**A HIÁNYZÓ LÁNCSZEM MEGTALÁLÁSA FOOTBALL MANAGER SEGÍTSÉGÉVEL**

**FIND THE MISSING LINK VIA FOOTBALL MANAGER**

**OE-NIK** Hallgató neve: **Munkácsi Zoltán**

**2021** Hallgató törzskönyvi száma: **T/006318/FI12904/N**

Tartalomjegyzék

[1. RÖVID TARTALMI ÖSSZEFOGLALÓ A TÉMA TERÜLETÉRŐL, A FELADATRÓL 5](#_Toc103358345)

[1.1. Bevezetés 5](#_Toc103358346)

[1.2 Terület bemutatása 6](#_Toc103358347)

[2. A MEGOLDANDÓ PROBLÉMA MEGFOGALMAZÁSA 8](#_Toc103358348)

[2.1. A hatékony játékosmegfigyelés fontossága 8](#_Toc103358349)

[2.2 Adatalapú döntéstámogatás megjelenése a footballban 8](#_Toc103358350)

[2.2.1. Adatalapú játékosértékelés 9](#_Toc103358351)

[2.2.2. Előnye a hagyományos módszerekkel szemben 10](#_Toc103358352)

[2.2.3. A játékosok piaci értékének megbecsülése adatok alapján 11](#_Toc103358353)

[2.3. Adatok forrásai 11](#_Toc103358354)

[2.3.1. Football Manager 12](#_Toc103358355)

[2.3.2. Adatok elérése 13](#_Toc103358356)

[3. A SZAKDOLGOZAT CÉLJA 14](#_Toc103358357)

[3.1. Célkitűzés 14](#_Toc103358358)

[3.2. Elérendő állapot 15](#_Toc103358359)

[3.3. Célként kitűzött felhasználási lehetőségek 15](#_Toc103358360)

[4. AZ IRODALOM ALAPJÁN A LEHETSÉGES MEGKÖZELÍTÉSI MÓDOK ÁTTEKINTÉSE ÉS KIVÁLASZTÁSA 15](#_Toc103358361)

[4.1. Megközelítési módok 15](#_Toc103358362)

[4.2. Piacon elérhető hasonló szolgáltatások 17](#_Toc103358363)

[4.3. Eltérés az eddigi megoldásokhoz képest 18](#_Toc103358364)

[5. A RÉSZLETES SPECIFIKÁCIÓ LEÍRÁSA 19](#_Toc103358365)

[5.1. Vázlatos rendszerterv 19](#_Toc103358366)

[5.2. Vázlatos időterv 20](#_Toc103358367)

[5.3. Adatbázis 21](#_Toc103358368)

[5.3.1. Lehetőségek áttekintése 21](#_Toc103358369)

[5.3.2. Választás 24](#_Toc103358370)

[5.4. Webalkalmazás 25](#_Toc103358371)

[5.4.1 Lehetőségek áttekintése 25](#_Toc103358372)

[5.4.2 Választás 27](#_Toc103358373)

[5.5. Gépi Tanulás Komponens 28](#_Toc103358374)

[5.5.1. Választás 28](#_Toc103358375)

[5.6 Adatbetöltő script 28](#_Toc103358376)

[6. A TERVEZÉS SORÁN VÉGZETT MUNKAFÁZISOK ÉS TAPASZTALATAIK LEÍRÁSA 29](#_Toc103358377)

[6.1 Adatbázis tervezése 29](#_Toc103358378)

[6.2 Django webalkalmazás tervezése 30](#_Toc103358379)

[6.3 Gépi tanuló komponens tervezése 30](#_Toc103358380)

[6.4 Adatbetöltő script tervezése 31](#_Toc103358381)

[7. MEGVALÓSÍTÁS LEÍRÁSA 32](#_Toc103358382)

[7.1 Adatbázis kialakítása 32](#_Toc103358383)

[7.2 Webalkalmazás kialakítása 33](#_Toc103358384)

[7.2.1 Környezet kialakítása 33](#_Toc103358385)

[7.2.2 Django projekt felépítése 34](#_Toc103358386)

[7.2.3 Tervezés 34](#_Toc103358387)

[7.2.4 Felhasználói felület 35](#_Toc103358388)

[7.2.5 Logika 37](#_Toc103358389)

[7.2.6 Működés szerint: 37](#_Toc103358390)

[7.3 Gépi tanuló komponens 37](#_Toc103358391)

