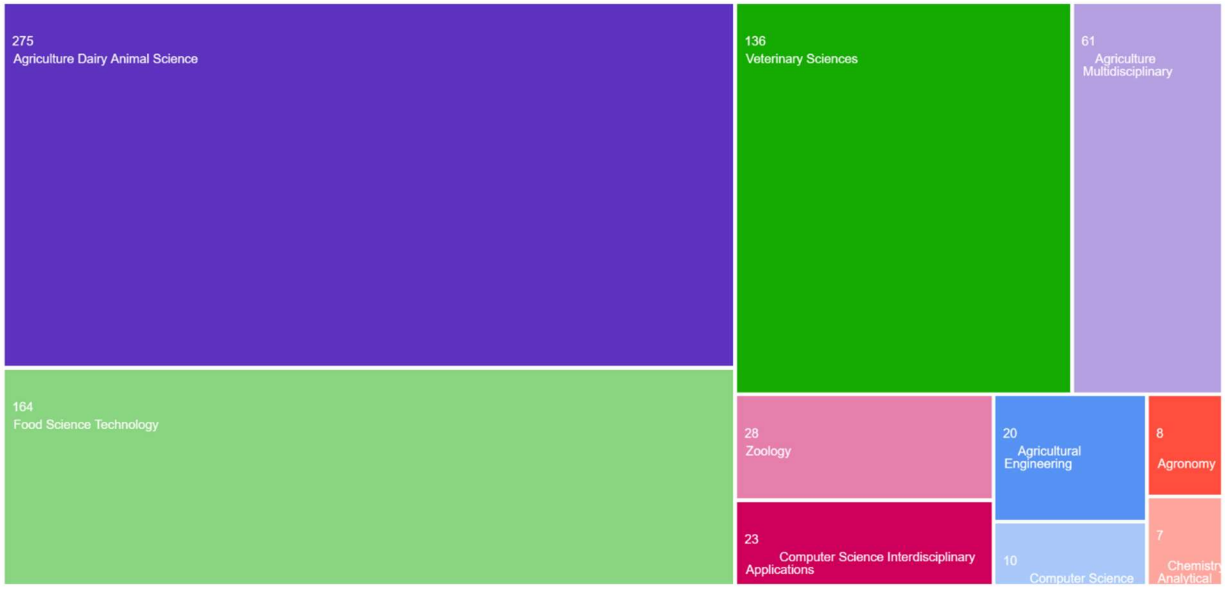
A keresés szűkítéséhez a kulcsszavak együttes előfordulásának klasztereiben feltárt, a témával szorosan nem összefüggő kifejezések kizárásával jutottam el. Ezen találati lista szolgált a későbbeikben a vizsgálat elmélet modelljének (módszertan, adatok, eljárások) alapjául.

2.1.2. Az irodalomkutatás eredményei

Áttekintés

A fent megadott (1.) keresőalgoritmus segítségével összegyűjtött tanulmányokra jellemző, hogy elsősorban az Agriculture Dairy Science (56,4 %), majd a Food Science Technology (33,6 %) és a Veterinary Science (27,9 %) szerzőközösségének cikkei foglalkoznak a fenti összefüggéssel (1. ábra).

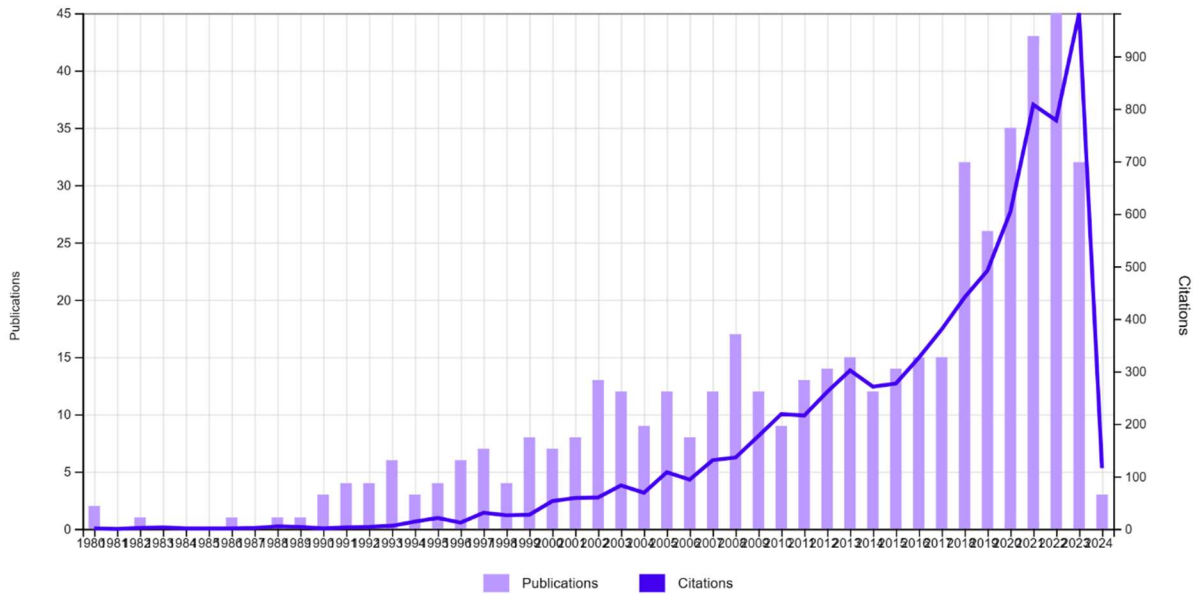
1. ábra: A teljes találati lista folyóirat szerinti megoszlása



Forrás: webofscience.com vizualizációja a találatokra

A tanulmányok megjelenésének és azokra vonatkozó hivatkozásszámok időbeliségét vizsgálva megállapítható továbbá, hogy a robotfejés és gépi tanulás összefüggésében írt tanulmányok népszerűsége a hivatkozásszám exponenciális növekedésével alátámasztható (2. ábra), a 2000-es évekhez képest megtízszereződött a téma iránti tudományos érdeklődés.

2. ábra: A teljes találati lista publikációinak és hivatkozásainak évenkénti megoszlása



Forrás: webofscience.com vizualizációja a találatokra

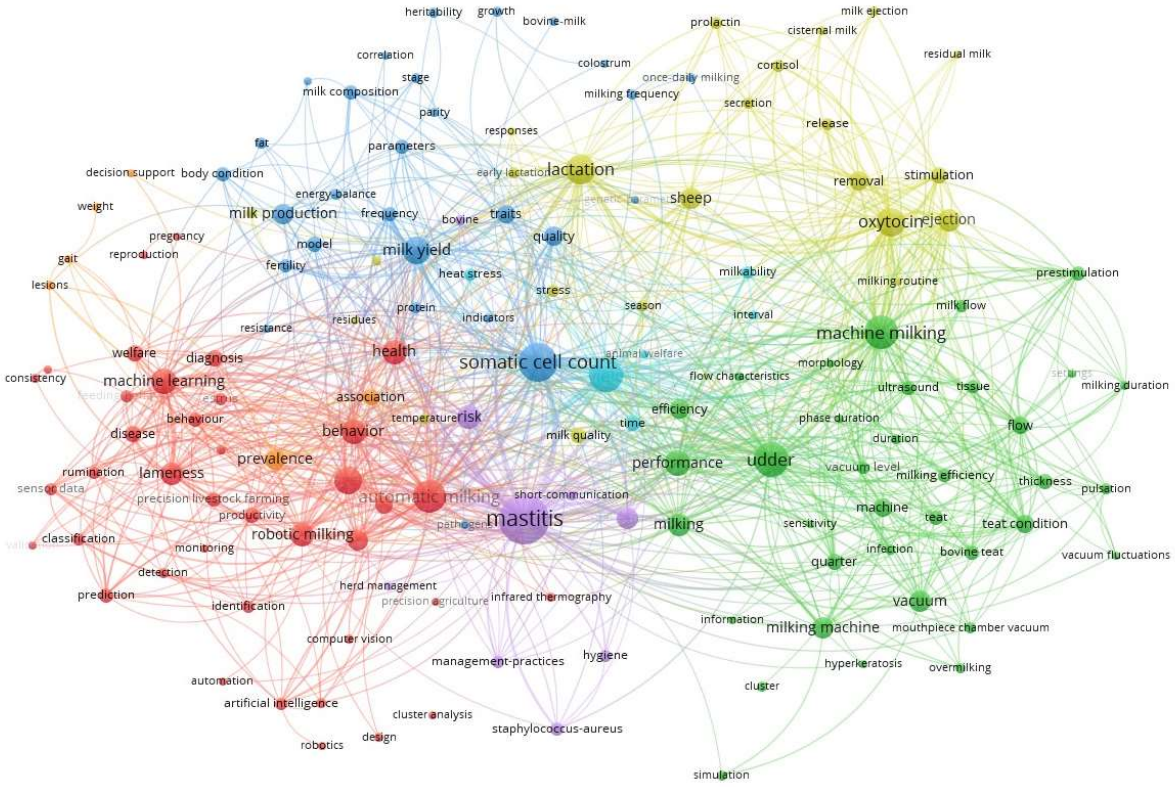
A következő lépésekben a szövegelemző szoftver segítségével vizsgáltam a találatok TS mezőjében (tehát témakör, absztrakt, cím és kulcsszavak listáiban) szereplő kulcsszavak együttes előfordulását (co-occurence, treshold=4), majd pedig a tanulmányok szerzőközösségét (co-authorship, n=92, treshold=3) és közös hivatkozási hálózatát (co-citation, n=70, treshold=10).

2.1.3. A szakirodalom hálózatmodellezése

A Thesaurus állomány segítségével 78 kifejezést helyettesítettem nagyon egyező tartalom, de eltérő írásmód miatt, illetve iktattam ki, mert a vizsgált témakörben alapvető szavak és kifejezések voltak. Az adatok VOSviewerben történő elemzéséhez már ezt a javított kifejezéslistát használtam.

Először a kifejezések együttes használatának (legalább 4 különböző cikkben fordul elő) hálózatát vizsgáltam (n=147 kifejezés az 1817-ből). Az együttes előfordulások hálózata alapján 5 modul rajzolódik ki (3. ábra).

3. ábra: A kulcskifejezések együttes előfordulása a tanulmányokban (n=147, treshold=4)



Forrás: saját szerkesztés

A lila színnel jelzett modul alapvetően az állománymenedzsment, egészségmenedzsment téma köré csoportosul; a részhálózat fő eleme a tőgygyulladás (mastitis), ami arra utal, hogy leginkább a tőgyegészség fenntartásával, higiéniával kapcsolatos menedzsment témájú tanulmányokat takar ez a modul.

A kék színnel jelölt modul csomópontjai a tejtermelés mennyiségi és minőségi paraméterei és a testösszetétel, valamint környezeti indikátorok tekintetében kapcsolódnak a szomatikus sejtszám kifejezéshez.

A sárga modul a laktáció biológiai jellemzőihez kapcsolódik, illetve ebben az almodulban jelenik meg a juh állatfajként.

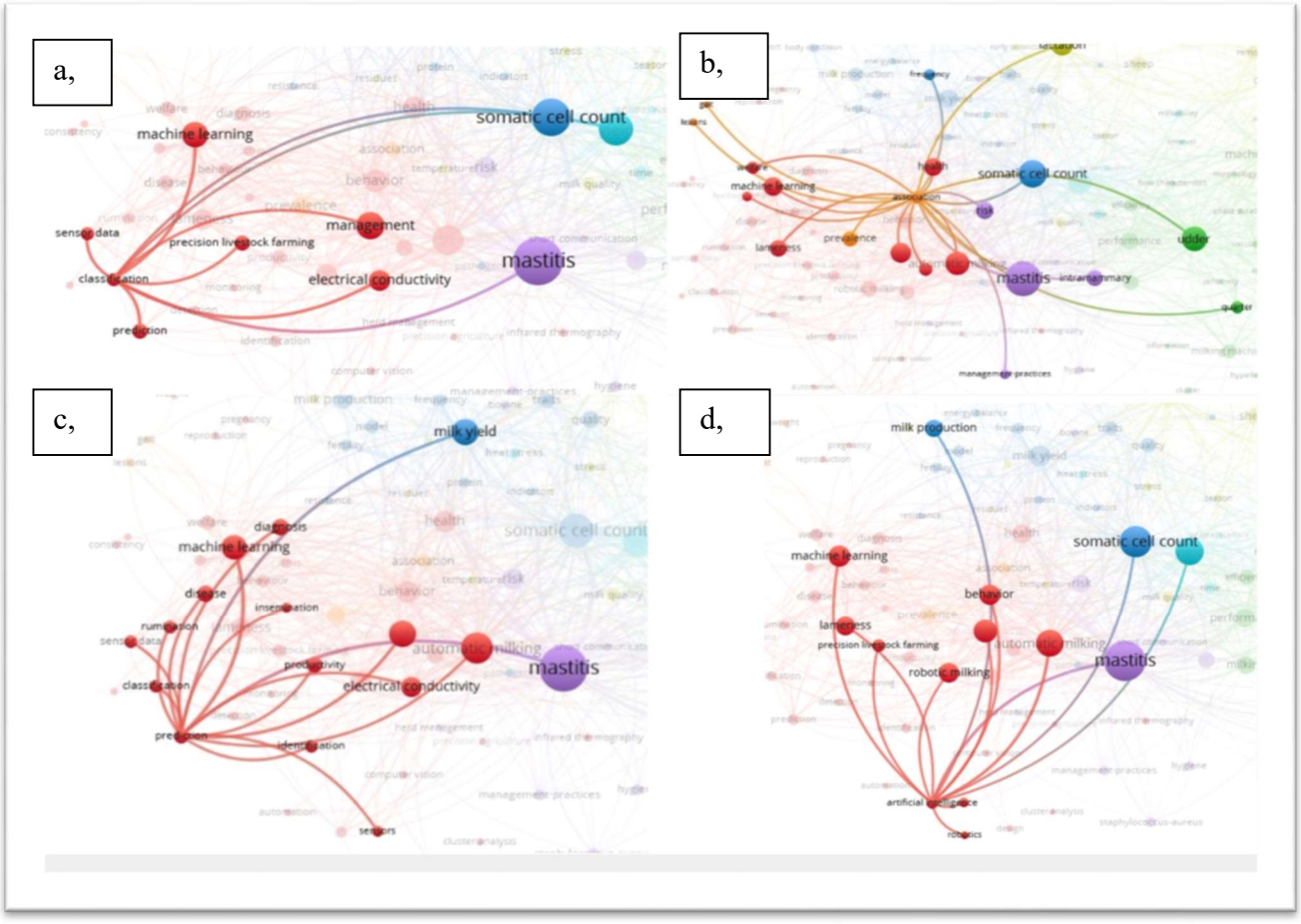
A vizsgálatom szempontjából legérdekesebb két csoport közül a piros színű modulban találhatók a gépi tanulás és ahhoz kapcsolódó eljárások (asszociáció, predikció, osztályozás, stb.) a robotfejés és automatikus fejés kifejezések kiemelkedő jelenléte mellett nehezen tipizálható egyéb kulcsszavakkal, ami arra utalhat, hogy számos terület vizsgálata merülhetett fel a mesterséges intelligencia és a robotfejés viszonylatában.

A másik jelentős modul a zölddel jelzett, a gépi fejhetőséghez kapcsolódó kulcsszavakat tömörítő klaszter, amelyben jellemzően a fejés fizikai paraméterei rajzolódnak ki. Ugyanakkor itt nem kapcsolódik gépi tanulási vagy mesterséges intelligencia kifejezés.

A szövegelemzés hálózattérképét felhasználva szűkíthető a feldolgozandó tanulmányok listája. A szűkítés érdekében vizsgáltam a legrelevánsabb modulon belüli kapcsolatok alakulását (4. ábra a-d). A részhálózatok kirajzolják a mesterséges intelligencia, vagy gépi tanulás és ezen eljárások (pl. osztályozás, előrejelzés, asszociáció, klaszterezés) kapcsolatát a robotizált fejőrendszerekkel a tanulmányokban.

Így az osztályozás, asszociáció, előrejelzés kifejezések a szenzoradatokhoz, például a robotfejőgépen (pl. vezetőképesség) és különböző, az állatok viselkedésének monitorozása során (kérődzés, sántaság) keletkező adatokhoz kapcsolódik. A mesterséges intelligencia kifejezéssel nem specifikus, hanem általános (pl. robotfejés, mastitis) szavak társulnak, a témakör tágabb megnyilvánulását jelzi ez az összefüggés.

4. ábra: A hálózat kiemelt részmoduljai: a, osztályozás - b, asszociáció – c, előrejelzés – d, mesterséges intelligencia

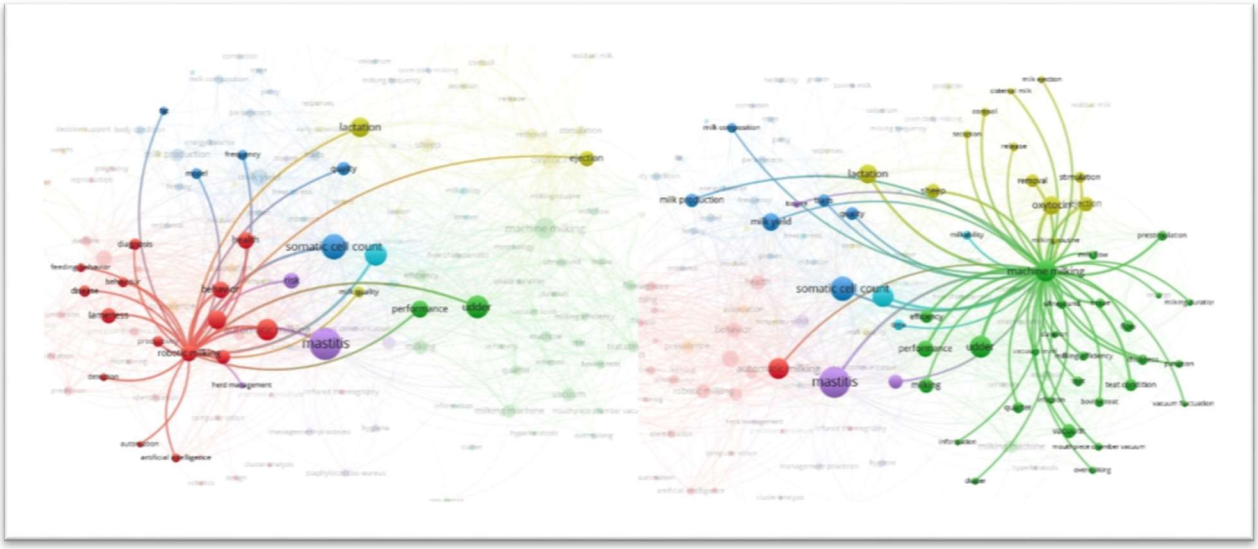


Forrás: saját szerkesztés

Habár a robotfejés és a gépi fejés szinonimái is lehetnek egymásnak, a két kifejezés körül teljesen más hálózati térkép rajzolódik ki (5. ábra a, b); az utóbbi a már említett, fiziológiás fejhetőségi paraméterekhez kapcsolódik (zöld modul), míg az előbbi a robotfejés és viselkedés kapcsolatára utal.

5. ábra: A robotfejés és a gépi fejés kifejezések együttes előfordulási kapcsolati részhálózatai

a, *b,*



Forrás: saját szerkesztés

A fent megfigyelt elkülönült részrendszerek alapján módosítottam a keresőalgoritmusokon a találatok későbbi szűkítése érdekében. Ezek eredményeire itt nem térek ki részletesen; a legszűkebb találati lista alapján a 3. kutatási feladatra léptem tovább.

A tanulmányok elemzésének következő szempontrendszere a vizsgálni kívánt kifejezéseket tartalmazó tanulmányok szerzőközösségének feltárása, valamint az egymásra épülő műhelyek azonosítása volt a robotfejés és gépi tanulás témakörben.

A cél érdekében először a közösen publikáló szerzőket, majd a közösen hivatkozott cikkek hálózatát vizsgáltam.

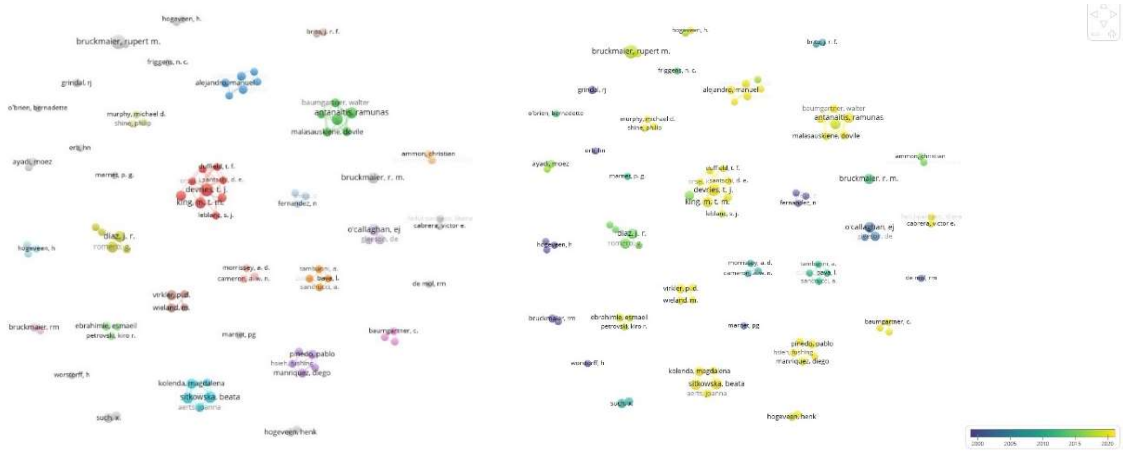
Szerzői és hivatkozási hálózatok

Kicsi közösségekre utal, hogy a mintegy 1777 szerző közül a legalább három dokumentumban történő megjelenés kritériumának 92 szerző felel meg. A tématerület tehát elaprózottnak tekinthető.

Mivel nem volt célom a kis szerzőközösségek kizárása, a közös szerzői hálózat (co-authorship, n=92, treshold=3) elkülönülő, kisméretű csoportokba tagozódik (6. ábra a, b). Az összesen 33 klaszterben 121 link, kapcsolat található.

6. ábra: A tanulmányok közös szerzői hálózata: a, és dinamikája: b

a, *b,*

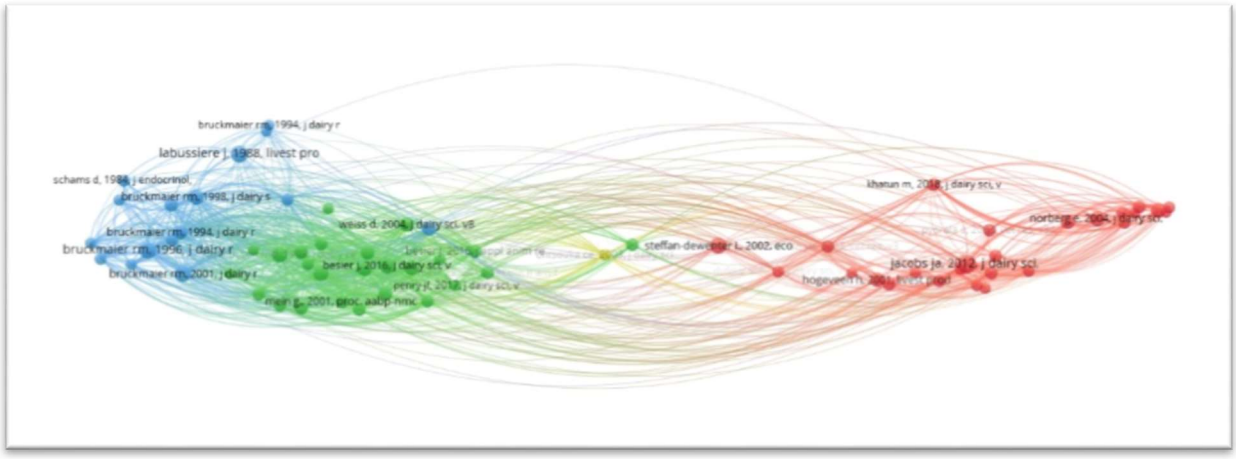


Forrás: saját szerkesztés

A leginkább kiterjedt szerzői csoportok újabbak is, kevés számú kötödéssel a korábbi időszak szerzőire. Mindazonáltal ezek a csoportok jelezhetik az adott tématerület felívelő dinamikáját. A tématerületben rejlő publikációs potenciál további vizsgálatára alkalmas bibliometrix R modul felhívhatja a figyelmet a területre.

A feldolgozott korábbi tanulmányok előfordulásának hálózata (co-citation) a legalább 10 dokumentum által hivatkozott tanulmányok közös előfordulását ábrázolja (6. ábra). Az összes előfordulásból (11529 db) mintegy 70 esetében találunk legalább 10 hivatkozást. Egy-egy csomópontba azokból a csomópontokból futnak be élek, ahol a két kapcsolódó tanulmányban legalább 10 közös hivatkozás található. Az élek száma 1095, a tanulmányok száma 68, azaz átlagosan 16 a páronként közösen hivatkozott tanulmányok száma.

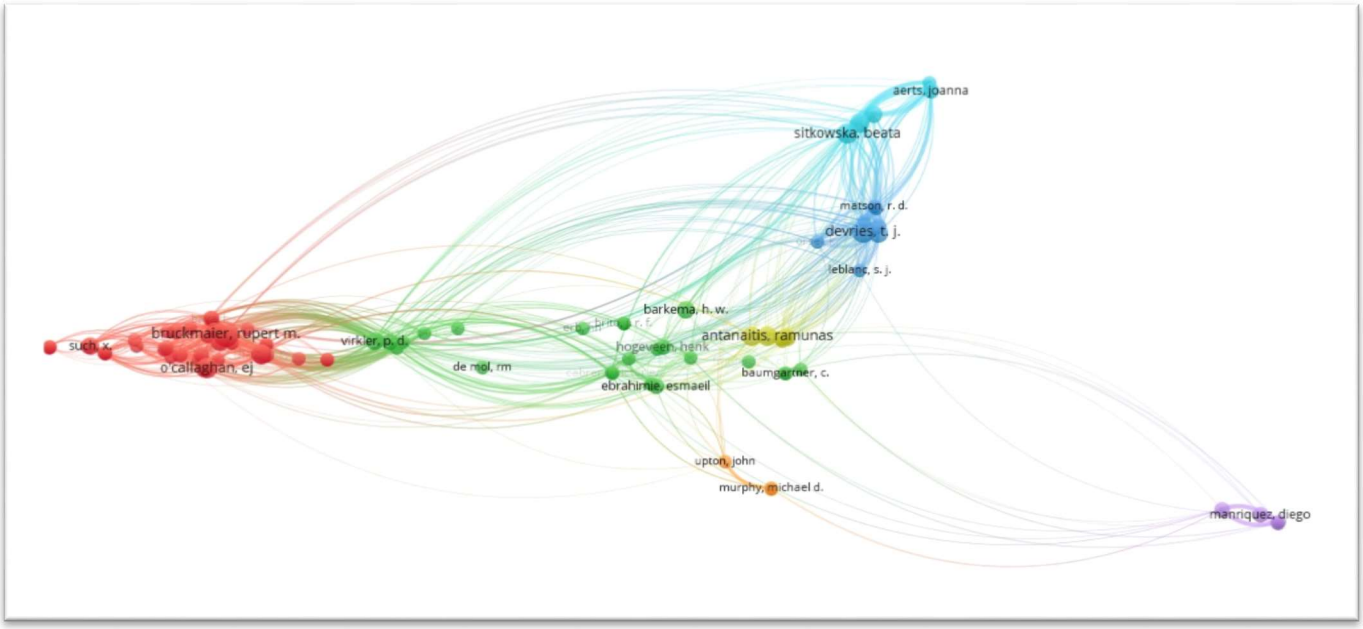
7. ábra: Közös hivatkozás hálózat (co-citation, n=70, treshold=10)



Forrás: saját szerkesztés

A szerzők közötti bibliográfiai kapcsolatok (bibliograpfic coupling, authors, treshold=3, n=92) egymást szorosabban hivatkozó 2-2 közösséget és lazábban kapcsolódó továbbiakat jelenít meg (8. ábra).

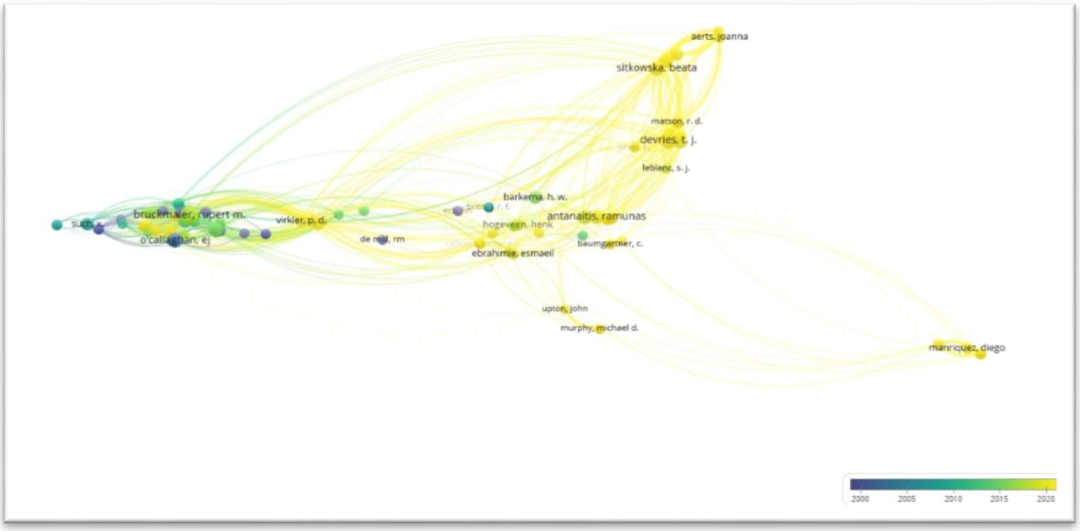
8. ábra: Bibliográfiai párosok (bibliograpfic coupling, authors, treshold=3, n=92)



Forrás: saját szerkesztés

Időbeli dinamikáját tekintve elmondható, hogy a fiatalabb publikációk szerzői jelentősen nagyobb számosságú bibliográfiai kapcsolatban állnak egymással, mint a régebbiek (9. ábra).

9. ábra: Bibliográfiai párosok - időtérkép (bibliograpfic coupling, authors, treshold=3, n=92)

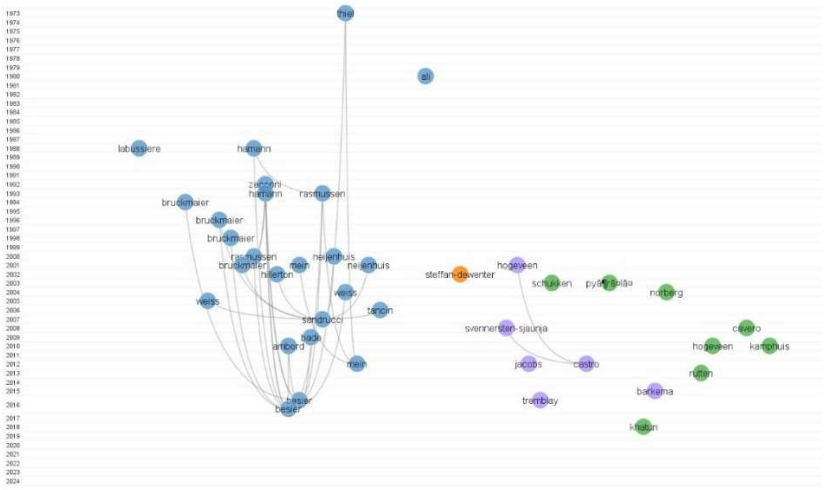


Forrás: saját szerkesztés

A szerzőközösségek dinamikája

A citnetexplorer vizualizációjának segítségével egyidejűleg bemutatható a hivatkozási közösségek fejlődése és ezek „alapító", a tématerületet először feldolgozó szerzői. A teljes találati lista (1. keresőalgoritmus) alapján 5 klaszter rajzolódik ki (10. ábra), amely 1973-2024 között keletkezett 543 publikációt tömörít legalább 10-es elemszámú csoportba (137 publikáció nem került csoportosításra a kritérium miatt).

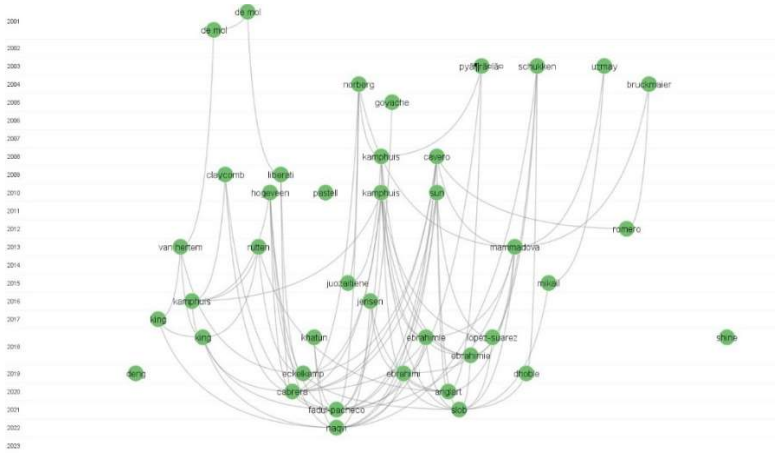
10. ábra: A hivatkozási közösségek időbelisége



Forrás: saját szerkesztés

Mélyfúrással feltárható, hogy a második legnagyobb (n=86) hivatkozási csoportba tartozó szerzők tanulmányai adják a dolgozat témájához közelebb álló publikációs kört (11. ábra).

11. ábra: Mélyfúrással elérhető közösségi hálózat időbeli kapcsolatai



Forrás: saját szerkesztés

A citnetexplorer eszköztára megalapozhatja egy feltáró, visszafejtő munka lehetőségét is a témakörben elterjedt definíciók használatának és eredetének fonalára vonatkozóan.

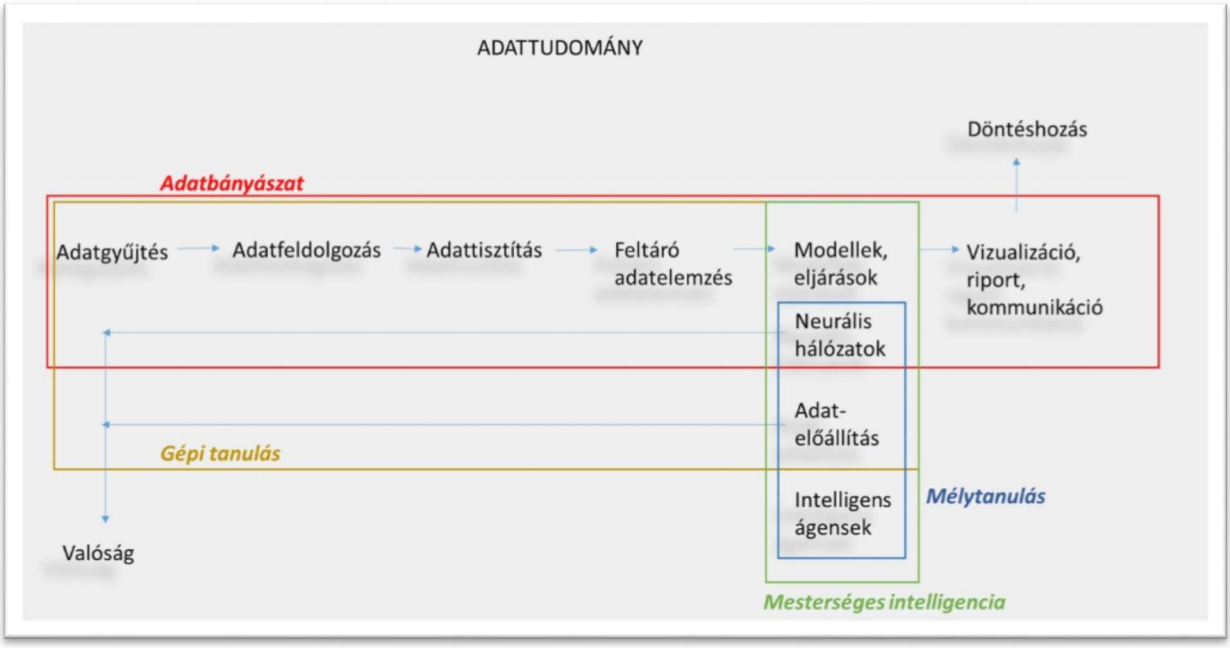
2.2. Gépi tanulási eljárások

A gépi tanulási eljárások fejezet tárgyalja a módszertani alapokat és a feltárt irodalomban azonosított eljárásokat.

2.2.1. A gépi tanulási eljárások

A gépi tanulás olyan mesterséges intelligencia eljárások területéhez tartozik, amelyek statisztikai algoritmusok fejlesztésén és tanulmányozásán keresztül adatokból tanulnak és általánosítanak, jellemzően explicit utasítások nélkül végezve el feladatokat. A gépi tanulás kifejezés Samuel 1959-es megalapozó munkáját követően jelenik meg a szakirodalomban (Koza és munkatársai, 1966). Ágens alapú tanulás specifikus területeként is hivatkozzák, amikor az ágens egy számítógép: megfigyel adatokat, ezekre támaszkodva modellt épít, majd ezt a modellt a világra vonatkozó hipotézisként használja, illetve szoftverként, amellyel problémákat tud megoldani (Tarr, 2024). A gépi tanulás definícióját az adattudományokon belül az adatok gyűjtése, rendszerezése és feldolgozása vonatkozásban (12. ábra) is kiterjesztik (Horváthné et al, 2024).

12. ábra: A gépi tanulás környezete az adattudományok terében



Forrás: Horváthné et al, 2024 (2. o) nyomán

Az adatgyűjtés során a szükséges nyers adatokat kell elérhetővé tenni, összekötni, tisztítani, hogy elegendően nagy és reprezentatív mintát adjunk a rendszer tanításához és teszteléséhez. Az adatok előfeldolgozása nagy szereppel bír annak érdekében, hogy jó teljesítményű modellt

nyerjünk. A jellemzőkinyerés a feltáró adatelemzés feladatköre. A gépi tanulási eljárás megválasztása az adatbázistól és az elvárt feladattól függ, továbbá iterálható: különböző beállítások vagy módszerek eredményességének összehasonlításával a legjobb modell elérése a célunk (Farkas et al., 2020).

Az adatfeldolgozás és előrejelzési modellek módszertana, valamint az előre jelzett adatok minőségének vizsgálata alapján számos eljárást ismerünk, ezeket egyfajta csoportosítás szerint kategóriára oszthatjuk. A felügyelt tanulási eljárások esetében (pl. osztályozás, regressziós eljárások) mindig tudjuk a modell alapjául szolgáló adatok ún. címkéit, azaz, hogy mi az eredményváltozó értéke. Nem felügyelt tanulási eljárások alatt azokat az algoritmusokat értjük, amikor a számítógép a kimenetek ismerete nélkül (ún. címkézetlen) adatok mintázatát tárja fel (pl. klaszterezési eljárások). Harmadik csoportként megkülönböztetjük a megerősítő tanuló eljárásokat. Előbbi három csoporton felül megjelennek még a félig-felügyelt eljárások, valamint egyes szerzők külön tárgyalják a mélytanuló eljárásokat (habár ezek alapvetően klasszifikációt valósítanak meg).

Felügyelet nélküli tanulási módszerek

A felügyelet nélküli, vagy nem felügyelt eljárások a címkézetlen adatok mintázatának, összefüggésének definícióját eredményezik. Ezek leggyakrabban (Farkas et al, 2020) a

a) klaszterezés, melynek során egy adatbázis címkézetlen egyedeinek olyan csoportjainak keressük, hogy az egy csoportban levő egyedek hasonlóbbak lesznek egymáshoz, mint a más csoportban levőkhöz; illetve a

b) dimenzió csökkentés, amelynek célja, hogy egy adatbázis egyedeinek jellemzőire adjon egy olyan transzformációt, amiben az egyedek egy jóval kisebb dimenziószámú térben írhatóak le (statisztikai eljárások közül a faktorelemzés és a PCA sorolható ide pl.).

A klaszterezési eljárások statisztikai alaptípusai is többfélék (k-közép [mean, median], hierarchikus, spektrális), és ezek is futtathatók különböző paraméterezéssel (pl. távolságmetrika) is (StataCorp, 2023).