[7.3.1 Játékos ajánlás 38](#_Toc103358392)

[7.3.2 Alulértékelt játékosok keresése 40](#_Toc103358393)

[7.4 Komponensek közötti kapcsolatok kialakítása 40](#_Toc103358394)

[7.4.1 Adatbázis – Webalkalmazás 40](#_Toc103358395)

[7.4.2 Adatbázis – Gépi tanuló komponens 41](#_Toc103358396)

[7.4.3 Webalkalmazás – Gépi tanuló komponens 42](#_Toc103358397)

[7.5 Adatbetöltő script 42](#_Toc103358398)

[8. TESZTELÉS 44](#_Toc103358399)

[8.1 Tervezés 44](#_Toc103358400)

[8.2 Adatbázis 44](#_Toc103358401)

[8.3 Webalkalmazás és Gépi Tanuló komponens tesztelése 44](#_Toc103358402)

[8.4 Adatbetöltő tesztelése 44](#_Toc103358403)

[9. AZ EREDMÉNYEK BEMUTATÁSA, ÉRTÉKELÉSE 46](#_Toc103358404)

[9.1 Eredmények 46](#_Toc103358405)

[9.1.1 Komponensek 46](#_Toc103358406)

[9.2 Értékelés 46](#_Toc103358407)

[9.2.1 Felhasználási lehetőségek 47](#_Toc103358408)

[10. MEGVALÓSÍTÁS ELEMZÉSE, TOVÁBBFEJLESZTÉSI LEHETŐSÉGEK 48](#_Toc103358409)

[Továbbfejlesztési lehetőségek 48](#_Toc103358410)

[11. A SZAKDOLGOZAT TARTALMI ÖSSZEFOGLALÓJA MAGYAR ÉS ANGOL NYELVEN 49](#_Toc103358411)

[Magyar nyelven 49](#_Toc103358412)

[In english 49](#_Toc103358413)

[12. IRODALOMJEGYZÉK 50](#_Toc103358414)

[13. MELLÉKLETEK 52](#_Toc103358415)

[1. Melléklet – Játékosok pontszámai 52](#_Toc103358416)

[2. Melléklet – Egységtesztek leírása, eredménye 54](#_Toc103358417)

[Webalkalmazás üzleti logikájának egységtesztjei 54](#_Toc103358418)

[Gépi Tanuló komponens üzleti logikájának egységtesztjei 55](#_Toc103358419)

# RÖVID TARTALMI ÖSSZEFOGLALÓ A TÉMA TERÜLETÉRŐL, A FELADATRÓL

## Bevezetés

A nagy számú statisztikai adatok gyökeresen megváltoztatták az utóbbi években a futball világát. Csapat szinten, játékosszinten és a globális folyamatokat tekintve is megváltozott a sikerhez vezető út.

Az első mérkőzés, amelyen mérték a játékosok pozíciójának adatait játék közben a 2006-os Világbajnoki döntő volt. Ezeknek az adatoknak a gyűjtésével és feldolgozásával az edzők sokkal jobban ki tudták használni a csapatuk képességeit. Azáltal, hogy képesek voltak minden játékos pontos mozgását követni, sokkal hatásosabb és részletesebb taktikát tudtak kidolgozni. [4]

A pozíciós mérések felhasználása túlmutat a játékosok és a labda helyzetének ismeretén. A mesterséges intelligenciáknak köszönhetően képesek vagyunk előrejelzést adni a meccsek kimenetére, játékos sérülésekre, hogy mennyire térül meg egy játékos megvásárlása befektetési szempontból és a taktikai döntések hatékonyságára is. Ezeket az mutatókat a csapatok edzői és vezetőségi szakemberei fel tudják használni az edzéseken, a csapatösszeállításban, taktika kiválasztásában és a pályán kívüli tevékenységekben is (rehabilitációs program kialakítása, utánpótlásnevelés). [5]

 A Grand View Research kutatása alapján a globális sportanalitikai piac értéke 2020-ban 1 milliárd dollár körül mozog, az előrejelzések szerint 2025-re akár a 4,5 milliárd dollárt is elérheti, ebből az európai labdarúgó elemzési piac értéke 100-200 millió € is lehet. Ebből látható, hogy a verseny egyre intenzívebb, és azon csapatok, melyek időben lépnek, komoly versenyelőnyre tehetnek szert, alulértékelt játékosokat vásárolnak, előre jelezhetik az ellenfél gyenge pontjait, követhetik a játékosok napi fizikai állapotát, megelőzhetik a sérüléseket, ezáltal felzárkózhatnak a nagyobb költségvetésű klubok mellé. [6]

Később kifejtésre fog kerülni, de a szakdolgozatom célja egy olyan döntéstámogató rendszer létrehozása, ahol az erre pénzügyi eszközöket ráfordítani képtelen csapatok is lehetőséget kapnak számukra értékes, de a piac által alulértékelt játékosok megtalálására és ezáltali szerződtetésére.