Az egyik legjellemzőbb eljárás a k-közép módszer, amelynek során célunk, hogy az egyedeket tetszőleges k darab csoportba soroljuk. Az eljárás hogy egy klaszterben levő pontok közelebb vannak saját klaszterük középpontjához (centroid), mint bármely más klaszter középpontjához. A klaszterezés eredményességének mutatói az ún. validációs mutatók (pl. Shilouette mutató).

Felügyelt tanulási eljárások

Formálisan megfogalmazva Tarr (2024) alapján az eljárás az alábbiak szerint definiálható.

Adott egy N db bemeneti – kimeneti példapárból álló **tanítóminta-halmaz**: (x_1, y_1), (x_2, y_2), . . . (x_N , y_N) , ahol mindegyik párt egy ismeretlen y = f(x) függvény generált. Találjunk meg egy olyan **h függvényt**, amelyik közelíti a keresett f függvényt.

A **h függvényt** a világra vonatkozó hipotézisnek nevezzük. A lehetséges függvények H halmazából, a hipotézistérből választjuk. A hipotézistér lehet előzetesen ismert, de lehet, hogy feltáró adatelemzést kell végeznünk annak érdekében, hogy előzetes ismereteket szerezzünk és meghatározzuk a választható függvények halmazát. Az is előfordulhat, hogy több hipotézisteret is kipróbálunk, hogy a legjobb megoldást találjuk meg. Feltáró adatelemzéssel, amelyek statisztikai (StataCorp, 2023) leíró, illetve vizualizációs eszközök lehetnek.

Az egyik leggyakoribb felügyelt gépi tanulási feladat az **osztályozás** (classification). Itt előre adott egy kategória-/osztályhalmaz és a célunk, hogy modellt/döntési szabályokat tanuljunk a tanító adatbázis alapján, ezt az eljárást pedig új adathalmazon alkalmazva, a nem címkézett megfigyeléseket is osztályozni tudjuk (Farkas 2020). Az osztályozási eljárást kategória típusú kimenet esetén alkalmazzuk; lehet bináris vagy több kimenetű (multi-class) típusú.

Folytonos numerikus változó kimenettel rendelkező adatbázis esetén **regressziós** eljárásokkal jutunk olyan modellhez, amely egy új adathalmaz jellemzői alapján előrejelzést ad a várható kimenetre.

Az, hogy a modellünk a tanulómintába nem tartozó, új bemeneti adatokon hogyan teljesít, a hipotézis jóságának mértéke. Azok a metrikák, amelyek kifejezik, hogy az ún. teszthalmaz (x_i,y_i) mintapárjain mennyire pontos predikciós teljesítményt nyújt a modell, az általánosíthatóságát adják meg. Ebből az aspektusból beszélünk alul- és túlillesztett modellekről. A modell torzítása az a variancia, amely a különböző tanítóhalmazokra illeszkedésre jellemző, míg a modell varianciáját a tesztadatokra történő illeszkedés pontossága adja (általánosíthatóság). A modell előrejelzési teljesítőképességét az ún. tévesztési mátrixból származtatott mérőszámok (pontosság – precision, fedés – recall, találati arány - accuracy , ROC, AUC) segítségével fejezzük ki (Orováné, 2024).

Osztályozási eljárások

Leggyakoribb osztályozási eljárások modelljei a döntési fa osztályozó, amely jellemzően kisebb és diszkrét tulajdonság esetén elterjedten használt modell, valamint az ún. lineáris gépek (Farkas et al., 2020) csoportja, ahol nagyszámú és folytonos változó tulajdonság mellett készíthetünk osztályozókat.

A döntési fa osztályozó modell egy gráf, amelynek csúcsaiban a megfigyelések egy-egy jellemzője található. A csúcsok elágazásai, az élek a szülőtől a gyerekhez, azaz a tulajdonság értékeihez vezetnek. A leveleken (tovább már nem elágazó alkotói a fának), osztálycímkék állnak. A fa elágazási mentén jutunk egy megfigyelés jellemzőin keresztül a megfigyelés osztályáig (Tarr, 2024). A predikció az az osztálycímke, ahova a belső tulajdonságok elágazásain eljutottunk. Nem kategória típusú jellemzők esetében a döntési fát regressziós fának nevezzük; az elágazások valamilyen folytonos változó alapján „terelik" az osztályok felé a döntési útvonalat. Vizuális megjelenése miatt jól értelmezhető az ilyen döntési fa modell. A klasszikus döntési fákat akkor érdemes választani, ha viszonylag kevés (pár száz), elsősorban diszkrét jellemzőnk van és feltételezzük, hogy bonyolultabb kapcsolat van közöttük. A döntési fák hátrányainak (kis elemszám, kevés és elsősorban diszkrét tulajdonság ismerete) megoldására vezették be az ún. erdő osztályozókat. Ezeknél a modell nem egyetlen döntési fa, hanem több viszonylag kis méretű döntési fa, amiknek a "szavazataiból" alakul ki az erdő végső predikciója. A kisebb fák különböző jellemzőket tartalmaznak, így az erdő fái az egyedet különböző nézőpontokból értékelik. A legismertebb erdő osztályozók a véletlen erdő (random forest) és a boostolt erdő (Farkas et al., 2020). A döntési fa modellek fejlesztése során gyakori eljárás a tanuló és teszt adathalmaz mellett ún. validációs adathalmazt is használni, így a **tanítóhalmaz** a modelljelölt tanítására szolgál, a **validációs halmaz**, más néven fejlesztési halmaz a jelölt modellek értékelésére és a legjobb kiválasztására, míg a **teszthalmaz** a legjobb modell egy végső, torzítatlan értékelésére szolgál (Tarr, 2024). A folytonos változókra épülő regressziós fa modellek esetén az elágazásban alkalmazott döntési algoritmus kiválasztása a modellszelekció. A modellszelekciót a validációs halmazon végzett osztályozás értékelése alapján hatjuk végre.

Lineáris gépek

A lineáris gépeket – a döntési fa modellel szemben - akkor érdemes választani, ha nagyon sok, elsősorban folytonos jellemzőnk van és feltételezzük, hogy a jellemzők lineáris függvénye elégséges a modellezéshez (Farkas et al., 2020), illetve olyan modellt szeretnénk, ahol minden jellemző, valamilyen mértékben, hozzá tud járulni az osztályozási döntés meghozatalához.

Egy lineáris gép a döntését az egyedet leíró jellemzők értékeinek és a minden jellemzőhöz rendelt súlyok lineáris kombinációja alapján számolja ki.

Ha van d számú jellemzőnk, akkor az i. folytonos jellemző értéke x_i. Bináris osztályozás esetén a lineáris gép modellje minden x_i jellemzőhöz rendel egy w_i súlyt. Predikciós időben a modell

kiszámolja a jellemzővektor és a modell súlyvektorának lineáris kombinációját (g() függvény) konstanst w_0 eltolás értéke mellett.

2. képlet: Diszkriminancia függvény

$$g(\vec{x}) = \sum_{i=1}^{d} w_i x_i + w_0$$

Ha a függvényérték>0, akkor az első osztályt, egyénként a második osztályt fogja predikálni a lineáris gép. Több osztályos esetben minden osztályhoz tartozik egy d+1 hosszú súlyvektor a lineáris gép modelljében. Minden osztályra kiszámolja a diszkriminancia függvény értékét és azt az osztályt fogja predikálni, amelyiknek a legnagyobb a diszkriminancia értéke (Farkas et al, 2020).

A leggyakrabban használt lineáris gépek a folytonos jellemzőkészletre tervezett sztochasztikus gradiens perceptron (SGD osztályozó) és a Support Vector Machines (SVM), illetve a diszkrét jellemzőkészletre használható Naive Bayes és Logisztikus Regresszió osztályozók. Ezeknél az eljárásoknál a tanulás módja különbözik, - azaz, hogyan állítják be a súlyvektorokat - egy tanító adatbázis alapján. Annyit érdemes még megjegyeznünk, hogy a Naive Bayes és Logisztikus Regresszió diszkrét jellemzőkészleten, míg az SGD és SVM módszerek folytonos jellemzőkészletre lettek tervezve, ezért ott érdemes őket használni.

A logisztikus regresszió olyan osztályozó, amelynek bemeneti adatai folytonos változók, míg a kimenet annak valószínűsége, hogy egy esemény bekövetkezik-e (osztály) (Orováné, 2024). Többváltozós kimenet esetén minden kimenet valószínűségét becsüljük.

A kimenet valószínűsége loglineárisan függ a bemeneti változóktól.

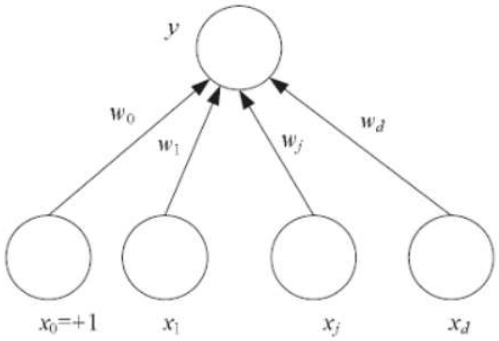
3. képlet: A logisztikus regresszió becslőegyenlete kétváltozós esetben

$$p = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

A Bayes hálózat egy irányított aciklikus valószínűségi grafikus modell, amely a véletlen változókat és köztük fennálló feltételes valószínűségeket reprezentálja. A modell megmutatja B válltozó bekövetkezésének valószínűségét A esemény fennállása mellett. Többváltozós esetben – függetlenséget feltételezve az Ai események bekövetkezte között, a Naiv Bayes modellt alkalmazzuk. Típusai: a multinomiális (előbb vázolt eset) és a Gauss-féle osztályozás (ha a változók eloszlása követi a Gauss eloszlást).

A háló alapú osztályozók a mélytanulás témakörébe tartozó, neurális háló eljárások. Alapja a perceptron (egy neuron), amelyben a belépő (input) réteg egy változó, és a kimeneti réteg egy output változóból áll. A neuron egy aktivációs függvény segítségével aktiválódik és a következő (rejtett) réteg neuronjához ér a jel. A többrétegű neurális háló modellek tervezésekor a rétegek és a node-ok (neuron) számának megadása szükséges. Az akivációs függvényeknek több típusa létezk, leggyakoribbak a szigmoid, a logisztikus regresszió, és a konvolúciós neurális hálózatokban (image processing) alkalmazott ReLU.

Egy neuron az x_i bemeneti értékek w_i-vel súlyozott lineáris kombinációját számolja ki, majd egy nem-lineáris f aktivációs függvényt alkalmaznak (13. ábra).



13. ábra: Egy neuron aktiválása

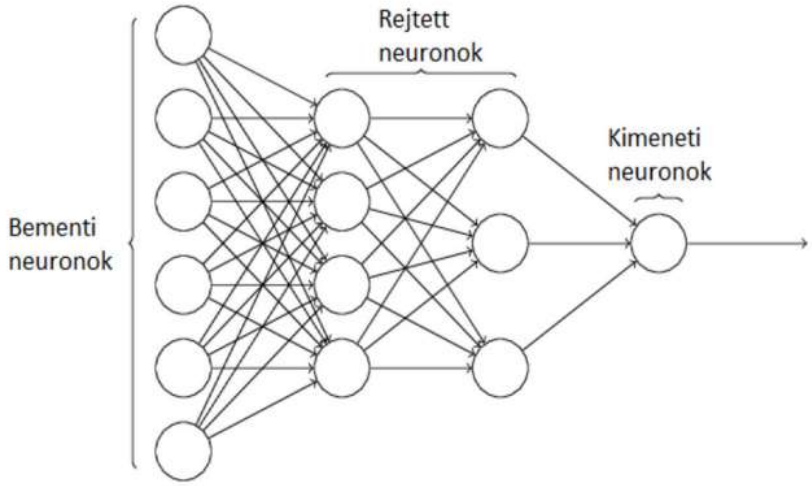
Forrás: Farkas et al, 2020

Azaz egy neuron y kimeneti értéke függ a bemeneti értékek vektorának és súlyvektorának lineáris kombinációjától, valamint az aktivációs függvénytől.

4. képlet: Egy neuron kimeneti értéke

$$y = f(\sum_{i=0}^{d} w_i a_i)$$

A neurális hálózatok a neuronok rétegeiből épülnek fel (14. ábra).



14. ábra: Többrétegű neurális háló

Az információ csak rétegről rétegre, egy irányba, előre (forward), a bemeneti rétegtől (input layer) a kimeneti réteg (output layer) felé, vagy hátrafelé (backward), a kimeneti rétegtől a bemeneti felé terjedhet. Ez utóbbi a neurális háló tanításának alapja; az aktuális súlyokat korrigáljuk a predikció hibájával.

A megerősítéses tanulás

A megerősítéses tanulás a gépitanulásnak olyan területe, amely azzal foglalkozik, hogyan kellene a szoftvernek egy tevékenységet végrehajtania annak érdekében, hogy adott környezetben a kummulált díjazás maximum legyen. Számos területen (pl. játékelméletben, multi-ágens rendszerekben, stb.) alkalmazzák, a legelterjedtebb környezet a Markov döntési eljárás. Általában nem ismerhető, vagy leírható egzakt matematikai algoritmusok és modellek hiányában alkalmazott eljárások (pl. önvezető autók).

Együttes tanulási módszerek

Az együttes tanulás a gépi tanulási eljárások hipotézisfüggvényeinek (h1, h2, h3…) kombinációja átlagolással, szavazással vagy más eljárásokkal (Tarr, 2024). Pl. a véletlen erdő modell kis döntési fák kombinációja. A legnépszerűbb együttes tanulási módszer a boosting, amikor nagyobb súllyal látjuk el azokat a mintahalmazokat, amelyek fontosabbak a tanítás során.

2.2.2. Gépi tanulási eljárások a tejtermelés modellezésében, döntéstámogatásban

Az eredeti (n=488) listában a „review” keresőszó címbeli találata alapján 11 áttekintő tanulmány volt azonosítható. Az áttekintő tanulmányok közül a cím és absztrakt átolvasása után Ozella és munkatársai (2023) és Slob és munkatársai (2021) munkáit választottam ki.
A következőkben a fenti két áttekintő cikk néhány eredményét ismertetem, amelyek a célul kitűzött vizsgálatot érdekében megalapozó információként szolgálnak.

Ozella és munkatársai 60, specifikusan az automata fejőrendszerek és a tejtermelő tehenek egészsége, termelése, viselkedése és menedzsmentje témát célzó tanulmányt tekintett át. A módszereket tekintve a tanulmányok többségében gépi tanulást (63%), míg jóval kisebb arányban statisztikai elemzést (14%), fuzzy (9%) és determinisztikus modelleket, valamint detekciós algoritmusokat (7-7%) alkalmaztak.
Mind Ozella, mind Slob kiemeli, hogy a tanulmányok nagy többsége a tőgygyulladást helyezve középpontba, a tehenek egészségi állapotának modellezésével foglalkozik, jóval kevesebb a tejtermeléssel és különböző aspektusaival, mutatóival. A gazdálkodás számára kézzelfogható, anyagi előnyt jelenthet a tejtermelési paraméterek és az elvárt, vagy átlagos mutatóktól eltérő hozamok feltárása, ez mégis alulkutatott terület (Ozella szerint mindössze 7%-a a tanulmányoknak). Slob egy másféle osztályozási megközelítéssel azt mutatta ki, hogy a tej termelési paramétereivel (66%), a tej beltartalmi (58%) és fizikai (32%) értékeivel jelentős számú, a mesterséges intelligencia algoritmusokat alkalmazó modellezés foglalkozik – függetlenül attól, hogy a tehenek egészségi állapota a vizsgált paraméterek között volt-e.

A Slob-féle feldolgozás a tejtermelő tehenészetek menedzsmentjét támogató gépi tanulás alkalmazásokat vizsgálta 38 tanulmány alapján. A hét kategóriába rendezett 71 változó közül témánk szempontjából azok a termelési és fejési paraméterek, amelyeket a robotfejőgépek szenzorai szolgáltatnak. A tejminőség kategóriában és a tejtermelés kategóriában végzett vizsgálatok megoszlása a tanulmányokban alkalmazott gépi tanulási eljárások szerint az alábbiak szerint alakult (1. táblázat).

1. táblázat: A tanulmányokban vizsgált problémakategóriák és alkalmazott gépi tanulási eljárások kereszttáblája

kategória	döntési fa	ANN	regresszió	egyebek
termelési paraméterek*	*3*	*8*	*5*	*2*
fejési paraméterek**	*3*	*6*	*4*	*1*

Forrás: saját szerkesztés Slob és mtsai (2021 6. oldal) alapján

** elektromos vezetőképesség, auto fluoreszcencia, raman spektrumok, szín, mozgóátlag vezetőképesség, áramlási citometria, spektrális adatok, ph, fd negyedek, fagyáspont, színes kép, tej bhb, foszfokolin, tejzsír, kazein, tej karbamid, zsírmentes szilárd anyagok, laktóz, glicerofoszfokolin, zsír/fehérje arány, fehérje, anyagcseretermékek*
*** tejhozam eltérése, tejmintavétel, tejhozam, fejési gyakoriság, csúcsáramlás, mozgóátlaghozam, tejmennyiség, fejési idő*

A betegségek előrejelzésében, modelljeiben a döntési fa (12) eljárások domináltak, a tejtermelés minőségi vagy mennyiségi jellemzőihez köthető eljárások a neurális hálók, regressziós eljárások vagy egyéb gépi tanulási módszerek.

Ozella (2023) feldolgozása alapján a különböző gépi tanulás eljárások az alábbi modellekben paraméterezhetők:

(1) Dinamikus lineáris modellezés célja lehet a tehenek egyedi tejtermelésének (fejésenkénti) előrejelzése, az előrejelzés és a tényleges értékek közötti eltérések, valamint a szomatikus sejtszámmal való kapcsolat feltárása.

(2) Döntési fa eljárások alkalmasak a tejhozam előrejelzésére, a fejési gyakoriság, a laktációszám, a termelési hónap figyelembe vételével, illetve ezen tényezők közötti interakciók feltárására, amely szelekciós döntési kritérium is lehet. Az ellés körüli időszak rögzített adatai is megfelelők lehetnek a tejtermelés hozamára CART döntési fa modell alapján.

(3) Véletlen erdő algoritmus segítségével a tejtermelés rövid és hosszabb távú trendje becsülhető a környezeti adatok függvényében.

(4) Neurális háló modell (Bayesi tanuló algoritmust alkalmazva) mikroklimatikus, takarmányozási és testtömeg adatokat felhasználva nagy megbízhatósággal modellezte a termelt tej mennyiségi és egyes beltartalmi paramétereit.

(5) A robotfejőgép szenzorai segítségével rögzített adatok felhasználásával többféle modellel is előrejelzés adható a tejhozamra, tej összetételére és a fejési gyakoriságra, amelynek gyakorlati hasznosítása a beavatkozásban, például a megfelelő tápanyagellátás és környezeti mikroklíma biztosításában rejlik.

2.2.3. Integrált fejőrobot rendszerek

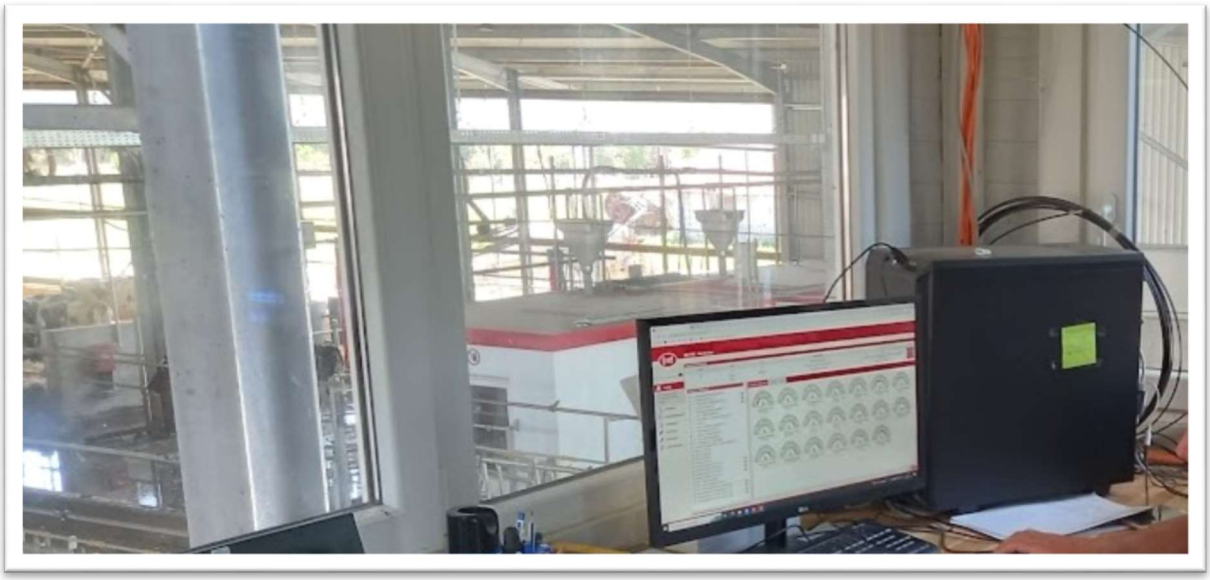
A robotfejőgépeket alkalmazó tejtermelő üzemekben az adatok sokasága (15. ábra) áll rendelkezésre a döntéshozásban. Automatikus fejőgép rendszereket (AMS) ma Magyarországon két fő vállalat telepít és támogatja üzembe helyezésüket – egyre növekszik ezeknek a rendszereknek az elterjedése.