## 1.2 Terület bemutatása

Egy sportcsapat játékosokkal foglalkozó személyzete az alábbi szerepkörökre bontható (az ellátószemélyzeten kívül):

1. Edzők és a Menedzsment
2. Sport Science[[1]](#footnote-1) és Orvosi stáb
3. Hatékonyság elemzés
4. Utánpótlás biztosítás

Az egyszerűsített adatfeldolgozási modell (1.1 ábra) alapján láthatjuk, hogy a 2, 3, 4 szerepkörök tagjai alakítják a kutatások eredményét és az összegyűjtött adatokat meglátásokká, tanácsokká. Ezeket a meglátásokat próbálják az Edzői és Menedzsment (1. szerepkör) csoport tagjainak használható formában átadni. Az első szerepkör szakemberei ezeket a meglátásokat figyelembe veszik az edzések tervezésekor, taktikai döntéseknél. Konkrét utasítás formájában jut el a játékosokhoz, akiknek gyakorláskor és meccsen már csak az edzői instrukciókra kell figyelniük. [5]

Ha valamilyen releváns adat összegyűjtésre, majd feldolgozásra kerül, az mindegyik szerepkör szakértőinek felelőssége, hogy eljuthasson az Edzők és Menedzsment csapathoz (1. szerepkör), akik alkalmazhatják az edzések során, a csapatösszeállításkor és a meccstaktikák megválasztásakor. [5]



* 1. Ábra - Szerepkörök és fázisok az adatfeldolgozásban [5]

A fenti ábrán látszódik, hogy a begyűjtött és feldolgozott adatok mekkora része és milyen formában jut el az egyes szerepkörökhöz.

Big Data és Üzleti Intelligencia szakirányos hallgatóként a háromszög aljára szeretnék koncentrálni, ami az adatokkal való munkát jelenti. Ezen belül megvizsgálom, hogy a Scoutok (azaz a játékosmegfigyelők) hogyan tudják felhasználni a rendelkezésükre álló nagy mennyiségű strukturált és strukturálatlan adatot munkájuk során

# A MEGOLDANDÓ PROBLÉMA MEGFOGALMAZÁSA

## 2.1. A hatékony játékosmegfigyelés fontossága

Egy csapat számára hatalmas helyzeti előnyt jelent, ha sikerül egy magas kvalitású játékost áron alul szerződtetniük, tehát kevesebb pénzért, mint amennyit a képességei vagy a benne rejlő lehetőségek miatt ér. Ez hosszú távon kifizetődő és kisebb csapatok esetében a klub pénzügyi fenntartásának egyik eszköze is lehet a megfelelő játékosok megszerzése/kinevelése, majd értékesítése.

Egy példa, amely mutatja ennek fontosságát:

A pénzügyi gondokkal küzdő angol Birmingham City FC 2020 júliusában a fiatalkorában szerződtetett és saját akadémiájukon nevelkedett, 17 éves játékosát, Jude Bellingham-et, a német Borussia Dortmund-hoz engedte el 25 Millió angol font ellenében. Az átigazolásért kapott összeg megmentette a csapatot a csődtől és a sportágban szokatlan módon visszavonultatták a játékos mezszámát hálájuk jeléül.[7]

## 2.2 Adatalapú döntéstámogatás megjelenése a futballban

Az utóbbi időben a legjobb sportcsapatok egyre több pénzt fektetnek az adatelemzésbe, ennek megfelelően egyre jobban fejlődnek sport elemzési technikák.

Egy nagy hiány ezen a területen jelenleg, egy olyan platform vagy alkalmazás megléte, ahol a sportszakemberek együtt alkalmazhatják az elemzési technikákat és a fejlett vizualizációkat. [5]

A mesterséges intelligencia bevonása egy újabb mérföldkövet jelent. Az ideális felhasználási módja főként az analitikai hatékonyság javítása és a nagyszámú lehetőségek szűrése. Ezenkívül az adatok puszta elemzése alapján olyan meglátások felé terelhetik a szakértőket, amelyeket lehet, hogy figyelmen kívül hagytak volna. Egy rugalmas és könnyen konfigurálható eszközt kell a sportszakemberek kezébe adni, amely a mesterséges intelligencia és az adatelemzés erejét felhasználva adatalapú döntések meghozásában segíti őket. [5]

Ezzel a megközelítéssel kockázatok is járnak, különösen a nem megfelelő analitikai módszerek lehetséges alkalmazása vagy a torz adatok használata. Ezek mérsékelhetők képzett elemző szakemberekkel bevonásával, vagy hosszú távon a legjobb módszer kitapasztalásával. [5]

A felhasználóknak képesnek kell lenniük arra, hogy a folyamat részeként más szakértőkkel felülvizsgálják vagy megerősítsék a következtetéseket. [5]

Sok futballklub felhasználja az adatokat a döntéshozatali folyamatai során, de ennek módja és mértéke klubonként nagyon eltérő. A legtöbb klub azonban inkább titokban tartja, hogy mit csinál az adatokkal kapcsolatban.

Az elitklubok előnyben részesítik a személyre szabott eszközöket, amelyek bizonyos területeket vizsgálnak, a kisebb kluboknál pedig hajlamosak olyan kész szoftvert választani, amely videofelvételeken alapul. [9]

### 2.2.1. Adatalapú játékosértékelés

A focisták rangsorolása és a teljesítményük adatalapú értékelése egyre központibb szerepet kap a futball világában. A játékosokat mérőszámok alapján pontozzák és ez alapján összehasonlíthatóvá válik a teljesítményük, rangsorolhatóvá válnak. Meg kell említeni, hogy ezeket a számokat időközönként frissíteni kell, különben gyorsan elavulttá válnak.

Egy ilyen pontozási rendszerben a játékosok teljesítményét el lehet helyezni a hasonló szerepkörben játszó játékosokhoz képest. Ennek több lehetséges felhasználási területe is van:

* A sporttal foglalkozó televíziós csatornák, weboldalak, videojátékok gyártói használják ezeket az adatokat a játékos teljesítmények összehasonlítására, a szurkolók magasabb minőségű szórakoztatása céljából.
* Az edzők és a csapat menedzsment is érdekeltek az adat-vezérelt eszközökben, mivel ezek képesek támogatni a taktikai elemzéseket és követhetővé válik a játékosok pályán nyújtott teljesítménye.
* A játékosmegfigyelők és teljesítmény elemzők folyton keresik a lehető legjobb adat-vezérelt eszközöket, amelyekkel növelhetik a keresett képességekkel rendelkező tehetséges játékosok felfedezésének hatékonyságát. [10]

A játékosmegfigyelők munkáját nagyon nagy mértékben könnyíti meg, ha számadatok alapján tudják összehasonlítani a játékosokat, akár visszamenőleg is. Ez alapján szűkíthetik a megfigyelt játékosok listáját, ezzel erőforrást spórolva. Az adatok másik fontos felhasználási módja, hogy alátámaszthatják a megérzéseiket vagy felülvizsgálatra késztethetik őket.

A szakdolgozatom keretein belül ezeket a számadatokat és az ez alapján elérhetővé váló összehasonlítás lehetőségét szeretném felhasználni, egy olyan rendszert létrehozva, ahol a játékosmegfigyelők képesek szűrni az általuk fontosnak tartott paraméterek alapján, illetve a gépi tanulás segítségével profilra szóló játékosajánlásokat is kaphatnak.

### 2.2.2. Előnye a hagyományos módszerekkel szemben

Ahhoz, hogy jobban megjósolhassuk, hogy egy játékos valóban rendelkezik-e azokkal a készségekkel, amelyekre a csapatnak szüksége van, az adatelemzés nélkülözhetetlen. A következő három állítás alátámasztja ezt:

* A „hagyományos”, nem adat alapú játékosmegfigyelésben sok a torzítás és az elfogultság. Egy példa erre: Ha egy játékosmegfigyelő, aki éppen a csapat új szabadrúgás rúgóját keresi kilátogatott volna Cristiano Ronaldo első bajnokok ligája meccsére a Real Madrid-ban, akkor minden bizonnyal őt tartaná a lehető legjobb választásnak, mivel két gólt is szerzett szabadrúgásból azon a meccsen. A számok ezzel szemben azt mutatják, hogy „átlagos” arányban képes értékesíteni a szabadrúgásokat, de erről az őt látó játékosmegfigyelőt már nehezen lehetne meggyőzni. [3]
* Az adatelemzés segítségével növelhető a játékosmegfigyelők hatékonysága. Erre jó példa az, ha a szakemberek minden tehetséges játékost meg szeretnének vizsgálni, az közel lehetetlen lenne pusztán azok száma miatt. Az adat-vezérelt technikák segítenek leszűkíteni a halmazt a megadott paraméterek szerint (Például: életkor, poszt). Már csak ez a művelet is sok időt és anyagi forrást szabadít fel, amit a még hatékonyabb elemzésekre tudnak fordítani a szakemberek. [10]
* Az adatelemzés nem növeli a klub költségeit jelentős módon, sőt csökkentheti azokat. A mai, magas átigazolási díjak és játékos bérek miatt a rossz igazolások anyagi kockázatot is jelentenek. A nem az elvárásoknak megfelelően teljesítő játékos posztja továbbra is hiány lesz, a bére terheli a csapat erőforrásait. Lehet, hogy tehetséges játékosról van szó, viszont, ha egy nem megfelelő csapatban hosszú távon rosszul teljesít, az a piaci árának az esését is jelenti, ami már pénzben kifejezhető veszteség a csapat számára.

Az játékosról gyűjtött adatok kulcsfontosságúak abban is, hogy a játékosmegfigyelő, aki kiválasztotta a megfelelő játékost, alá tudja támasztani az érvelését a döntéshozók számára. Ez azt jelenti, hogy a játékos vásárlásában érdekelt tulajdonosoknak, igazgatóknak is betekintést kell kapniuk a felderítés eredményeibe és ez elképzelhetetlen a megfelelő adatok nélkül. [11]

### 2.2.3. A játékosok piaci értékének megbecsülése adatok alapján

A menedzserek egyik legfontosabb döntései az átigazolásokhoz kapcsolódnak. A játékosok szerződtetésekor, eladásakor az átigazolási díjak és a piaci értékek meghatározása mindig nehéz feladat.

A piaci értékek összességében az átigazolási díjak becsléseiként - ez az a díj, amit a játékos szerződtetésekor meg kell fizetni az ő játékjogát korábban tulajdonló csapatnak - foghatók fel, így ezek fontos szerepet játszanak a tárgyalásokban.

Korábban ezeket az értékeket futball szakértők becsülték meg, de a közönség egyre inkább beleszól az árak alakulásába. Nem elhanyagolható szempont a játékos reklámértéke, nemzetisége, vallása, ezek a tényezők új szurkolókat vonzhatnak be a klubhoz. Az új szurkolók egyenesen arányosak az új nézőkkel, az új vásárlókkal, ami a beáramló tőke növekedését jelenti.

De a szakértők és tömegek általi értékelések vegyítése nem átlátható folyamat, ritkán frissíthető és nem reprodukálható. Az adatelemzés tehát itt is egy megalapozott alternatívát vagy kiegészítő megközelítést jelenthet.

Az eredmények azt sugallják, hogy a piaci érték adat-vezérelt becslése képes felülkerekedni a tömegek számos gyakorlati korlátján, miközben összehasonlíthatóan pontos számokat állítanak elő. Eredményeink fontos hatással vannak a futballmenedzserekre és a játékosmegfigyelőkre, mivel az adatelemzés lehetővé teszi a piaci érték pontos, objektív és megbízható becslését, amely bármikor frissíthető. [12]

## 2.3. Adatok forrásai

A játékosmegfigyeléshez nagy mennyiségű strukturálatlan adat áll rendelkezésre: videók az egyes játékosok teljesítményéről, írott riportok, különböző statisztikák elemző weboldalakon. Ezek közül a videók figyelése mára a felderítés utolsó szakaszába tolódott, ahol már az adatok igazolták a játékosban rejlő lehetőségeket és a szakember megnézi a saját szemével is. A statisztikákat és írott riportokat pedig különböző módszerek szerint már számítógépes programok elemzik.

Érdekesség, hogy a mérkőzések statisztikáinak készítésekor sokszor emberi erővel jegyzik fel a történéseket. Ez a feladat még nincs teljesen automatizálva, ám már dolgoznak a megfelelő videó elemző szoftvereken, amelyek megbízható pontossággal rögzítik az eseményeket.