15. ábra: Egy fejőrobot box és a nyerhető adatok köre

Forrás: https://www.delaval.com/

A szoftveres támogatás főbb jellemzői, hogy a beteg egyedekre azonnal figyelmeztet (szomatikus sejtszám emelkedés), a rendellenes viselkedési mutatók alapján jelzést küld pl. a visszaivarzó egyedekről vagy sikertelen fejési kísérletek megjelenéséről.

1. kép: A robotfejőrendszer számítógépes információs felülete egy istálló irodájában



Forrás: saját kép, készítve: 2022. július 4.

A takarmányozási költség után a második legnagyobb közvetlen ráfordítás az állategészségügyi, és a meg nem termelt tej vesztesége is jelentős gazdasági hatással jár.
Minden olyan információ, amely a gazdálkodás hatékonyságának javítását lehetővé teszi – így a robotfejőgépek és egyedi monitoring eszközökből származó adatok hasznosítása a gazdálkodás érdeke.
A nagy adatokon végezhető vizsgálatok informatikai háttere (mind a tárolókapacitás, mind a műveleti kapacitás) rendkívüli mértékben nőtt köszönhetően az elérhető cloud technológiáknak. A vállalkozásoknál azonban nem áll rendelkezésre az információ kinyerésének, az algoritmusok, műveletek elvégzésének és a modelleredmények értelmezésének készsége. A technológiát nyújtó szolgáltatók szaktanácsadói, kutatóműhelyek – a gazdasági szereplőkkel együttműködésben – kezdték feldolgozni az adatokat (Horváthné et al, 2024) és vizsgálják a termelési döntéstámogatásban betöltött szerepét, lehetőségeit.
A rögzített adatok kinyerésére is létezik eljárás, napi összesített, vagy egyedi, illetve fejésenkénti rekordokat exportálhatunk ki az informatikai rendszerből továbbfeldolgozás céljára.

2.3. A vizsgálat elméleti modelljének megalapozása

A szövegelemzés hálózattérképét felhasználva szűkíthetővé vált a feldolgozandó tanulmányok listája. A találatok szűkítése érdekében több lépésben módosítottam a keresőalgoritmuson (2. táblázat).
A következő két keresőalgoritmusból kikerült az általános megközelítést tükröző „mesterséges intelligencia” kifejezés, majd tiltottam a sántaság és a mastitis szavakat, mivel a tanulmány célja, hogy a tejtermelés szenzorral mérhető paramétereinek felhasználásával termelési csoportokat azonosító, a termelés perzisztenciáját osztályozó, esetleg előrejelző modelleket alkalmazó tanulmányokat találjak.

2. táblázat: A szűkítést elérő keresőalgoritmusok összevetése

	Keresőkifejezések (TS mezőkben)	**kizárt kifejezések**	**találatok száma**
1. keresőalgoritmus	*robot* milking, automated milking, machine milking, ML, machine learning, classification, regression, cluster*, neural network, fuzzy, artificial intelligence, AI*		*488*
2. keresőalgoritmus	*classification, regression, cluster, neural network, fuzzy, robot* milking, automated milking, machine milking*		*113*
3. keresőalgoritmus	*2. algoritmus kifejezései + production*	*mastitis, lameness*	*65*
4. keresőalgoritmus	*3. algoritmus kifejezései + cow*		*29*

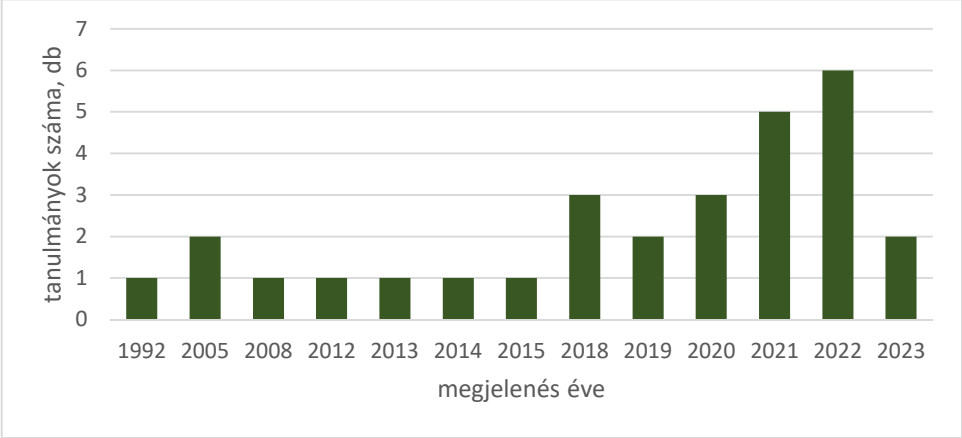
Forrás: saját szerkesztés

A szűkített találati lista tanulmányainak áttekintése

A tanulmányokat 18%-ban a Journal of Dairy Science, 14-14%-ban a Journal of Animal Science és az Animals publikálta. Jelentős még a témakörre tekintettel a Computers and electronics in Agriculture publikációk részaránya (11%). A dolgozat szűkebb témakörében feltárt tanulmányok a fentieken kívül a Biosystems Engineering, Sensors, Polish Journal Of Veterinary Sciences, Applied Animal Behaviour Science, Journal Of Food Composition And Analysis, Preventive Veterinary Medicine, Translational Animal Science, Animal, Agricultural And Food Science folyóiratokban, illetve két konferenciakötetben jelent meg.

A megjelenések dinamikáját a 16. ábra jelzi.

16. ábra: A szűkített találati lista tanulmányainak eloszlása a megjelenés éve szerint



Forrás: saját szerkesztés

A szűkített találati lista tanulmányai a 2. és 3. feladatként megfogalmazott lépések alapját képezik. A témában megjelent áttekintő tanulmányok eredményei alapján az adatelemzés megtervezése, vizsgálati célok, paraméterek, modellek kijelölése; valamint a benchmark tanulmány alapján az adatgyűjtés megtervezése, az adatbázis beszerzése képezte a következő lépéseket.

3. Módszertan

3.1. A kutatás célja és a bevont változók körének kiválasztása

A fent leírt keresési eljárás alapján nyert 29 tanulmányból (1. sz. melléklet) 22 került további kizárásra, amelynek okai nagyrészt a tanulmány célkitűzése (tématerületen kívüli: 18 db) volt vagy mert nem (csak) robotfejőrendszerből származó adatokra épülő vizsgálatra vonatkozott (2 db). Két olyan tanulmány referenciaként (3. táblázat *-gal jelöltek) a feldolgozásba bekerült ((test)hőmérséklet, illetve pedigree adatokat is modellbe építő tanulmány), amely a fenti módszertani kiválasztási kritériumnak ugyan nem felelt meg, de jelen dolgozatban felhasználható összefüggésekre mutatott rá. További egy áttekintő tanulmány (Slob et al, 2021), amelynek feldolgozása korábbi fejezetrészben már megtörtént, így nem itt kerül bemutatásra.
A dolgozat témájához szorosan kapcsolódó kutatások eredményeit és módszertanát az 5 fennmaradt és a 2 további(*) dokumentum alapján dolgoztam fel (3. táblázat).

3. táblázat: Módszertani benchmark tanulmányok

s. sz	folyóirat	szerző	a kutatás célja és eljárás	területi lefedettség
1	*Journal of Dairy Science*	*Castro, A et al. 2012*	*AMS rendszerek kapacitásának meghatározása többváltozós lineáris regresszióval*	*Észak-Olaszország (farm=29)*
2	*Animals*	*Aerts, J et al. 2022*	*döntési fa eljárással a fejési hatékonyságra ható specifikus tényezők feltárása*	*Lengyelország (farm=20, egyed=1823)*
3	*Biosystems Engineering*	*Ji, BY et al. 2020**	*hőstressz, viselkedés és robotfejési teljesítmény kapcsolata többváltozós regresszióval*	*Ausztrália (farm=1)*
4	*Biosystems Engineering*	*Bonora, F et al. 2018*	*AMS alapú többváltozós idősor adatok alapján állományszegmentáció és kapcsolatmodellezés klaszter-gráf modellel*	*Olaszország (nyár) (farm=1, egyed=65)*
5	*Computers and Electronics in Agriculture*	*Rebuli, KB et al. 2023*	*tejtermelő képesség jellemzői és stabilitása multi-algoritmikus klaszterezési eljárással*	*Olaszország (farm=1)*
6	*Polish Journal of Veterinary Sciences*	*Antanaitis, R et al. 2018*	*a kérődzési idő összefüggése a tejtermeléssel és minőségi paraméterekkel lineáris regresszióval*	*Litvánia (farm=1, egyed=728)*
7	*Animals*	*Aerts, J et al. 2021**	*genetikai és öröklődési korreláció feltárása a tejhozam, fejési gyakoriság, és fejési sebesség értékmérő tulajdonságokra véletlen regresszió modellel*	*Lengyelország (farm=21, egyed=1713)*

Forrás: saját összeállítás

A dolgozat céljára kiválasztandó változók és modellezési eljárások azonosítása érdekében a fenti tanulmányokban kidolgozott modellek eredményeiből indultam ki. Ezeket a 4. táblázatban foglaltam össze.

4. táblázat: A benchmark tanulmányok változókészlete és eredményei

szerző	függő (vagy vizsgált) változó	független változók	eredmények
Castro et al. 2012	*éves összes tejhozam robotfejőgépenként*	*tehenek száma AMS-enként* *fejések száma tehenenként naponta* *tejáramlási sebesség* *elutasítások száma AMS-enként évente*	*Lineáris regresszió: 33,7% tejhozam növekedés érhető el átlagosan a robotfejőgépeken, ha a kapacitás 16 tehénnel nő, az átlagos napi fejésszám 2.69-ról 2.48-re csökkenhet. Legerősebb prediktor az AMS-re jutó tehénlétszám és a tejáramlási sebesség. Maximális tejhozam akkor érhető el, ha az átlagos fejésszám 2.40 és 2.60 között van.*
Aerts et al. 2022	*tehenenkénti fejési hatékonyság (kg/perc)*	*AMS használati évek (1, 2, 3)* *robotonkénti tehenek száma (45–50; 51–55; 56–60; 61–75)* *laktáció szám (1 vagy több: 2 vagy 3)* *ellési időszak (ősz, tavasz, nyár vagy tél)* *életkor első elléskor*	*Döntési fa: a legmagasabb fejési hatékonyság tehenenként (2,01 kg/perc) ott érhető el, ahol napi termelés 45 kg-nál több, naponta négynél kevesebbszer fejték, rövid volt a felhelyezési és fejési idő (<7,65 s), továbbá 56 tehénnél kisebb létszámú robotokban.*
Ji et al. 2020	*kérődzési idő (hőstressz proxy változója)*	*tejhőmérséklet* *napi tejhozam* *fejési idő* *fejési gyakoriság* *fejés időpontja* *fejési sebesség* *egy fejésre jutó tejhozam*	*többváltozós szegmentált regresszió: a nap első fejési eseményének késleltetése és a fejési intervallumok 4 óránál rövidebbre csökkentése javítja a kérődzési hatékonysági indexet és a robotfejés teljesítményét*
Bonora et al. 2018	*klasztertagság összefüggése hálózattérkép alapján*	*k-közép klaszterekbe tartozás* *Napi fejések száma* *Paritás* *Átlagos napi aktivitás* *A fejési események közötti időintervallumok szórása a vizsgálati időszakban* *Tehén testtömeg szerint*	*kiemelt klaszterjellemzők:* *1. klaszter: kevés eredményes AMS látogatás (napi átlag 2), gyenge rendszerességgel (a látogatások közötti időintervallum szórása közel 5 óra). Ez a klaszter az állatállomány közel felét foglalta magában.* *3. klaszter: a napi AMS látogatások hármat meghaladó átlagos száma, rendszeresség, az időintervallumok szórásával adva körülbelül fele volt az 1-es klaszterének.*
Rebuli et al. 2023	*klaszterbe tartozás stabilitása*	*napi tejtermelés (laktációként)*	*különböző termelési hatékonyságú csoportokba tartozó egyedek laktációkon keresztüli csoporttartása*
Anatanaitis et al. 2018	*kérődzési idő*	*modellalkotók:* *napi tejhozam (6 osztály)* *testtömeg* *tejösszetétel (zsír, fehérje) (3 osztály)* *szaporodásbiológiai állapot (5 osztály)* *egyéb:* *laktóz* *szomatikus sejtszám*	*a legalacsonyabb termelés a vemhes teheneknél a legmagasabb a frissen ellettekél* *leghosszabb kérődzési idő a termékenyített teheneknél (szignifikánsan magasabb, mint a nem vemhes teheneknél)* *kérődzési idő pozitív korrelációt mutatott a termelékenységgel és szorosabbá vált a laktációszámmal* *lineáris regresszió: $y = 38{,}02x + 232$, $R^2 = 0{,}721$ ($p<0{,}001$).* *szaporodásbiológiai állapot és tejzsír-fehérje arány között statisztikailag szignifikáns összefüggés* *a leghosszabb kérődzési időt a termékenyített teheneknél (1-35 nappal a termékenyítés után), a legrövidebbet a nem vemhes teheneknél figyelték meg*

Aerts et al. 2021	*genetikai tényező (h^2)*	*pedigree adatok* *tejhozam* *fejési gyakoriság* *fejési sebesség* *időszak: fejési napok száma (a tesztidőszak 5. napjától a 305. napig)*	*kétváltozós véletlen regressziós modellek:* *A tejhozam és fejési gyakoriság változókra a véletlen regressziós modell 2-es rendű heterogén reziduális varianciával, a fejési sebességre a véletlen regressziós 1-es rendű heterogén reziduális variancia modell volt a legjobban teljesítő.* *a napi fejési sebesség becsült öröklődése mérsékelt, míg a napi fejési gyakoriság és a tejhozam alacsony volt* *A magas, pozitív genetikai korreláció a napi fejési gyakoriság és a tejhozam között rámutat, hogy a fejőrobot gyakoribb látogatása a tejhozam genetikai alapja lehet.*

Forrás: saját összeállítás

Összegezve megállapítható, hogy a tanulmányok érdeklődési köre részben a) a robotfejőgépek kihasználtságának (összkapacitásának vagy az egy tehénre jutó fejési hatékonyságnak) a javítása, részben b) az állomány termelésében rejlő tartalékok feltárása, akár szelekciós előrehaladás, akár a termelés csoportos perzisztenciájának megismerése révén. Két tanulmány a kérődzésszám modellezése (mint az egyedek jóllétének indikátora, illetve mint a termelési hatékonyság prediktora) felé fordult.

A modellekben felhasznált adatok köre – az AMS integrált rendszerében rögzítetten elérhetőkkel összhangban – a következők.

5. táblázat: Felhasznált adatok köre

Fejőgép adatai:	***Egyed adatai:***	***Az adatrögzítés körülményei:***
Tehenek száma robotonként	*Azonosító*	*Fejés időpontja*
Elléskori életkor	*Laktáció száma*	*A mérés dátuma*
Fejt napok száma	*Laktációban töltött napok száma*	
Fejési gyakoriság	*Vemhesség*	
Csatlakozási idő	*Vemhességi napok száma*	
Boxban töltött idő		
Tejáramlás sebessége		
Tejhozam		
Hátsó tőgynegyed tejhozam aránya		
Fejési hatékonyság		
Fejési idő		
Tejzsír%		
Tejfehérje%		
SSC		
Elutasítások száma		

Forrás: saját összeállítás

A szakirodalomban alkalmazott eljárások, a gépi tanulási eljárások (klaszterezés, döntési fa, lineáris regresszió, véletlen regresszió) mellett, statisztikai (összefüggésvizsgálat kategóriaváltozókra, minták közötti különbségek feltárása) és egyéb módszereket (hálózatelemzés) foglalnak magukban.

3.2. A vizsgálati modell elméleti kerete

A dolgozatban megvalósítandó kutatás célkitűzése olyan egyedcsoportok azonosítása, amelyekben lévő egyedek tejhozama hasonló, de más csoportoktól eltérő jellemzőkkel rendelkezik.

A tejhozam alakulását az egyedek laktációs görbéje (tejtermelési időszakban fejt napi tejmennyiség) írja le. A laktációs görbe élettani szakaszai alapján a termelés ellést követő felfutása, a termelési csúcs időszaka, a laktáció nagy részét kitevő plató és az apasztásig tartó, a vemhesség előrehaladásával fokozatosan csökkenő termelés szakaszai különíthetők el. (Anatanaitis és munkatársai (2018) tanulmányukban igazolták a szaporodásbiológiai állapot hatását, amely a laktáció egyes szakaszaival is összefüggésben áll.)

6. táblázat: A vizsgált laktációs szakaszok

Laktációs szakasz	Szakaszra jellemző termelési mintázat	Termelésben töltött napok száma (DIM)	
		üszők	tehenek
1. szakasz	*ellést követő termelésfelfutás*	*5-35*	*5-49*
2. szakasz	*termelési csúcs*	*36-49*	*50-63*
3. szakasz	*termelési plató*	*50-175*	*64-175*
4. szakasz	*termelés csökkenése*	*176-260*	

Forrás: saját összeállítás

A tejtermelés vizsgálatát és az egyedek csoportokba sorolását ezen szakaszokban végzem el, majd megfigyelem, hogy az egyedek változtatják-e a csoportok között a helyzetüket a laktáció szakaszaiban. Ehhez Bonora et al. 2018 tanulmányában alkalmazott klaszter-gráf módszert alkalmazható a későbbiekben.

Mivel az elsőborjas üszők és a többször ellett tehenek laktációja különbözik, az állományt két csoportra érdemes bontani (Aerts et al. 2022).

A termelési csoportokat felügyelet nélküli tanulási eljárások közé tartozó klaszterképzéssel hozom létre, a megfelelő klaszterszámot a Shiluette együtthatóval határozom meg. Az eljárások közül a k-közép klaszterezés módszert alkalmazom (Bonora et al. 2018 és Rebuli et al. 2023 eredményesen alkalmazta), amelynek során a következő változókat vonom be a modellbe: napi tejtermelés összege és varianciája, a napi fejések számának összege és varianciája és a napi termelés maximális értéke a laktáció szakaszaiban. A változókat a robotfejő rendszer fejésenként regisztrált adataiból származtatom.

3.3. Adatok

A dolgozatban felhasznált adatok szaktanácsadói hozzáféréssel, a gazdaság vezetőjével egyeztetetten elérhető adatbázisból származnak. A gazdaságban Holstein Fríz állomány van, 8 db robotfejőgép több mint egy éve működik.

Az adatokat a telepi integrált robotfejőrendszerben 2024. január 1-től 2024. március 31-ig regisztrált fejésekre 30 kezdeti változókészlettel és 437 tehénre vonatkozóan, összesen 133692 fejési alkalom adataival csv formátumban letöltött fájlként kaptam.

Az elméleti modell alapján az alábbi változókat (7. táblázat) jelöltem meg a vizsgálat céljára.

7. táblázat: Klaszterképzésbe vont változók

változó megnevezése	változó jelölése	változó származtatott értékének alapja	változó alapjának mértékegysége
napi tejtermelés összege	*sMY*	*egyedenként és fejésenként regisztrált tejhozam napi összege*	*kg/nap/egyed*
napi tejtermelés varianciája	*vMY*		
a napi termelés maximális értéke	*mxMY*		
a napi termelés varianciája	*vMY*	*egyedenként regisztrált sikeres fejőrobot látogatások napi száma*	*db/nap/egyed*
napi fejések számának átlaga	*sMF*		

Forrás: saját összeállítás

Az adattáblában összesítettem egyedenként és naponként a fejések számát (count_Milking), a fejt mennyiséget (MilkYield és Max_MilkYield) és megtartottam a fejés dátuma (short date), hónap (MO), nap (DA), boxban töltött idő (robotTime, mp), a fejési sebesség (Mspeed, l/p), laktációszám (Parity), fejt napok száma (MIDs), és a laktációs szakasz (Lact_segm) adatokat.