A statisztikák elemzése egy adat-vezérelt teljesítményértékelési mutatók kiszámítását jelenti, amely számszerűsíti a játékos teljesítményének minőségét egy adott mérkőzésen. Ez egy összetett feladat, mivel a teljesítmény minőségének nincs objektív és közös definíciója. A szakirodalomban eddig számos adat-vezérelt rangsoroló és kiértékelő algoritmussal találkozhatunk. [10]

A projektem során ilyen számszerűsített teljesítményekhez hasonló adatokkal fogok dolgozni. A Football Manager nevű videójáték korábbi adatait fogom felhasználni.

### 2.3.1. Football Manager

A Football Manager egy foci menedzser szimulációs játék, ahol a játékosok közel 60 szellemi és fizikai képessége számszerűsítve szerepel, amit több mint 1300 játékosmegfigyelő munkája segít. Ennyi ember sokkal átfogóbb információt tud összegyűjteni a játékosokról, mint amilyet képes lenne bármelyik klub egyedül kezelni. Figyelik a játékosokat edzéseken és mérkőzéseken, összehasonlítják és pontozzák őket. Ha egy játékos kiemelkedik a többiek közül magasabb pontszámot kap, amivel felfigyelhet rá egy nagyobb klub. Visszajelzések alapján folyamatosan javítják a pontozási hibákat a játék aktuális kiadása előtt, ezért az adatbázis valós képet ad a játékosok tudásáról. [13]

Míg a legjobb ligákban játszó csapatoknak megvan az a lehetősége, hogy jelentős felderítési költségvetéssel dolgozzanak, a Football Manager értékes erőforrás lehet a kisebb klubok számára. [13]

A futballfelderítők megfigyeléseket végeznek a pályán, aztán elküldik ezeket a jelentéseket a vezetőségnek és ők ellenőrzik őket. Vagy lekérnek adatokat egy adatbázisból, ami visszaad nekik játékosokat, akik megfelelnek bizonyos kritériumoknak. Aztán felkérik a felderítőket ezeknek a játékosoknak a megfigyelésére.

Mindkét módszer működik, felderítés legjobb módja keverni őket. [13]

A felderítés első szakasza nagy merítéssel történik, ahol sok potenciális név kerül a kalapba. Ezen a ponton nem igazán fontos, hogy honnan származik az információ, érkezhet ügynököktől is, vagy egy számítógépes játék adatbázisából. Ezután az érdekes játékosokat online elemzik. Akik átmennek ezen a szűrőn, azért kiküldenek egy játékosmegfigyelőt, hogy megnézze élőben is. [13]

### 2.3.2. Adatok elérése

A Football Manager nevű játék 2017-es kiadásának játékos-pontszámokat tartalmazó adathalmaza elérhető a Kaggle.com oldalon. Ez fogja képezni a rendszerem alapját játékosok teljesítményére vonatkozó pontszámokat. (<https://www.kaggle.com/ajinkyablaze/football-manager-data>), Utoljára megtekintve: 2021.12.08.

A fenti adathalmaz frissített verziója (2020-as) szintén a Kaggle.com oldalon érhető el. Felhasználom a benne szereplő személyes adatokat a játékosokról, hogy színesebb képet lehessen kapni róluk, illetve több szempont szerint lehessen szűrni őket.   
(<https://www.kaggle.com/datasets/ktyptorio/football-manager-2020>). Utoljára megtekintve: 2022.03.10.

Ez a két *.csv* fájl, egy táblává alakítva fogja adni a rendszerem adatbázisát. Mivel mindkettő relációs adatokat tartalmaz, ezért egy relációs adatbáziskezelő rendszert fogok választani a tárolásához.

# A SZAKDOLGOZAT CÉLJA

## 3.1. Célkitűzés

Ahogy az ismertetett szakirodalomból kiderül: az, hogy a játékosok mennyire képesek jól együtt játszani, meghatározza a csapat sikerességét. Másképp mondva, a csapat sikeressége azon múlik, hogy egymáshoz és az edző taktikájához mennyire jól illő játékosokat tudnak megszerezni.

Ez azt jelenti, hogy a jó igazolások és a rossz igazolások arányát a lehető legmagasabbra kell emelni. Az, hogy egy igazolás jó-e vagy rossz-e, bevált vagy nem vált be, nem lehet egyértelműen kijelenteni, ha nincs számokkal alátámasztva.

Nem könnyű feladat kiválasztani, hogy egy csapat melyik játékost szeretné szerződtetni, bármilyen posztról legyen is szó. Ha van egy nagyszerű játékos, akinek a világon mindenki tisztában van a kiemelkedő képességeivel, valószínűleg több csapat is szeretné megszerezni, ami extrém magas átigazolási díjat is generálhat.

A játékosmegfigyelők célja az, hogy még a többiek előtt megtalálják a nagy tehetségeket, illetve az ígéretes fiatal játékosokat, ezzel elérve, hogy a piaci értékénél olcsóbban tudják szerződtetni. Az utóbbi időben kezdett elterjedni, hogy ezek a szakemberek adatelemzőkkel együttműködve nagy adathalmazokból kinyert értékes információkkal tudják javítani a munkájuk hatékonyságát. Elképesztő eredményeket lehet elérni, pusztán azzal, hogy az adatelemzést figyelembe veszik.

Ezzel szemben nem teljeskörű az adatalapú játékosmegfigyelés térhódítása. Azok a csapatok, akik nem tudnak megfelelő mennyiségű erőforrást allokálni a játékosmegfigyelők döntéseinek adatalapú támogatására, nem tudják kihasználni ennek az igazi előnyeit. Úgy gondolom, hogy amíg el nem jut oda minden csapat, hogy kellő hatékonysággal tudják használni az interneten fellelhető strukturált és strukturálatlan adatokat, addig ezek a csapatok hátrányban lesznek a többiekhez képest.

A szakdolgozatom keretein belül az ő munkájuk segítésére szeretnék létrehozni egy felületet, amit ingyenesen használhatnak a megfelelő játékosok keresésekor, illetve ajánlásokat is kaphatnak a rendszertől.

Egy olyan adatbázist fogok használni ehhez, ahol a játékosok képességei diszkrét pontszámokkal vannak jellemezve, ezzel elérve, hogy egymáshoz hasonlíthatók lesznek különböző szempontok alapján. A szakemberek számára számok formájában is látszódni fog, hogy melyik játékos miben erős, illetve gyenge.

## 3.2. Elérendő állapot

Szeretném elérni a szakdolgozatom késznek tekintéséhez, hogy

* feldolgozzam és bemutassam a rendelkezésre álló forrás adatokat,
* eltároljam őket egy adatbázisban,
* a lekérdezések optimalizálása érdekében megfelelő indexeket hozzak létre,
* létrehozzak egy webes felületet, ahol a felhasználók elérhetik ezeket az adatokat,
* a felületen lehetőségük legyen a paraméterek szerint szűrni, rendezni a játékosokat,
* gépi tanulás segítségével lehetőséget kapjanak játékos ajánlásokat kapni.

## 3.3. Célként kitűzött felhasználási lehetőségek

A felhasználók képesek lesznek az adatbázisban található összes játékos pontszámait megtekinteni, és ezeket összehasonlítani más játékosok pontszámaival, ezáltal egyfajta rangsort alkotva közöttük.

A további eltárolt adatok arra is alkalmasak, hogy rendezzük, illetve szűrjük a játékosokat a piaci áruk, nemzetiségük, magasságuk vagy bármely más tulajdonságuk alapján.

Ezek a lehetőségek nagy mértékben le fogják csökkenteni azoknak a játékosoknak a számát, akiket továbbra is érdemes lehet megfigyelni. Ha a szakemberek fel tudnak állítani olyan szempontokat, amelyek alapján kizárhatók a játékosok a további megfigyelésből, akkor ezeket alkalmazni tudják majd itt. Ez egyértelműen sok megtakarított időt és másra fordítható energiát jelent a számukra.

Ezek mellett lesz lehetőség arra, hogy a rendszer, beállítható paraméterek alapján, képes legyen játékosokat ajánlani, azok profilja alapján. Ezzel a módszerrel olyan játékosokkal találkozhatnak a felhasználók, akiket nem biztos, hogy figyelembe vettek volna a keresés során.

# AZ IRODALOM ALAPJÁN A LEHETSÉGES MEGKÖZELÍTÉSI MÓDOK ÁTTEKINTÉSE ÉS KIVÁLASZTÁSA

## Megközelítési módok

Többféle megközelítés jöhet szóba:

* Statisztikák felhasználása, elemzése

Sok csapat használja ezt a módszert az adatelemző csapatuk segítségével. Biztos vagyok benne, hogy ezzel lehet a legnagyobb eredményeket elérni jelenleg, de sajnos nekem nem áll rendelkezésemre az a sportszakmai tudás, amivel megfelelő szempontok alapján tudnám elemezni a statisztikákat.