A kapott dataframe dimenziója: (33846, 12).

Az adattáblát felhasználva, az elemzés során, k-közép eljárással klasztereket hoztam létre, amely alapján klaszterekbe tartozó egyedeknek – laktációs szakaszonként és paritás bontásban – a robotfejő rendszerben rögzített adatai felhasználásával definiálhatók a csoportok közötti különbségek az egyes további paraméterek mentén (pl. fejési sebesség, robotban töltött idő, tejösszetevők, stb.).

4. Eredmények és értékelésük

A fejezet az adattábla előkészítésének műveleteitől a feldolgozásának bemutatásán keresztül, a kapott eredmények közül kiválasztottak értelmezése és a kutatás kiterjeszthetőségére történő kitekintést öleli fel.

4.1. Az adattábla előkészítésének műveletei, adattisztítás

Az adatelőkészítés műveleteit a csv fájl MS Excelbe importálásával, a szükséges beállítások és felesleges karakterek eltávolításával kezdtem. Az alapadatkészlet: 133692 x 30 mátrixban tartalmazta a letöltött adatokat.

A dolgozat vizsgálati témájának szempontjából fontos volt, hogy az adatok csak a sikeres fejéseket tartalmazza. A robotlátogatások eredménye szerinti megoszlását a 8. táblázat mutatja be. (A sikertelen fejések vizsgálata további kutatás alapját képezheti.)

8. táblázat: A robotlátogatások eredményeinek összesítése

látogatás eredménye	darabszám	megoszlás
fejési időköz nem felelt meg	*36559*	*27.18%*
sikertelen fejések	*855*	*0.64%*
egyéb	*342*	*0.25%*
csatlakozási kísérletek (tőgynegyedhez)	*271*	*0.20%*
csatlakozási idő	*264*	*0.20%*
holt fejési idő (tőgynegyeden)	*208*	*0.15%*
egyéb	*75*	*0.14%*
sikeres	*95880*	*71.28%*

Forrás: saját összeállítás

A további elemzések a sikeres és befejezett fejési kísérletek adatbázisára vonatkoznak.

Az adatok a táblázatban fejésenkénti rekordként álltak rendelkezésre, azaz egy egyedtől több rekord egy-egy termelési napra vonatkozóan. Így összesíthetővé váltak az egyedekre vonatkozó napi jellemzők, pl. a (sikeres) fejések gyakorisága, átlagos napi mennyisége és mennyiségének varianciája, vagy a fejt mennyiségek napi maximum értéke az adott egyed esetében. Az összegzést az adattábla előkészítése során egy összetett változó (egyedi sorazonosító) képzésével (állatnap) a megfelelő oszlopok táblázatkezelőben történő összefűzését (concatenate művelet) követően kimutatáskészítővel (sorváltozó: állatazonosító) értem el. Adatbázisműveletként értelmezve szelekciót és projekciót végeztem az adatokon.

A feldolgozandó adattábla a 9. táblázat szerinti oszlopokat tartalmazta. Az adattábla 33845 rekordot tartalmaz (megfigyelt állatnapok száma n=33845).

9. táblázat: Az adattábla változókészlete

változó megnevezése	jelentése
cowDay	*állatnap*
cowID	*egyedi ID*
Mspeed	*fejési sebesség átlagos napi értéke*
robotTime	*boxban töltött idő átlagos napi értéke*
Parity	*az egyed laktációszáma*
count_Milking	*a napi fejések száma*
MilkYield	*átlagos fejési tejhozam*
Var_MilkYield	*a fejési tejhozam napi varianciája*
MIDs	*az egyed laktációban töltött napok száma a regisztrált adat idején*
Lact_segm	*az egyed laktációs görbéje szerinti szakasz száma*
short_date	*az adott rekord regisztrációjának napi dátuma*
MO	*az adott rekord regisztrációjának hónapja*
DA	*az adott rekord regisztrációjának naptári napja*
Max_MilkYield	*a napi fejési hozamok maximumértéke*

Forrás: saját összeállítás

A következő megoldandó probléma a hiányzó adatok kezelése volt (ahol a napi fejésszám alacsony értéke miatt nem volt értelmezhető a variancia). Többféle módszer (pydata.org) elérhető a hiányzó adatok pótlására, számításos alapú megoldásokat azért nem alkalmaztam, mert ebben a fázisban még nem megmondható az a megfigyelés-csoport, amelyből akár átlagolással, lineáris vagy más extrapolációval, egyéb művelettel számítható értékkel helyettesíthető lett volna az adat. Másrészről az ilyen egyedek csoportja az üzemi menedzsment számára értelmezhető, ezért a rekordok eltávolítása sem kerülhetett szóba. Konstanssal történő helyettesítéssel a hiányzó értékeket 0 értékre állítottam, ami ebben az esetben nem a 0 varianciájú megfigyeléseket jelenti.

4.2. Az adatfeltárás műveletei

A sikeres fejések adatainak további előkésztése a valószínűsíthető adatmintázatok előzetes feltárása, és a nem megfelelő rekordok kiszűrése (kiugró értékek) érdekében valósult meg.

Az adattábla beolvasását kitallózható módon oldottam meg. Importáltam a szükséges I/O műveleteket lehetővé tevő modulokat.

2. kép: kódrészlet: Az adattábla beolvasása Jupyter notebook platformon

Milking_data_analy.py

```
# -*- coding: utf-8 -*-
"""K_means_új.ipynb

Automatically generated by Colab.

Original file is located at
    https://colab.research.google.com/drive/1HLA99GEcZnw2Ib7IW_9207bVlXhRV_5-

"""

import pandas as pd

from google.colab import files
from google.colab import drive
drive.mount('/content/drive')
uploaded=files.upload()

import io

df=pd.read_excel(io.BytesIO(uploaded['ujtabla1NaN.xlsx']), header=0, decimal='.')
```

Forrás: saját összeállítás

Pandas modul fájlbeolvasás műveletével dataframe-be töltöttem az adattáblát.

A nem megfelelő adatok sorainak (fejési átlag nem érte el az 5 litert) eltávolítása után alapstatisztikákat számítottam. Mind az alapstatisztika táblázatot, mind a szűrt, tisztított dataframe-et xlsx formátumban mentettem.

3. kép: kódrészlet: A beolvasott dataframe szűrése és lementése

```
"""alapstatisztikák számítása"""
sumstat=df_szurt[['Mspeed','robotTime','Parity','count_Milking','MilkYield','Var_MilkYield','MIDs',
                  'Lact_segm','short_date','MO','DA','Max_MilkYield']].describe()
sumstat.to_excel("sumstat_df_szurt.xlsx", sheet_name="Sheet1")
sumstat
print(df_szurt.columns)
df_szurt.shape
df_szurt.to_excel("df_szurt.xlsx", sheet_name='Sheet')
```

Forrás: saját összeállítás

A tisztított adattábla statisztikai összesítésének exportjából látható a megfigyelt tulajdonságok eloszlását jellemző főbb mutatók (10. táblázat).

10. táblázat: A sikeres fejések összesítő statisztikája (szélsőséges adatok nélkül)

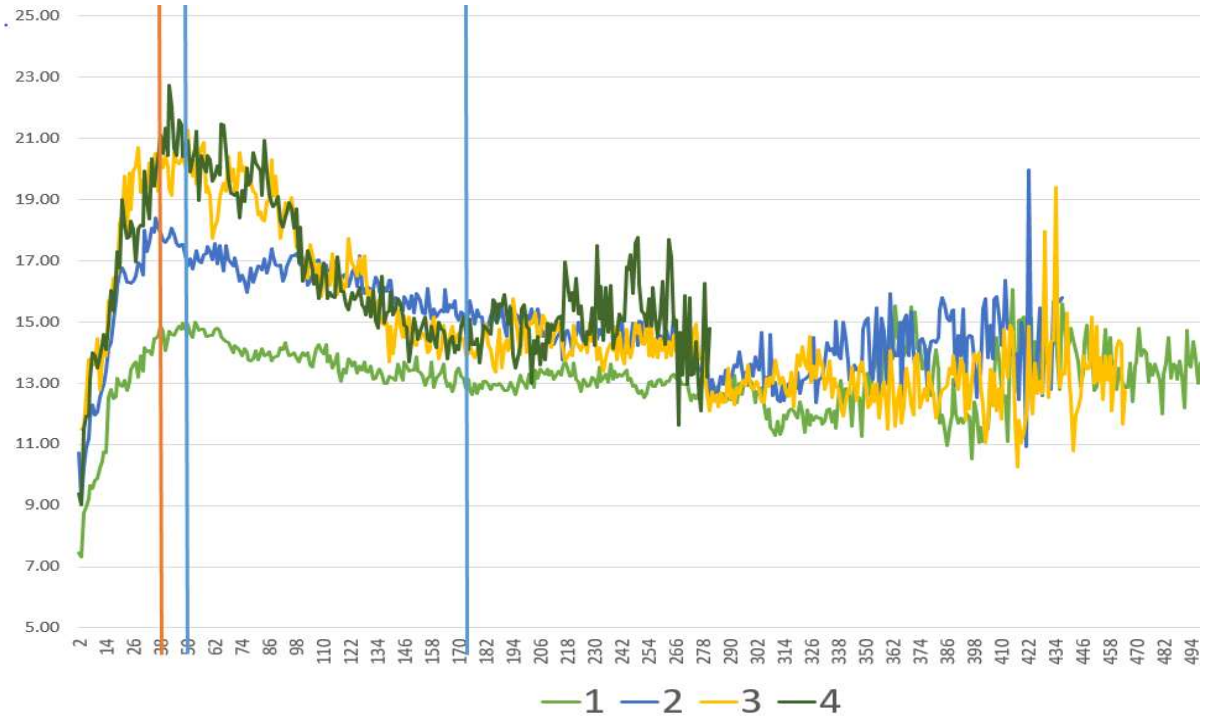
	db	átlag	min	25%	50%	75%	max	std
Mspeed		3.28	0.40	2.58	3.20	3.87	9.10	1.00
robotTime		0.31	0.10	0.25	0.29	0.35	2.50	0.09
Parity		2.01	1.00	1.00	2.00	3.00	7.00	1.12
count_Milking		2.83	1	2	3	3	6	0.75
MilkYield	*33796*	15.12	5.03	12.50	14.50	17.20	40.55	3.78
Var_MilkYield		8.18	0.00	1.13	3.74	9.68	564.48	15.61
MIDs		154.06	2.00	76.00	147.00	214.00	566.00	97.46
Lact_segm		3.09	1	3	3	4	4	0.98
Max_MilkYield		17.24	5.10	14.10	16.70	19.90	53.90	4.43

Forrás: saját összeállítás

A vizsgálat célkitűzésének megfelelően megfogalmazott elméleti módszertani keret alapján a klaszterképző változók (32. oldal táblázat) alapját adó mutatók, azaz a fejések száma és a napi egyedenkénti átlagos tejhozam alakulása az egyedek laktációszámának és a laktációs görbe szakaszainak bontásában érdekes.

Az adatok mintázatának áttekintése érdekében a fejési hozamok alakulását vonaldiagramon ábrázoltam (17. ábra) az első négy laktációs csoportban.

17. ábra: A fejési hozam alakulása laktációs csoportok (1, 2, 3, 4) szerinti bontásban



Forrás: saját összeállítás

Az ábrán is szemléletesen kirajzolódik a laktációs csoportok hozamának eltérő alakulása; a függőleges vonalakkal a jellemző szakaszok határait jelöltem.
Az élettani laktációs szakaszokra (lásd: 6. táblázat) számítottam a napi fejések számát és a napi összes tejhozam alakulását a különböző laktációban termelő egyedekre és összevetettem a teljes állomány ugyanezen jellemzőivel.

11. táblázat: A napi fejések száma és a napi tejhozam alakulása laktációszám (1, 2, több) szerint a laktációs görbe szakaszaiban

	1 szakasz		2 szakasz		3 szakasz		4 szakasz		összesen	
	napi átl. fejés szám	*napi össz tejhozam*	*napi átl. fejés szám*	*napi össz tejhozam*	*napi átl. fejés szám*	*napi össz tejhozam*	*napi átl. fejés szám*	*napi össz tejhozam*	*napi átl. fejés szám*	*napi össz tejhozam*
minden csoport	*5.8*	*88.2*	*5.7*	*96.6*	*5.8*	*89.4*	*5.4*	*74.8*	*5.7*	*83.7*
1. laktáció	5.2	62.8	5.2	76.8	5.6	77.4	5.3	68.3	5.4	72.0
2. laktáció	6.3	97.5	6.6	113.5	6.1	99.2	5.5	80.4	5.9	92.3
3+ laktáció	5.8	104.2	5.5	109.5	5.9	98.2	5.7	80.1	5.8	92.6

Forrás: saját összeállítás

Az adatok várható mintázatának áttekintése megnyugtató információt nyújtott a szakirodalom alapján megtervezett módszertani kerethez illeszkedés vonatkozásában.
Az adatfeldolgozás során a fő célom a klaszterképző változók alapján meghatározni az egyedek jellemző csoportjait az egyes termelési szakaszokban.

4.3. Az adatfeldolgozás

A rendelkezésemre álló CSV formátumú adatbázis feldolgozását Jupiter notebookban, a google research collaboration térben végeztem el, egyes lépésekben, a fent bemutatott módon (előkészítés, előfeldolgozás) az adattábla MS Excelben való módosításával.
A beolvasás és dataframe-mé alakítás, adattisztítás műveleteit követően a részfeladatoknak megfelelő kódrészeket hoztam létre.
Minden egyes (vizsgálatba vont) laktációban és azon belül a laktációs görbe 4 szakaszában külön-külön volt szükség elvégezni a klaszterezési eljárást.
A scikit-learn python modul K-közép eljárását alkalmaztam 16 részadatbázison, négy egyenkénti változóval. A részadatbázisok előállításán kívül, megfelelő metrikát választva, a legjobban teljesítő klaszterszám meghatározása is célom volt. Az eredményeket tároló eljárásokat is terveztem.
A kód felépítését (4. kép) így három ciklus (laktációszám, laktációs szegmens és klaszterszám) határozta meg.

4. kép: kódrészlet: A részadattáblákat előállító egymásba ágyazott ciklusok

```
from sklearn.cluster import KMeans
from sklearn import metrics

for i in range(1,5): #4 db laktációszámon fusson végig
  global laktsz
  laktsz=i
  for j in range(1,5): #4 db laktációs szakaszon fusson végig
    global laktsegm
    laktsegm=j
    df_name=f"laktsz{i}_laktsegm{j}.xlsx"
    df_part.to_excel(df_name)
    sumstat=df_part.describe()
    sumstat.to_excel(f"sumstat_{df_name}")
```

Forrás: saját összeállítás

A kódrészlet a klaszthereljárás futtatásához szükséges modulok meghívását, a részadattábla kettős ciklusmagban való elkészítését és a (i,j)-dik részadattábla összesítő statisztikájának dataframe-be majd excel fájlba mentését mutatja be.

A ciklusokon belül az eredmények tárolása érdekében string literálokat, dataframe-ket és az ún. Shilouette mutató változóját hoztam létre.

A string literálokat a ciklusváltozók ciklusmagban felvett értékeinek kombinálásával az eredményfájlok és dataframek, valamint a klasztereredmény mező elnevezésére hoztam létre. Ezek általános alakja:

(1) df_name=f"laktsz{i}_laktsegm{j}.xlsx"

dataframe elnevezése adott laktációszám x laktációs szegmens mátrixban, az így képzett részadattábla a klaszthereljárás ciklusában került feldolgozásra

(2) cluster_label= f"{var}_CL{klaszter}_laktsz{laktsz}_lsegm{laktsegm}"

klaszter címke formázott, változókat tartalmazó string kifejezése

A klaszthereljárás ciklusban történő futtatása a 2-5 számú klaszterezés Shilouette mutatójának értékelése miatt szükséges, amely alapján az eljárást azokra a k klaszterszámokra futtatam, ahol a legmagasabb volt a mutató értéke.

A következő kódrészlet (5. kép) a klaszthereljárás ciklusmagban futtatását mutatja be. Ezt megelőzően az éppen vizsgált klaszterképző változó és a ciklusmagban a változót tartalmazó X

df megadása történik meg. A gyakorlatban a Shilouette mutató előállítása egy külön kódrészletben valósult meg, ezt követően az egész eljárást már csak a definiált klaszterszámmal futtattam újból és tároltam az eredményeket. Így – bár ez eljárást kétszer kellett futtatni, a jóval kevesebb számú dataframe-kiírási szükséglet miatt a korábbi 20-25 perces futásidő 3-7 percre csökkent. Nem említve azt a hozadékot, hogy a későbbi elemzés céljára előállt táblázatokat nem kell a változónként 80 iteráció eredménytáblái közül válogatni, ami felhasználó barátabbá teszi a kódot. Fejlesztési lehetőség lenne egy olyan megoldás megvalósítása, amikor a számított Shilouette mutatók és alaptábla azonosító alapján pl. szótárban tárolt vektorból a legmagasabb SM értékű kulcshoz tartozó változókkal fut automatikusan a klaszterezési eljárás.

5. kép: kódrészlet: A k-közép klaszterezési eljárás futtatása és az eredmények elmentése, kiírása

```
    var='MilkYield' #var_MilkYield, Max_MilkYield, count_MilkYield
    for k in range(2,6):
      global klaszter
      klaszter=k

      cluster_label= f"{var}_CL{klaszter}_laktsz{laktsz}_lsegm{laktsegm}" #Var_CLk_lszi_lszegj
      #print(f"df_part:{df_name}_\n{df_part}\n") # ellenőrzés

  #klaszterek
      km = KMeans(n_clusters=klaszter, random_state=1)
      X=df_part.drop(columns=['cowDay','cowID','Mspeed','robotTime','Parity','count_Milking','Var_MilkYield','MIDs',
                              'Lact_segm','short_date','MO','DA','Max_MilkYield'],axis=0) #klaszterképző változó
      #print(X)
      km.fit(X)
      km.labels_
      df_part[cluster_label] = km.labels_
      #print(f"{var}_CL{klaszter}_laktsz{laktsz}_lsegm{laktsegm}: kész") #ellenőrzés
      count=df_part.groupby(cluster_label).count().to_excel(f"db{var}_{df_name}")
      centers=df_part.groupby(cluster_label).mean().to_excel(f"centers_{var}_{df_name}")

      SM=metrics.silhouette_score(X, km.labels_)
      print(f"{var}_CL{klaszter}_laktsz{laktsz}_lsegm{laktsegm}:{SM}")
    df_part.to_excel(f"klaszteres_{var}_{df_name}")
```

Forrás: saját összeállítás

A k-közép klaszterezés eredményeit az adott változók szerinti klaszter címke elnevezésű oszlop (klasztertagság), a klaszterekbe tartozó egyedek száma (count() függvény) és a későbbi ábrázolásra felhasználható klaszterközéppontok táblázata tárolja (dinamikus módon, az adott ciklusváltozó-hármas állása mellett egyesével automatikusan mentve).

Ahogy erről szó volt, az eredményeket létrehozó kód külön cellában futott azt követően, hogy megállapítottam a for ciklusban előállított Shilouette mutatók alapján a megfelelő klaszterszámot. Ezt elágazással oldottam meg, adott laktáciszámban és adott laktációs szegmensben leghatékonyabb k klaszterszámmal futott az eljárás.

Erre példa a 6. képen bemutatott kódrészlet.

6. kép: kódrészlet: A SM mutató alapján megállapított leghatékonyabb k klaszterszám megadása

```
142         var='MilkYield' #var_MilkYield, Max_MilkYield, count_MilkYield
143       # for k in range(2,6):
144         global klaszter
145         if (laktsz==1 & laktsegm==1):
146           klaszter=3
147         elif (laktsz==2 & laktsegm==2):
148           klaszter=3
149         else:
150           klaszter=4
```

Forrás: saját összeállítás

Az eredményfájlokat (centerek, számosságok, klasztertagság) a google.colab változóiban tároltam, majd a futások végeztével letöltöttem.

7. kép: kódrészlet: Az eredményfájlok letöltése

```
import pandas as pd
import os
import re
import io
import shutil
from google.colab import drive
drive.mount('/content/drive')

folder_path = '/content/'  # Change this to your actual folder path
print(folder_path)
folder_dl_path='/content/drive/MyDrive/'

for filename in os.listdir(folder_path):
    file_path = os.path.join(folder_path, filename)
    folder_drive_file=os.path.join(folder_dl_path, filename)
    print(folder_drive_file)
    shutil.copy(file_path, folder_drive_file)
```

Forrás: saját összeállítás

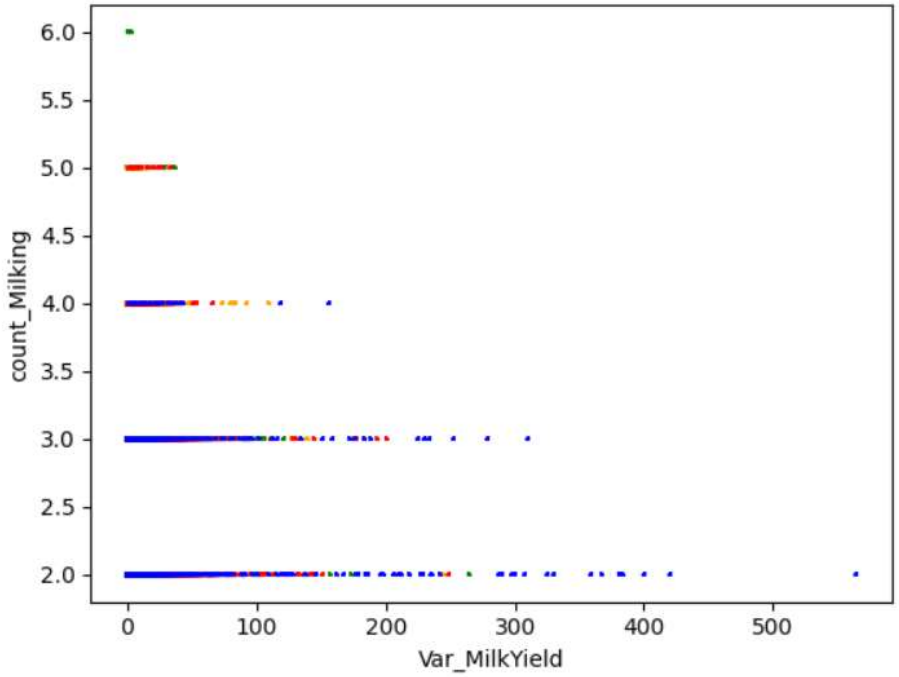
A kapott klaszterek ábrázolását is megoldottam, bár a nagyon nagyszámú adatpont miatt a klaszterezettség kevésbé megfigyelhető. Ide vonatkozóan példaként az alábbi programrészlet (8. kép) és a készített 18. ábra állhat.

8. kép: kódrészlet: Klaszterek ábrázolása

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib inline # Google Colab miatt csak

# A klaszterek középpontjait tartalmazó DataFrame- hoz változó rendelése
df_part[cluster_label] = km.labels_

centers=df_part.groupby(cluster_label).mean()
print(centers)

import numpy as np
colors = np.array(['red', 'green','blue','orange']) # ha több k több: 'blue', 'black','yellow',

# pontdiagram
plt.scatter(df.Var_MilkYield, df.count_Milking, c=colors[list(df.cluster)], s=5, marker="+")

plt.xlabel('Var_MilkYield')
plt.ylabel('count_Milking')
```

Forrás: saját összeállítás

18. ábra: Klasztercsoportok vizuális megjelenítése



Forrás: saját összeállítás

Az ábra egy négyklaszteres (k= 4) csoportosítás eredményét illusztrálja. Megállapítható, hogy:
- ebbe a csoportba tartozó egyedekre a napi 2-3 fejési alkalom jellemző;
- a napi kétszeri fejések esetén a legnagyobb a fejt tejmennyiség szóródása, illetve a fejések napi számának növekedésével a szóródás csökken;
- a klaszterképző változó alapján kialakított csoportokba eső megfigyelések legnagyobb számban a kékkel jelzett klaszterbe tartoznak, és ezek inkább az alacsonyabb napi fejésszámhoz tartoznak, míg pl. a piros, illetve narancs színű klaszter elemei a többszöri napi fejésekkel jellemezhetők.

A dolgozat terjedelmi korlátaira tekintettel nem térek ki a csoportok egyenkénti bemutatására itt.

4.4. Az egyedek jellemző klaszterei és a termelési csoportba tartozás dinamikájának vizsgálata

A vizsgált tejtermelő tehén állomány bizonyos termelési csoportjai a kapott eredmények alapján elkülöníthetőek, akár az állománymenedzsment gyakorlat fejlesztésének szükségességére is alapot adhat. Az egyedek csoportok közötti mozgása a laktációs szakaszok szerinti termelési időszakban, vagy akár – hosszabb távú megfigyelés esetén – a laktáció perzisztencijának változására utaló csoportváltás a klaszterek között izgalmas megfigyelés lehet az egyedek relatív pozíciójának változásáról, amelyből akár a szociális csoportdinamika, versengés hatására is következtethet a menedzsment. A további vizsgálatok alapját ezért a csoportba tartozás és annak változása képezheti.

A következőkben, laktációszám szerinti tagolásban, először bemutatom a csoportdinamika alakulását az egyedek klaszterhez való tartozása alapján.

Később a jellemző egyedcsoportokba tartozó (klaszterek) egyedek együttes tulajdonságai alapján a teljesség igénye nélkül néhány példán keresztül érzékeltetem, hogy a választott felügyelet nélküli tanítási modell milyen potenciállal bírhat még az üzem döntéshozói számára.

4.4.1. Termelési csoportok

A következő részben a k-közép klaszterezési eljárással kapott eredményeket, az egyedek csoportba tartozását dolgozom fel laktációszám szerint és a laktációs görbe szakaszai között. Az egyes csoportok számosságára tekintettel az eredmények részletesen nem bemutathatók, összefoglaló táblázatokkal igyekszem betekintést nyújtani.

A leghatékonyabb számú klaszter meghatározásának alapjaként az ún. Shilouette mutatót használtam (12. táblázat). Minden laktációszám x laktációs szakasz csoportban meghatároztam

az értékét. Összefoglalóan a legjobban teljesítő klaszterszámhoz tartozó értékeket mutatom be (13-16. táblázat).

12. táblázat: Shilouette mutató értékei

résztábla		*klaszterképző változók*							
		MilkYield		*count_Milking*		*Max_MilkYield*		*Var_MilkYield*	
laktáció száma	*laktációs szakasz száma*	*SM*	*k*	*SM*	*k*	*SM*	*k*	*SM*	*k*
1	1	0.68	3	0.65	3	0.73	5	0.81	2
1	2	0.74	4	0.71	4	0.79	4	0.62	2
1	3	0.74	4	0.70	4	0.78	4	0.70	2
1	4	0.74	4	0.70	4	0.77	4	0.78	2
2	1	0.71	4	0.67	4	0.77	5	0.93	2
2	2	0.72	5	0.67	3	0.78	5	0.94	2
2	3	0.71	4	0.68	4	0.79	4	0.66	2
2	4	0.72	4	0.67	4	0.79	4	0.63	2
3	1	0.67	4	0.65	4	0.78	5	0.91	2
3	2	0.74	4	0.71	4	0.79	4	0.93	2
3	3	0.71	4	0.67	4	0.81	5	0.92	2
3	4	0.73	4	0.67	4	0.79	4	0.67	2
4	1	0.66	4	0.64	4	0.78	5	0.93	2
4	2	0.72	4	0.70	4	0.78	4	0.79	2
4	3	0.69	4	0.66	4	0.80	5	0.65	2
4	4	0.70	4	0.66	4	0.79	4	0.62	2

Forrás: saját számítások

A táblázatban megadott *k* értékek jelölik, hogy az adott résztáblában milyen számú klasztert alakítottam ki a változók alapján.

Átlagos jellemzők összehasonlítása a klaszterek mentén

A négy változón (fejések száma, fejésenkénti hozam és ennek maximális értéke, illetve varianciája) kialakult klaszterek középpontjai és a klaszterbe sorolt megfigyelések további tulajdonságainak vizsgálata a laktációs görbe szakaszainak sajátosságaira hívja fel a figyelmet üzemi körülmények között.

A négy változó alapján kialakított klaszterek átlagos jellemzőit az alábbi táblázatok foglalják össze.

13. táblázat: Napi fejések száma alapján kialakított klaszterek átlagos jellemzői az első laktációban termelő egyedek esetén

CL3_ lsegm1	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	Max_MilkYield	MIDs
0	2.72	0.32	2.71	12.67	4.62	14.24	20.56
1	2.81	0.36	2.28	16.54	11.31	18.72	25.03
2	2.62	0.27	2.75	8.98	1.70	9.89	12.10

CL4_ lsegm2	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	Max_MilkYield	MIDs
0	2.36	0.34	3.07	11.28	1.25	12.14	41.87
1	2.98	0.38	2.28	18.79	12.07	21.21	42.75
2	2.93	0.32	2.80	13.83	2.80	15.10	41.96
3	2.84	0.36	2.40	16.19	5.45	17.91	41.86

CL3_ lsegm3	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	Max_MilkYield	MIDs
0	3.19	0.31	2.80	14.33	4.72	16.07	109.71
1	3.07	0.28	3.02	11.64	1.51	12.62	114.63
2	3.31	0.34	2.46	17.60	11.18	20.14	93.29

CL4_ lsegm4	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	Max_MilkYield	MIDs
0	3.30	0.29	2.69	12.73	3.07	14.02	262.11
1	3.11	0.26	2.70	10.41	1.40	11.21	252.20
2	3.24	0.35	2.36	17.65	15.56	20.51	247.29
3	3.28	0.31	2.57	14.77	6.59	16.77	260.14

Forrás: saját számítások

Az első és a harmadik laktációs szakaszban 3-3, a másik kettőben 4-4 csoportban voltak a leghatékonyabban elkülöníthetők a megfigyelések a napi fejések száma alapján.

Az első laktációs szakaszban napi fejések száma változó mentén kialakított csoportok közül az 1. számú klaszterben a legalacsonyabb a napi fejésszám, ez magas fejésenkénti hozammal és ebben nagy szóródással kapcsolódik. Mind a fejési idő, mid pedig a fejési sebesség alacsony az 2. számú csoportban, amely alacsony hozamokkal társul, ez a csoport termel a legrövidebb idő óta.

A második szakaszban a 2. számú klaszterben a legmagasabb, ez a legmagasabb fejési sebességgel, varianciával és legkisebb napi fejésszámmal jár együtt.

A harmadik szakaszban a legnagyobb fejési sebesség, leghosszabb boxban töltött idő a legnagyobb hozammal és a csoportok között átlagosan mintegy 10 nappal kevesebb ideje termelő egyedek adatait látjuk a 2. számú klaszter esetében.

A negyedik szakaszban a relatív gyors tejleadás mellett hosszabb boxidő, a legnagyobb hozamok és a legalacsonyabb napifejésszám jellemzi a 2. számú klaszter egyedeit. A sokszori

látogatás, alacsony hozamok mellett jelentkezik az 1. számú csoportban, akik a második legkevesebb ideje termelő egyedeket tartalmazza.

14. táblázat: Fejésenkénti tejhozam alapján kialakított klaszterek átlagos jellemzői az első laktációban termelő egyedek esetén

CL3_lsegm1	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	MIDs	Max_MilkYield
0	2.79	0.37	2.20	16.61	6.63	25.34	18.15
1	2.73	0.31	2.75	12.43	5.33	19.96	14.09
2	2.61	0.27	2.80	8.74	3.49	11.87	9.92

CL4_lsegm2	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	MIDs	Max_MilkYield
0	2.38	0.33	3.20	10.65	2.17	42.26	11.79
1	2.90	0.35	2.40	16.23	6.18	41.86	17.97
2	2.83	0.33	2.91	13.38	4.38	41.93	14.94
3	2.90	0.40	2.05	19.59	6.90	42.69	21.08

CL3_lsegm3	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	MIDs	Max_MilkYield
0	3.32	0.34	2.31	18.12	7.45	91.62	19.90
1	3.07	0.28	3.07	11.49	3.19	113.87	12.82
2	3.19	0.31	2.80	14.37	5.16	110.32	16.17

CL4_lsegm4	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	MIDs	Max_MilkYield
0	3.12	0.26	2.73	10.17	3.14	252.24	11.30
1	3.22	0.32	2.50	14.98	6.32	257.82	16.75
2	3.32	0.28	2.74	12.65	3.68	262.30	14.10
3	3.23	0.36	2.20	18.35	9.86	249.04	20.31

Forrás: saját számítások

A fejésenkénti tejhozam alapján történt csoportosítás még szembetűnőbben mutatja a fejési gyakorisággal való negatív összefüggést. A rosszabb teljesítményű egyedek csoportjára a kevesebb ideje tartó termelés a jellemző az első szakaszban, míg ez a tendencia megfordul a harmadik szakaszra. A negyedik szakaszban a termelés visszatér nagyjából a második szakasz szintjére, de nagyobb tejleadási sebesség jellemző.

15. táblázat: Fejési hozam maximum értéke alapján kialakított klaszterek átlagos jellemzői az első laktációban termelő egyedek esetén

CL5_lsegm1	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	Max_MilkYield	MIDs
0	2.71	0.32	2.73	12.35	4.37	13.89	20.09
1	2.69	0.28	2.87	10.12	2.39	11.25	14.57
2	2.82	0.38	2.24	17.74	14.54	20.27	26.26
3	2.49	0.27	2.65	7.81	1.11	8.48	9.91
4	2.80	0.34	2.42	14.90	7.35	16.75	23.51

CL4_lsegm2	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	Max_MilkYield	MIDs
0	2.91	0.33	2.81	13.77	2.53	15.00	42.02
1	2.82	0.36	2.42	16.12	5.59	17.87	41.79
2	2.98	0.38	2.28	18.79	12.07	21.21	42.75
3	2.40	0.34	3.06	11.18	1.25	12.04	41.89

CL5_lsegm3	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	Max_MilkYield	MIDs
0	3.16	0.29	3.01	12.65	2.07	13.84	115.11
1	3.26	0.33	2.57	16.32	8.28	18.56	96.75
2	3.19	0.31	2.80	14.33	4.65	16.09	110.47
3	2.92	0.27	2.95	10.55	1.15	11.34	113.61
4	3.41	0.35	2.39	19.04	15.03	21.94	91.38

CL4_lsegm4	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	Max_MilkYield	MIDs
0	3.28	0.31	2.59	14.69	6.37	16.67	260.26
1	3.11	0.26	2.70	10.41	1.40	11.21	252.20
2	3.24	0.34	2.37	17.51	15.17	20.33	247.18
3	3.30	0.29	2.69	12.72	3.03	14.00	262.42

Forrás: saját számítások

A fejésenkénti hozam alapján 5 jól elkülöníthető csoportokban a termelésben töltött idővel növekszik a termelés és a tejleadás sebessége az első szakaszban. A harmadik szakaszban a másodikhoz képest gyakoribbak és gyorsabbak a fejések, a negyedik szakaszban egy gyengébben teljesítő csoport rajzolódik kis, lassú a tejleadás, gyakrabban látogatják a fejőgépet és kevesebb időt is töltenek ott.

16. táblázat: Fejési hozam varianciája alapján kialakított klaszterek átlagos jellemzői az első laktációban termelő egyedek esetén

CL3_lsegm1	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	MIDs	Max_MilkYield
0	2.83	0.33	2.33	13.98	24.68	20.68	17.74
1	3.53	0.24	2.00	9.27	156.91	18.33	18.03
2	2.69	0.31	2.67	12.07	2.70	18.45	13.31

CL3_lsegm2	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	MIDs	Max_MilkYield
0	2.78	0.35	2.66	14.76	2.42	42.03	15.93
1	2.95	0.35	2.51	16.32	16.80	42.28	19.72
2	3.04	0.35	2.00	17.26	68.64	42.75	23.03

CL3_lsegm3	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	MIDs	Max_MilkYield
0	3.15	0.30	2.84	13.79	2.53	109.12	15.08
1	3.33	0.31	2.62	15.69	16.56	102.11	19.22
2	2.85	0.33	2.41	13.28	76.23	105.76	19.75

CL3_lsegm4	Mspeed	robotTime	count_Milking	MilkYield	Var_MilkYield	MIDs	Max_MilkYield
0	3.23	0.29	2.67	12.83	2.21	257.15	13.98
1	3.23	0.32	2.24	14.70	62.69	251.94	20.62
2	3.32	0.31	2.44	14.70	15.08	260.50	17.88

Forrás: saját számítások

A fejési hozam varianciájával képzett klaszterekben hasonló termelési átlagú, de jelentősen eltérő szóródású adatok figyelhetők meg.

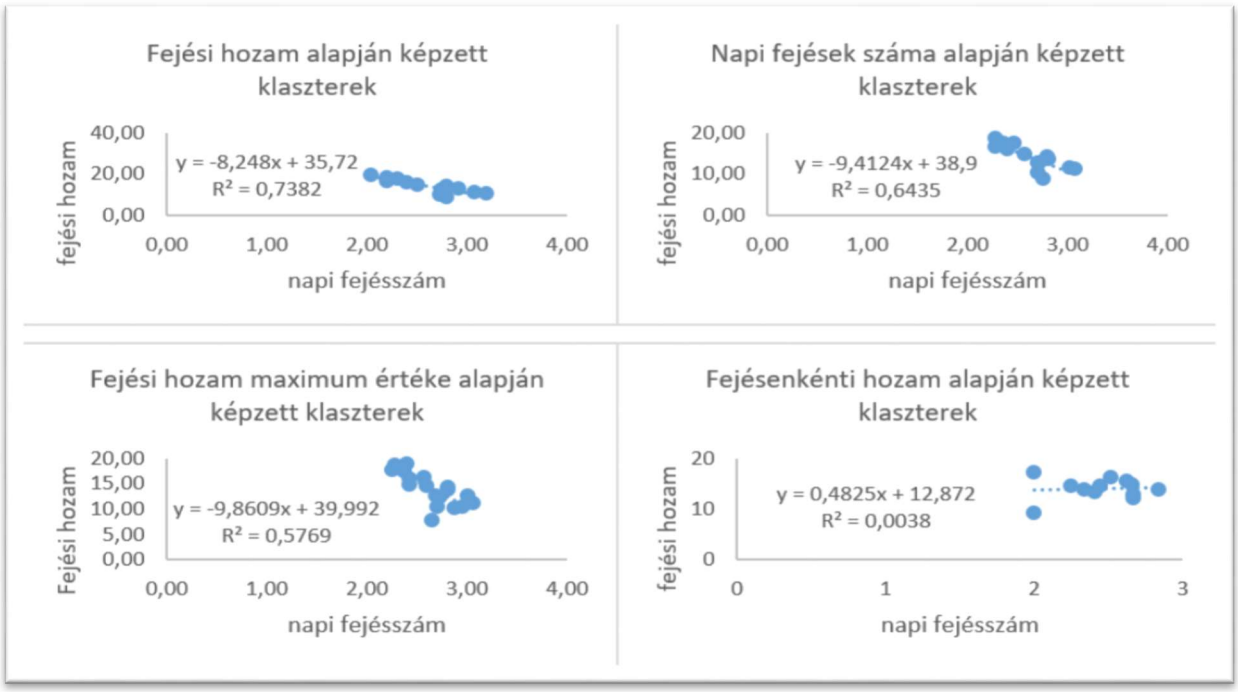
Fentiekhez hasonlóan elvégezhető a tipikus tulajdonságú csoportok leírása a további laktációkban is.

Statisztikai eljárásokkal (pl.egyutas varianciaanalízissel) tovább jellemezhető a csoportok közötti különbségek szignifikanciaszintje. Jelen esetben az klaszterezések eredményei közötti különbségre mutatok rá két változó (a fejések napi száma és a tejhozam) összefüggése (R2 és lineáris regresszió együtthatója) alapján.

Laktációk szerinti eredmények összevetése a klaszterezési eljárás eredményei alapján

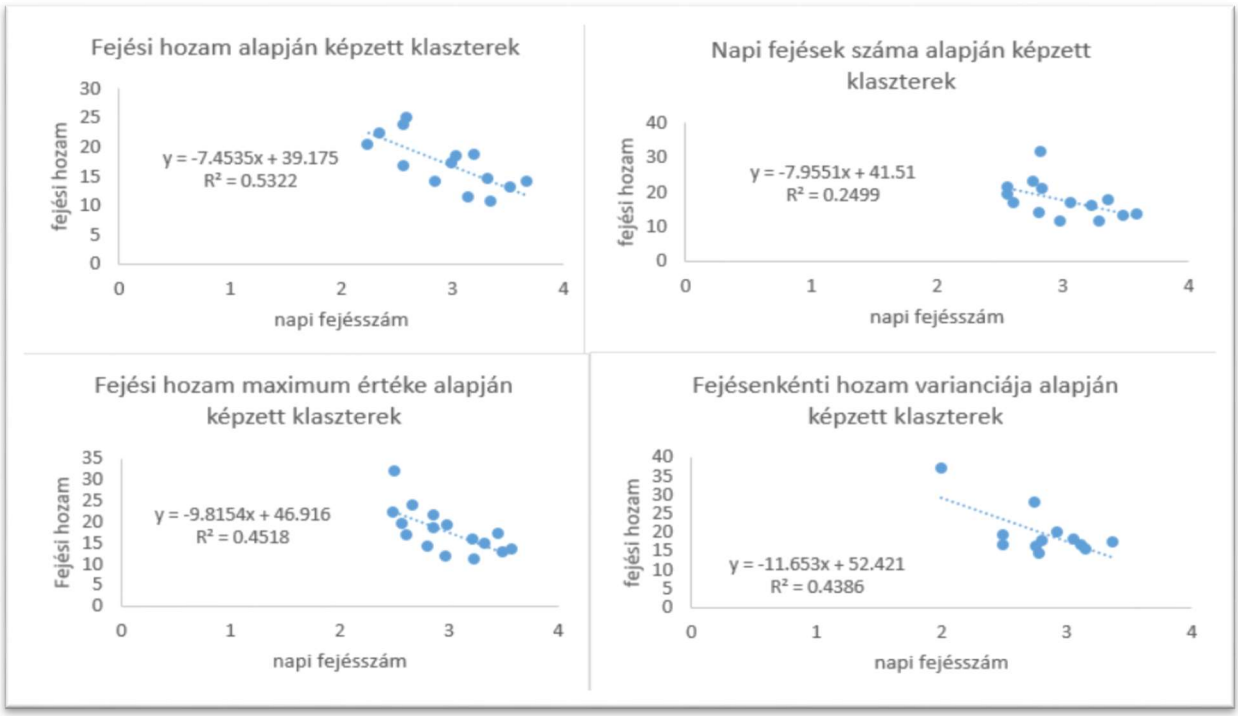
Lineáris modelleket illesztettem rendre a klaszterátlagokra. Az eredmények az adott laktációban termelő egyedek laktációs szakaszaira jellemző mintázatokat jelezhetik, ha összevetésre kerülnek a további termelési ciklusokban levő egyedcsoportok mintázataival. Betekintési céllal a 19. és 20. ábrákon az első és a második laktációs időszak szakaszaira illesztett lineáris trendet és egyenletét, valamint determinációs együtthatóját mutatom be.

19. ábra: A napi fejésszám és fejésenkénti tejhozam összefüggése a négy változó alapján képzett klaszterekben, első laktáció



Forrás: saját számítások és ábrázolás

20. ábra: A napi fejésszám és fejésenkénti tejhozam összefüggése a négy változó alapján képzett klaszterekben, második laktáció



Forrás: saját számítások és ábrázolás

A fenti eredmények (19. és 20. ábrák) alapján tehető főbb megállapítások az alábbiak.
A napi fejésszám alapján klaszterezett megfigyelések esetén a napi fejésszám és a fejésenkénti hozam között az első laktációban erősebb és kifejezettebb kapcsolat áll fenn, mint a második laktációban (R-négyzet 0,75, és 0,53, béta értéke -8,4 és -7,3). A fejési hozam maximuma alapján képzett csoportokban a tendencia hasonló (béta értéke -8,9 és -8,8), de a második laktációban erősebben szóródó adatok miatt a kapcsolatszorosság mértéke alacsonyabb (R-négyzet 0,58 és 0,45). A nap fejések száma alapján kialakított klaszterekben még érdekesebb megfigyelés, hogy a második laktációban mindössze 24%-os, míg az elsőben 64%-os a modell magyarázó ereje (r-négyzet alapján). Ez azt jelezheti, hogy egyöntetűbb az állomány fejési gyakoriságának mintázata, de szórtabb a teljesítménye az első laktációhoz képest. Ezt a megfigyelést látszik igazolni a fejési átlagok varianciája alapján képzett klaszterbe a fejési gyakoriság és a fejésenkénti hozam közötti összefüggés (r-négyzet értékei 0,004, 0,44).
A vizsgálat tovább vihető a csoportok más paramétereinek összevetése és további laktációk eredményeivel való vizuális és leíró összehasonlítás elvégzésével. Ezeket az eredményeket részletesen nem célom a dolgozat terjedelmi korlátai miatt bemutatni.

4.4.2. A klaszterhez tartozás mintázata

A vizsgálatban nagyszámú adathalmazból képzett részadathalmazok elemszáma nehézkessé teszi a megfigyelések klaszterhez tartozásának konkrét, leíró bemutatását. Az ide vonatkozó eredmények 4 laktációs csoport és laktációnként 4 termelési szakasz adattábláiban a vizsgálat alapján 3-5 klaszterszámhoz tartozó egyedeket tartalmazzák (összesen 234 db klaszter).
A klaszterekbe tartozó megfigyelések számát a 17. táblázat mutatja be.

17. táblázat: Klaszterekbe tartozó megfigyelések elemszámának összege laktációk és laktációs szakaszok szerint

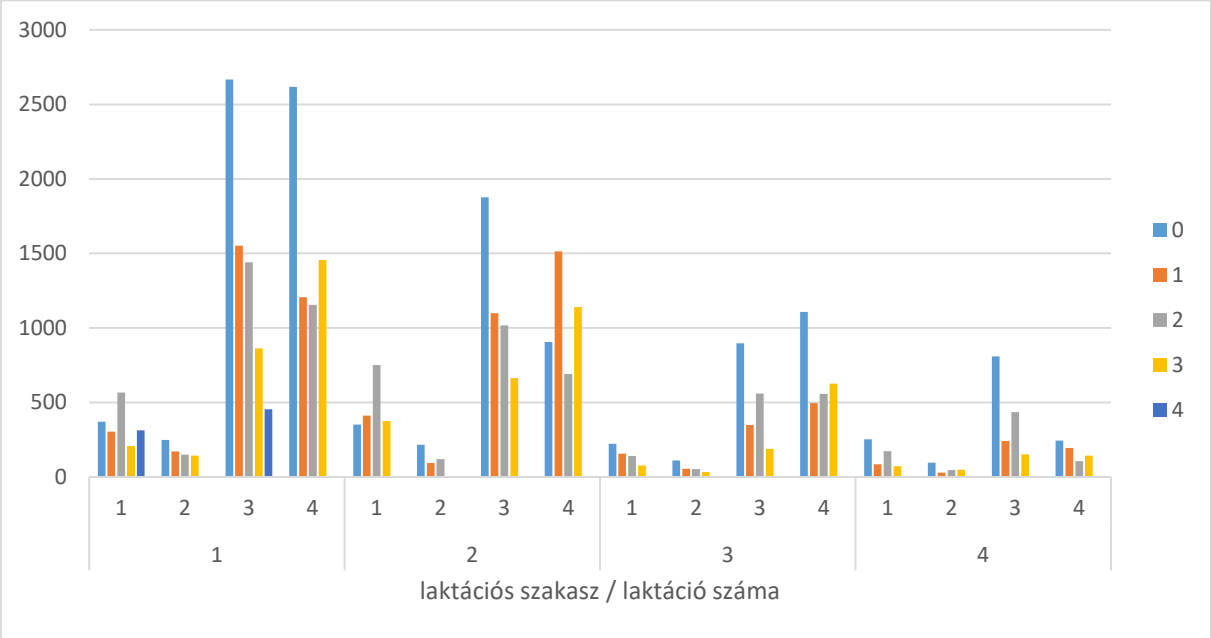
		klaszterek száma					
laktáció száma	***laktációs szakasz***	***0***	***1***	***2***	***3***	***4***	***Összesen***
1	***1***	*1483*	*1214*	*2267*	*207*	*313*	*5484*
	2	*994*	*681*	*600*	*429*		*2704*
	3	*10671*	*6207*	*5762*	*862*	*454*	*23956*
	4	*10467*	*4822*	*4618*	*4365*		*24272*
	Összesen	*23615*	*12924*	*13247*	*5863*	*767*	*56416*
2	***1***	*1407*	*1649*	*3005*	*1123*		*7184*
	2	*865*	*376*	*479*	*4*		*1724*
	3	*7508*	*4393*	*4071*	*664*		*16636*
	4	*3627*	*6051*	*2763*	*3423*		*15864*
	Összesen	*13407*	*12469*	*10318*	*5214*		*41408*
3	***1***	*891*	*627*	*563*	*227*		*2308*

	2	*448*	*225*	*215*	*104*		*992*
	3	*3586*	*1398*	*2238*	*562*		*7784*
	4	*4430*	*1979*	*2235*	*1880*		*10524*
	Összesen	*9355*	*4229*	*5251*	*2773*		*21608*
4	***1***	*1011*	*343*	*691*	*215*		*2260*
	2	*385*	*120*	*184*	*147*		*836*
	3	*3234*	*972*	*1745*	*453*		*6404*
	4	*978*	*783*	*431*	*432*		*2624*
	Összesen	*5608*	*2218*	*3051*	*1247*		*12124*
Összesen		*51985*	*31840*	*31867*	*15097*	*767*	*131556*

Forrás: saját számítások

A megfigyelések csoportonkénti és klaszterenkénti átlagos száma mintegy 500 db (21. ábra), így ezek tételes, táblázatos bemutatásától el kell tekinteni. A 234 db klaszter közötti legkisebb elemszámúak jellemzően a laktáció 2. szakaszának (csúcs) vagy a nagyobb laktációszámú egyedek 1. és 2. laktációs szakaszának klasztercsoportjai.

21. ábra: Klaszterekbe tartozó megfigyelések elemszámának átlaga laktációk és laktációs szakaszok szerint



Forrás: saját számítások és ábrázolás

Tovább lebontva az egyes klaszterképző változók szerint a laktációk és laktációs szakaszok által definiált részadatokra létrehozott csoportokat, (18. táblázat a, b, c, d) látjuk, hogy mennyire szelektív az adott változó a részadathalmazban levő megfigyelésekre.

18. táblázat: Klaszterekbe tartozó megfigyelések eloszlása laktációk és laktációs szakaszok szerint a klaszterképző változókra bontva

a,
változó: fejések napi száma

csoportba tartozó megfigyelések száma, db)

	1. laktáció					2. laktáció					3. laktáció					4. laktáció				
klaszterek	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.
0	45.1%	18.8%	44.7%	38.2%	40.7%	38.3%	51.5%	30.7%	25.6%	30.9%	24.3%	30.6%	39.7%	24.9%	30.5%	31.9%	21.1%	43.0%	16.0%	33.6%
1	21.2%	19.8%	35.3%	26.1%	29.2%	29.1%	18.8%	25.5%	27.8%	26.7%	35.4%	27.4%	30.0%	12.3%	21.8%	23.7%	28.2%	24.8%	34.0%	26.8%
2	33.7%	32.0%	20.1%	9.6%	17.5%	32.2%	29.7%	43.8%	12.7%	29.3%	38.0%	37.9%	29.1%	27.9%	29.9%	41.6%	4.3%	31.0%	34.8%	31.9%
3		29.4%		26.0%	12.6%	0.3%			34.0%	13.1%	2.4%	4.0%	1.2%	34.9%	17.8%	2.8%	46.4%	1.2%	15.2%	7.7%

b,
változó: fejési hozam maximális értéke

csoportba tartozó megfigyelések száma, db)

	1. laktáció					2. laktáció					3. laktáció					4. laktáció				
klaszterek	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.
0	30.2%	32.1%	29.4%	25.6%	28.0%	10.1%	48.3%	28.7%	33.5%	28.1%	24.3%	26.6%	28.7%	36.0%	31.7%	30.4%	42.6%	43.1%	32.3%	38.4%
1	22.6%	30.5%	19.6%	26.1%	23.2%	33.6%	26.7%	21.8%	12.6%	20.5%	35.5%	30.6%	29.3%	11.6%	21.4%	23.7%	3.8%	24.7%	36.4%	25.6%
2	9.3%	19.8%	29.1%	10.6%	18.8%	27.0%	24.1%	33.5%	26.3%	29.2%	37.8%	38.7%	40.8%	25.8%	33.1%	43.0%	32.5%	31.0%	14.9%	29.9%
3	15.1%	17.6%	14.4%	37.6%	24.6%	29.3%	0.9%	16.0%	27.6%	22.1%	2.4%	4.0%	1.1%	26.6%	13.8%	2.8%	21.1%	1.2%	16.3%	6.1%
4	22.8%		7.6%		5.4%															

c,
változó: fejési hozam

csoportba tartozó megfigyelések száma, db)

	1. laktáció					2. laktáció					3. laktáció					4. laktáció				
klaszterek	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.
0	23.6%	12.9%	18.3%	24.3%	21.1%	13.9%	11.8%	37.2%	11.6%	22.3%	8.1%	29.4%	26.9%	22.0%	22.6%	34.2%	37.8%	34.5%	26.7%	33.0%
1	44.6%	34.3%	35.3%	26.1%	32.2%	28.7%	41.5%	42.8%	36.0%	37.7%	35.7%	27.4%	12.1%	36.8%	27.4%	12.0%	23.4%	10.9%	27.7%	15.6%
2	31.8%	36.4%	46.4%	41.3%	42.3%	24.6%	46.6%	20.0%	27.6%	24.8%	21.7%	9.3%	34.4%	31.2%	30.3%	21.4%	35.9%	28.7%	11.3%	24.1%
3		16.4%		8.3%	4.4%	32.9%			24.8%	15.2%	34.5%	33.9%	26.6%	10.0%	19.7%	32.4%	2.9%	25.9%	34.3%	27.4%

d,
változó: fejési hozam varianciája

csoportba tartozó megfigyelések száma, db)

	1. laktáció					2. laktáció					3. laktáció					4. laktáció				
klaszterek	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.	1. szakasz	2. szakasz	3. szakasz	4. szakasz	össz.
0	9.2%	83.3%	85.9%	84.3%	77.6%	16.0%	89.1%	83.9%	20.8%	48.2%	97.7%	94.0%	89.0%	85.4%	88.4%	82.5%	82.8%	81.4%	74.1%	80.1%
1	0.2%	16.1%	13.5%	1.1%	7.0%	0.4%	0.2%	15.5%	76.2%	35.5%	2.1%	5.2%	0.4%	14.5%	7.7%	1.2%	1.9%	0.2%	21.2%	5.1%
2	90.6%	0.6%	0.6%	14.6%	15.4%	83.5%	10.7%	0.5%	3.1%	16.3%	0.2%	0.8%	10.6%	0.0%	3.9%	16.3%	15.3%	18.4%	4.7%	14.8%

Forrás: saját számítások

Azokban a részadathalmazokban, ahol a klasztercsoportok aránya nagyobb szóródást mutat, értelmezhetjük a klaszterképző változó nagyobb mértékű szelektív tulajdonságát. Ez fennáll például a 2. laktáció 2. szakaszában és a 3. laktáció 1. szakaszában is minden klaszterképző változó esetében, az 1. laktáció 4. szakaszában és a 4. laktáció 2. szakaszában pedig három klaszterképző változó esetében. Lényeges információkat nyerhetünk az üzemi menedzsment számára a megfigyelt változókra vonatkozóan nagyon eltérő mintázatú tulajdonsággal rendelkező egyedek csoportjaira.

4.5. Perzisztencia és csoportállandóság vizsgálatok megalapozása

Az eddig bemutatott lépések és részeredmények vezetnek ahhoz a végső célkitűzéshez, hogy a az egyedek a tejtermelésük (gyakoriság, átlagos fejési hozam, variancia, max termelés) alapján történt csoportbasorolásukat (klaszterben ahol vannak) egymáshoz képest megtartják-e a

termelés során; vagy megfordítva, van-e átrendeződés a megelőző laktációs szakaszok klaszterei között (pl. a kezdeti szakasz jó, illetve rosszabb valamint a csúcs, illetve a plató és leszálló szakaszok különböző termelési tulajdonságú csoportjai között). A végső eredmények arra engednek következtetni majd, hogy az egyedek relatív termelőképessége hogyan változik, mely tulajdonságok határozzák ezt meg az adott csoportban.

Annak érdekében, hogy a Bonora és munkatársai (2018) által eredményesen használt klaszter-gráf modellt a későbbiekben alkalmazzam, még arra volt szükség, hogy meghatározzam azoknak az egyedeknek a listáit, amelyekre vonatkozóan ismertek több laktációs szakaszban a termelési adatai. Másrészt, a továbbiakban fel kell tárni az azonos klaszterekbe tartozó egyedeket. Bonora és munkatársai módszere alapján a közös klaszterbe tartozó egyedek (csomópontok) közötti kapcsolatokat (élek) kell létrehozni. Ennek a feladatnak a megoldása a jelen dolgozat területén kívül esik, a következőkben a több szakaszban ismert termelési adatokkal rendelkező egyedek csoportjainak meghatározását mutatom be.

A programom létrehozta azokat a fájlokat, amelyek a klaszter címkéket részadattáblánként tartalmazták. Majd arra volt szükségem, hogy ezeket egy közös fájlba összefűzzem és kikeressem azokat az azonosítókat, amelyek több szakaszban is előfordultak.

Első lépésként a fájlok összefűzésére és az összefűzött dataframe excelben történő mentésére készített kódot. A programrész a drive-on az adott mappában levő fájlokat éri el (9. kép), végighalad és beolvassa dataframekbe, illetve összefűzi azokat (10. kép).

9. kép: kódrészlet: A fájlok elérése és közös oszlopfejléc lista létrehozása

```
"""elkészült eredmények összefűzése"""

import pandas as pd
import os
import re
from google.colab import drive
drive.mount('/content/drive')

folder_path = '/content/drive/MyDrive/eredmenyek/'

# Load the master dataframe to extract column names
master_df = pd.read_excel('/content/drive/MyDrive/eredmenyek/master.xlsx')

# Extract column labels from the master dataframe
column_labels = master_df.columns.tolist()

modified_subset_dfs=[]
```

Forrás: saját összeállítás

10. kép: kódrészlet: A fájlok df-be olvasása és összefűzése

```
# iterálás a drive mappájában levő fájlokkal
for filename in os.listdir(folder_path):
    if filename.endswith('.xlsx') and "master" not in filename.lower() and "combined_data" not in filename.lower():
        file_path = os.path.join(folder_path, filename)

        # a résztáblák beolvasása
        subset_df = pd.read_excel(file_path, skiprows=0)  # 2. sortól kezdődnek az adatok

        # Az egységes mezőnevek létrehozása
        subset_df.columns = column_labels[:len(subset_df.columns)]

         # A résztábla elnevezésének hozzáadása (fájlévből)
        subset_df['subset'] = filename

        # a beolvasott résztábla df listába rendezése
        modified_subset_dfs.append(subset_df)

# a fájllista elemeinek összefűzése egyetlen df-be
combined_df_ = pd.concat(modified_subset_dfs, ignore_index=True)

combined_df_.columns
```

Forrás: saját összeállítás

Ezt követően letöltöttem az összefűzött, minden egyedre, minden laktációs szakaszban, minden klaszterejárás eredményét (klaszterhez tartozás) tartalmazó fájlt (adatmérete: 32889 x 20).

A következő lépésben kereszttáblát készítettem az egyedek azonosítójára laktációszám (sorváltozó) és laktációs szakasz (oszlopváltozó) alapján, és összegyűjtöttem a páronként egymást követő laktációs szakaszban szereplő azonosítókat. A csoportokat jelöléssel láttam el, pl. 1. laktáció 1. és 2. szakasza: l1_1-2 jelet kapta. Az eljárás végén az alábbi csoportokat azonosítottam (19. táblázat).

19. táblázat: A vizsgált időszakban páronként egymást követő laktációs szakaszban termelő egyedek

Az egyes laktációk alatti egymást követő szakaszok											
l1_1-2	**l1_2-3**	**l1_3-4**	**l2_1-2**	**l2_2-3**	**l2_3-4**	**l3_1-2**	**l3_2-3**	**l3_3-4**	**l4_1-2**	**l4_2-3**	**l4_3-4**
7972	7929	7835	7388	7268	7223	6820	6834	6539	6144	6447	6379
8024	7923	7820	7095	7307	7277	6907	6864	6622	6507	6313	6468
8016	7919	7750	7256	7385	7237	6968	6942	6880	6513	6261	6141
8014	7916	7851	7340	7278	7339	6814	6921	6697	6426	6302	6340
7982	7903	7706	7341	7141	7172	7302	6562	6808	6469	5838	6323
8021	7758	7743	7416	7140	7016	6672	6772	6603	6381	6486	6139
7908	7963	7752	7418	7150	7220	6766	6792	6818	6300	6558	6311
7912	7845	7813	7462	7300	7265	6867	6983	6664	6159	6159	6190
7858	7948	7639	7381	7468	7088	6903	6903	6663	6558	6300	6377
7927	7976	7664	7481	7485	7238	6983	6867	6747	6486	6381	6308
7961	7902	7777	7054	7396	7335	6792	6766	6652	5838	6507	6198
7931	7825	7629	7480	7367	7073	6772	6672	6653	6261	6426	5973
7981	7806	7681	7402	7401	7234	6921	6820	6701	6302	6469	6153
7942	7857	7718	7463	7283	7110	6562	6907	6545	6313	6513	6157
7945	7955	7483	7229	7256	7121	6942	6968	6694	6447	6590	
7964	7925	7657	7288	7462	7139	6864	7302	6826			
7952	7867	7738	7301	7439	7222	6834	6756	6557			
8009	7829	7739	7439	7229	7231		6844	6789			
7995	7940	7763	7284	7301	7235			6638			
7767	7972	7711	7428	7340	7117						
7986	7908	7759	7283	7416	7236						
7977	7935	7709	7401	7418	7255						
7935	7767	7600	7367	7388	7230						
8015	7931	7723	7396	7341	7209						
7999	7981	7598	7485	7288	7170						
7829	7986	7710	7150	7381	7276						
7940	7987	7771	7300	7189	7296						
7987	7942	7674	7468	7197	7013						
7960	7912	7672	7140	7348	7201						
7867	7961	7766	7509	7380	7076						
7925	7927	7694	7141	7334	7108						
7955	7858	7734	7278		7025						
7857	7977	7693	7385		7205						
7806	7945	7697	7307		7113						
7825	7964	7651	7268		7145						
7902	8021	7707			7146						
7976	8024	7638			7097						
7948	7888	7690			6956						
7845	7801	7615			7116						
7963	7979	7631			7132						
7903	7649	7642			7165						
7758	7855	7684									
7916	7907	7526									

7919	7810 7889 7736 7727 7868 7897 7862	7647 7687 7645 7488 7712									
44 db	*50 db*	*48 db*	*35 db*	*31 db*	*41 db*	*17 db*	*18 db*	*19 db*	*15 db*	*15 db*	*14 db*

A további lépések annak érdekében, hogy a klaszter-gráf módszertan szerinti elemzést majd elvégezzem, az azonos klaszterekbe tartozó egyedek feltárása lesz. Ezt követően létre kell hozni a hálózat alapjait képező megfelelő node és edges táblákat. Bonora és munkatársai módszere alapján a közös klaszterbe tartozó egyedek a csomópontok (node), a közöttük lévő kapcsolatokat (edges, élek) az együttes klaszterbe tartozás jelenti. Ennek a feladatnak a megoldása már a jelen dolgozat területén kívül esik.

Összefoglalóan, a perzisztencia és csoportállandóság vizsgálatok megalapozása azokat a lépéseket tartalmazta, amellyel eljutottam vizsgálati végső célkitűzés megalapozásához. A kapott eredmények alkalmasak, hogy klaszter-gráf módszertan alkalmazásával kimutassam, az egyedek a tejtermelésük (gyakoriság, átlagos fejési hozam, variancia, max termelés) alapján történt csoportba sorolásukat (klaszterben, ahol vannak) egymáshoz képest megtartják-e a termelés során. Illetve, van-e átrendeződés a megelőző laktációs szakaszok klaszterei között (pl. a kezdeti szakasz jó, illetve rosszabb valamint a csúcs, illetve a plató és leszálló szakaszok különböző termelési tulajdonságú csoportjai között). A végső eredmények arra engednek következtetni majd, hogy az egyedek relatív termelőképessége hogyan változik, mely tulajdonságok határozzák ezt meg az adott csoportban.

5. Következtetések és javaslatok

A fejezetben a dolgozat résztémakörei szerint haladva fogalmazom meg következtetéseimet, javaslataimat.

A szakirodalomkutatás megerősítette, hogy a robotfejőrendszerek terjedésével a gazdaságokban nagy adattömegek keletkeznek, amelyek felhasználására szaktanácsadók és gazdálkodók keresik azokat a módszereket, amelyekkel hatékonyabb termék előállítás érhető el. A gépi tanulási eszközök alkalmazása túlmutat az integrált fejőrobot rendszereken alapuló, napi szintű vezetői döntéstámogatáson. A robotfejő rendszerekben keletkező adatok nagy része azonban ma nem hasznosul, pedig a cloud technológia által teremtett lehetőségek, a gépi tanulási eljárások és programozói környezetük révén eddig nem, vagy általánosságban ismert, összefüggések üzemi körülmények között felparaméterezhető egyedi döntéstámogató információkat szolgáltathat.

A robotfejőrendszerek adatainak gépi tanulási eljárásokkal való elemzése nagyságrenddel nagyobb a közvetlen gazdasági hatással járó sántaság és mastitis azonosításának témakörében, és jóval kevesebb a robotok gazdasági hatékonyságát befolyásoló tényezők vizsgálatának területén. Kevés azoknak a tanulmányoknak a száma, amelyek üzemi adatok alapján feltárható termelési mintázatok azonosítása révén a menedzsment szintű gyakorlatok átgondolását tehetik lehetővé. Egyik ilyen mintázat a termelő tehenek csoportállandósága, amely a szakirodalom alapján az egyedek viselkedési dinamizmusát befolyásolhatja (végső soron a robotfejőgéphez való hozzáférést, így a termelés hatékonysági mutatóit).

A dolgozat célja olyan eljárásnak az azonosítása és módszertani kereteinek definiálása, amely az egyedcsoportok tejhozam mutatói alapján képes leírni az egyedek termelési csoportok közötti relatív egyedi pozíciójának a megváltozását. Az alkalmazott szakirodalomkutatási módszertan (keresőalgoritmusok, és módosításuk, hivatkozási hálózat feltárása) megfelelően azonosította a témában benchmarkként felhasználható tanulmányokat, amelyek a módszertani elméleti keretet definiálták. Az elemzést python programnyelven, Jupiter notebookon, a colab.research platformon meg tudtam valósítani. A program további fejlesztése még járhat a hatékonyság javulásával, különösen a helyi df változókból történő közvetlen eredmények kinyerésén keresztül, amit most a dataframe változók excelben mentésével és kereszttábla elemzés alkalmazásával oldottam meg, több esetben az excel eredményfájlok visszaolvasásával és manipulálásával.

A vizsgálatok során a tejtermelő egyedek négy laktációs csoportján belül további négy laktációs szakaszban, különböző termelési tulajdonságok mentén nagyon jó SM statisztikával jellemezhető klasztereit tudtam elkülöníteni. A fejlesztett program alkalmas a vizsgálati célkitűzésnek megfelelő algoritmus megoldást nyújtani.
A dolgozat terjedelmébe tartozó végső eredmények definiálták azokat az egyedcsoportokat, ahol a csoportba sorolás laktációs szakaszonkénti változása vizsgálható a későbbiekben. Ez megalapozza azt, hogy következtetni lehet az egyedek relatív termelőképességének változásaira és az egyes csoportok közötti dinamikára.
A kutatás korlátai között megnevezhető a rendelkezésre álló termelési idősor rövidsége (90 termelési nap). További vizsgálatok széles köre vázolható, mind a képezett csoportok részletes jellemzése, az egyes mutatók csoportok közötti szignifikáns eltéréseinek vizsgálata, mind pedig az elemzések automatizáltságának javítása terén, ami a jelenleg táblázatkezelőben megoldott feladatok python nyelvre történő megoldásának kidolgozásába fogalmazható meg. A dolgozatban elért eredmények alapján tervezem megvalósítani a klaszter-gráf modellt a későbbiekben, amelyet tudományos publikációban tervezek közölni.

6. Összefoglalás

A dolgozat a robotfejőrendszerek terjedésével egyre nagyobb mennyiségben és több üzemben elérhetővé váló nagy adatmennyiség feldolgozásában rejlő potenciálokkal, ezen belül a termelési hatékonyságot várhatóan befolyásoló termelési csoportok azonosításával foglalkozik.
A vizsgálati célkitűzés az, hogy benchmark szakirodalmi tanulmányok módszertanának adaptálásával azonosítson és jellemezhetővé tegyen egyedcsoportokat bizonyos termelési mutatói alapján és a termelési csoportok közötti perzisztencia vizsgálatát megalapozó adatbázist hozzon létre.
A szakdolgozat e cél elérése érdekében foglalkozik a szakirodalom gyűjtésével, strukturált elemzésével és a módszertani elméleti keret definiálásával; továbbá lényegi részét képezi az elemzéseket megvalósító eljárások, az alkalmazott nem felügyelt gépi tanulási módszernek a felhasznált adatbázison történő alkalmazásához szükséges program fejlesztése.
Kitekintést nyújt a program fejleszthetőségének lehetőségeire, illetve a kutatás tovább vitelének vonalára. A végső eredmények hozzájárulnak olyan menedzsment döntések megalapozásához, vagy átgondolásához, amely az egyedek relatív termelőképességét, illetve annak változásait figyelembe veszi.

7. Irodalomjegyzék

1. Aerts, J., Kolenda, M., Piwczynski, D., Sitkowska, B., Onder, H., 2022. Forecasting Milking Efficiency of Dairy Cows Milked in an Automatic Milking System Using the Decision Tree Technique. Animals 12, 1040. https://doi.org/10.3390/ani12081040
2. Antanaitis, R., Zilaitis, V., Juozaitiene, V., Noreika, A., Rutkauskas, A., 2018. Evaluation of rumination time, subsequent yield, and milk trait changes dependent on the period of lactation and reproductive status of dairy cows. Pol. J. Vet. Sci. 21, 567–572. https://doi.org/10.24425/124291
3. Application of machine learning to improve dairy farm management: A systematic literature review Naftali Slob a , Cagatay Catal b , Ayalew Kassahun a, * a Information Technology Group, Wageningen University & Research, Wageningen, The Netherlands b Department of Computer Engineering, Bahcesehir University, Istanbul, Turkey
4. Bonora, F., Benni, S., Barbaresi, A., Tassinari, P., Torreggiani, D., 2018. A cluster-graph model for herd characterisation in dairy farms equipped with an automatic milking system. Biosyst. Eng. 167, 1–7. https://doi.org/10.1016/j.biosystemseng.2017.12.007
5. Castro, A., Pereira, J.M., Amiama, C., Bueno, J., 2012. Estimating efficiency in automatic milking systems. J. Dairy Sci. 95, 929–936. https://doi.org/10.3168/jds.2010-3912
6. Farkas, G., Magyar, P., Molnár, A., & Zubor-Nemes, A. (2020). Adatbányászati módszerek alkalmazás a mezőgazdaságban–a gépi tanulás felhasználási lehetőségei. Gazdálkodás: Scientific Journal on Agricultural Economics, 64(1), 15-24.
7. Horváthné Kovács, B., Zörög, Z., & Bus, B. (2024). Mesterséges intelligencia a precíziós állattartók vezetői döntéseiben: Bíznak-e a gazdák az adatokban? Gazdálkodás, Scientific Journal on Agricultural Economics, 68(1), 18-42.
8. Ji, B., Banhazi, T., Ghahramani, A., Bowtell, L., Wang, C., Li, B., 2020. Modelling of heat stress in a robotic dairy farm. Part 3: Rumination and milking performance. Biosyst. Eng. 199, 58–72. https://doi.org/10.1016/j.biosystemseng.2020.02.006
9. Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A. (1996). "Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming". Artificial Intelligence in Design '96. Artificial Intelligence in Design '96. Springer, Dordrecht. pp. 151–170. doi:10.1007/978-94-009-0279-4_9. ISBN 978-94-010-6610-5.

10. Orováné (2024): Bevezetés az adattudományba. Egyetemi oktatási segédanyag. Magyar Agrár- és Élettudományi Egyetem, Műszaki Intézet
11. Ozella, L.; Brotto Rebuli, K.; Forte, C.; Giacobini, M. A Literature Review of Modeling Approaches Applied to Data Collected in Automatic Milking Systems. Animals 2023, 13, 1916. https://doi.org/ 10.3390/ani13121916
12. Rebuli, K.B., Ozella, L., Vanneschi, L., Giacobini, M., 2023. Multi-algorithm clustering analysis for characterizing cow productivity on automatic milking systems over lactation periods. Comput. Electron. Agric. 211, 108002. https://doi.org/10.1016/j.compag.2023.108002
13. Shine, P., Murphy, M.D., 2022. Over 20 Years of Machine Learning Applications on Dairy Farms: A Comprehensive Mapping Study. SENSORS. https://doi.org/10.3390/s22010052
14. Tarr B. (2024): Gépi tanulás. Egyetemi oktatási segédanyag. Magyar Agrár- és Élettudományi Egyetem, Műszaki Intézet

Szoftverek, programok, eljárások

1. Bibliometrix - Download
2. CitNetExplorer - Download
3. https://www.vosviewer.com/downloads/VOSviewer_1.6.20_exe.zip
4. Jupiter notebook (google research collaboration)
5. pandas documentation — pandas 2.2.2 documentation (pydata.org) Date: Apr 10, 2024 Version: 2.2.2 (pandas, scikit-learn, IO,
6. StataCorp. 2023. Stata 18 Multivariate Statistics Reference Manual. College Station, TX: Stata Press.

8. Ábrák és táblázatok jegyzéke

Képek jegyzéke

Táblázatok jegyzéke

Ábrák jegyzéke

9. Hallgatói nyilatkozat

NYILATKOZAT

a szakdolgozat nyilvános hozzáféréséről és eredetiségéről

A hallgató neve: Horváthné dr. Kovács Bernadett
A Hallgató Neptun kódja: D062JY
A dolgozat címe: MINTÁZATFELTÁRÁS NEM FELÜGYELT TANULÁSI MÓDSZERREL ROBOTIZÁLT ÜZEMI TEJTERMELÉSBEN
A megjelenés éve: 2024
A konzulens intézetének neve: Műszaki Intézet
A konzulens tanszékének a neve: Mérnökinformatika Tanszék

Kijelentem, hogy az általam benyújtott szakdolgozat egyéni, eredeti jellegű, saját szellemi alkotásom. Azon részeket, melyeket más szerzők munkájából vettem át, egyértelműen megjelöltem, és az irodalomjegyzékben szerepeltettem.

Ha a fenti nyilatkozattal valótlant állítottam, tudomásul veszem, hogy a záróvizsga-bizottság a záróvizsgából kizár és a záróvizsgát csak új dolgozat készítése után tehetek.

A leadott dolgozat, mely PDF dokumentum, szerkesztését nem, megtekintését és nyomtatását engedélyezem.

Tudomásul veszem, hogy az általam készített dolgozatra, mint szellemi alkotás felhasználására, hasznosítására a Magyar Agrár- és Élettudományi Egyetem mindenkori szellemitulajdon-kezelési szabályzatában megfogalmazottak érvényesek.

Tudomásul veszem, hogy dolgozatom elektronikus változata feltöltésre kerül a Magyar Agrár- és Élettudományi Egyetem MATER Hallgatói Dolgozatok repozitoriumába. Tudomásul veszem, hogy a megvédett és

- nem titkosított dolgozat a védést követően
- titkosításra engedélyezett dolgozat a benyújtásától számított 5 év eltelte után

nyilvánosan elérhető és kereshető lesz az Egyetem MATER Hallgatói Dolgozatok repozitoriumában.

Kelt: Kaposvár, 2024 év április hó 20 nap

Hallgató aláírása

10. Konzulensi nyilatkozat

NYILATKOZAT

Horváthné dr. Kovács Bernadett (név) (hallgató Neptun azonosítója: D062JY) konzulenseként nyilatkozom arról, hogy a szakdolgozatot áttekintettem, a hallgatót az irodalmi források korrekt kezelésének követelményeiről, jogi és etikai szabályairól tájékoztattam.

A záródolgozatot/szakdolgozatot/diplomadolgozatot/portfóliót a záróvizsgán történő védésre **javaslom / nem javaslom**[1].

A dolgozat állam- vagy szolgálati titkot tartalmaz: igen nem[2]

Kelt: Gödöllő, 2024. év április hó 28 nap

belső konzulens

[1] A megfelelő aláhúzandó.
[2] A megfelelő aláhúzandó.

11. Mellékletek

1. sz. melléklet: A szűkített találat szakirodalom listájából kizárt tételek

folyóirat	első szerző	cím	évszám	kizárás oka
ANIMALS	Lessire, F	Systematic Review and Meta-Analysis: Identification of Factors	2020	grazing
ANIMALS	Wieland, M	Risk Factors of Forced Take-Off in Dairy Cows Milked Three Times per Day	2021	nem AMS
JOURNAL OF ANIMAL SCIENCE	Rodrigues, PF	Milk yield and composition from Angus and Angus-cross beef cows raised	2014	out of topic: beef cows
TRANSLATIONAL ANIMAL SCIENCE	Lancaster, PA	Relationships among feed efficiency traits across production segments	2021	out of topic: beef; RFI
JOURNAL OF ANIMAL SCIENCE	Bruckmaier, RM	Induction of milk ejection and milk removal in different production	2008	out of topic: biologicalfactors
JOURNAL OF DAIRY SCIENCE	Mäntysaari, P	Modeling of daily body weights and body weight changes of Nordic Red	2015	out of topic: Bweight
JOURNAL OF DAIRY SCIENCE	Tangorra, FM	Assessment of technical-productive aspects in Italian dairy farms	2022	out of topic: classification of farms
JOURNAL OF DAIRY SCIENCE	Churakov, M	Proposed methods for estimating loss of saleable milk in a cow-calf	2023	out of topic: different herd management
JOURNAL OF ANIMAL SCIENCE	MOODY, DE	CONCENTRATION OF PLASMA-CHOLESTEROL IN BEEF-COWS AND CALVES,	1992	out of topic: biologicalfactors
COMPUTERS AND ELECTRONICS IN AGRICULTURE	Shine, P	Machine-learning algorithms for predicting on-farm direct water and	2018	out of topic: energy&water consumption
JOURNAL OF ANIMAL SCIENCE	Brown, MA	Relationship of sire expected progeny differences to milk yield in	2005	out of topic: genetics
COMPUTERS AND ELECTRONICS IN AGRICULTURE	Tuan, SA	Frequency modulated continuous wave radar-based system for monitoring	2022	out of topic: heat stress
SENSORS	Fuentes, S	Biometric Physiological Responses from Dairy Cows Measured by Visible	2021	out of topic: image analysis
AGRICULTURAL AND FOOD SCIENCE	Koskela, O	Deep learning image recognition of cow behavior and an open data set	2022	out of topic: image analysis
CAIP 2019, PT II	Bhole, A	A Computer Vision Pipeline that Uses Thermal and RGB Images for the	2019	out of topic: image analysis
ANIMAL	Odorcic, M	Review: Milking machine settings, teat condition and milking efficiency	2019	out of topic: mechanikai faktorok hatása a szövetre
2022 33RD (ISSC)	O'Leary, C	An Evaluation of Machine Learning Approaches for Milk Volume Prediction	2022	out of topic: national level
JOURNAL OF DAIRY SCIENCE	Edwards, JP	Milking efficiency for grazing dairy cows can be improved by increasing	2013	out of topic: prestimulation
APPLIED ANIMAL BEHAVIOUR SCIENCE	Marumo, JL	Social associations in lactating dairy cows housed in a robotic milking	2022	out of topic: social associations
JOURNAL OF FOOD COMPOSITION AND ANALYSIS	Muñiz, R	Milk quality control requirement evaluation using a handheld near	2020	speciális eszköz, technika eredményei
JOURNAL OF DAIRY SCIENCE	Magliaro, AL	Automatic cluster remover setting affects milk yield and machine-on time	2005	technikai beállítás vizsgálata