* Videós adatok, írott jegyzetek elemzése

A videóanyagok képfeldolgozás segítségével történő elemzése még a jövőbe mutató irányzat, de mindenképpen nagyon hatásosnak gondolom. A folyószöveg feldolgozására az adott nyelvet értő mesterséges intelligenciák képesek, egy ilyen elkészítése túlmutatna a szakdolgozatom keretein.

* Kiszámított számértékek használata

Ez a legkevésbé komplex megoldás és ha megfelelő minőségűek az adatok, akkor ezzel a módszerrel is jelentős eredményeket lehet elérni.

A három közül az utolsó megoldást fogom választani, mivel ehhez van meg a megfelelő tudásom és ilyen típusú rendszert nem találtam még.

A megoldás lényege, hogy a játékosokat az kiszámított számértékek alapján lehet rangsorolni különböző szempontok szerint. Ha nem is pontos statisztikákról van szó, arra alkalmas a megoldás, hogy megmutassa a játékosok közti különbségeket és alapot biztosítson a szűréseknek.

## Piacon elérhető hasonló szolgáltatások

* **WyScout**: Egy olasz cég, amelynek a fő profilja a futball játékosok felderítésének, a meccselemzésnek és az átigazolási dinamikák megértésének segítése. Videóelemző eszközöket és digitális adatbázisokat biztosítanak futball edzőknek, csapatoknak és játékosoknak. A céljuk, hogy a felhasználó részletes képet kapjon nagy számú játékos teljesítményéről a játékmintákról és a taktikákról.

Leginkább videóelemzések tekintetében kimagaslóak, nagyon részletes videó-adatbázisuk van a meccsekről, amely nagy segítséget jelent az elemzések során.

Ez egy fizetős alkalmazás, a legolcsóbb elérhető csomag 250 euró évente, ez havonta 50 percnyi videó elérését biztosítja.

* **InStat**: Egy orosz cég, amelynek a profiljába több sportág is tartozik. Az InStat Scout egy webes platform csapatok, játékosok és játékvezetők teljesítményelemzésére. Hozzáférést biztosít az egyéni statisztikák széles skálájához.

A platform innovációja, hogy minden statisztika videóepizódokhoz kapcsolódik. Ez azt jelenti, ha a felhasználó kiválasztja egy játékos utolsó öt gólját, akkor egyből nézheti is a róluk készült képanyagot. A statisztikák nagy adatbázisának és a videóknak köszönhetően, játékosfelderítési lehetőséget is biztosít.

Az InStat Scout költsége 2000 euró évente.

* **Transfermarkt**: Egy német cég által készített weboldal, amely egy aktuális általános futball tudásbázis. Nagyon sok információ megtalálható rajta a gólokkal, eredményekkel, statisztikákkal és játékosokkal kapcsolatban. Nagy hangsúlyt fektetnek az átigazolásokra és az átigazolási hírekre is. Kutatások alapján az ott megjelenő átigazolási „pletykák” nagy arányban helytállónak bizonyulnak.

A játékosok piaci árát havonta frissítik különböző mutatók alapján és többnyire meg is egyezik a tényleges értékükkel.

A Transfermarkt ingyenesen használható.

## Eltérés az eddigi megoldásokhoz képest

Az eddigi tapasztalataim alapján a hasonló feladatot betöltő weboldalak és alkalmazások mind szoftver által feldolgozott statisztikákon alapulnak, ez alapján teszik lehetővé a játékosok teljesítményének megismerését. A videók és statisztikák elérhetősége is fontos eleme a piacon jelenlévő megoldásoknak.

Az én rendszerem ezzel szemben más megközelítést alkalmaz és más a betölteni kívánt szerepe is. Én szakértők és a közösség által meghatározott pontszámot fogok alkalmazni, amely nem felel meg a statisztákat felhasználó megoldások szerepének, hanem csak a megfelelő potenciállal rendelkező játékosokat lehet vele azonosítani. Ezek után érdemes a kiválasztott játékos statisztikáit is megvizsgálni, de arra vannak elérhető eszközök.

Ezenkívül a gépi tanulás alapú játékos ajánlást tekintem a megoldásom nagy erősségének, amire nem láttam még eddig példát, viszont fontos szerepe lehet a megfelelő profilú játékosok azonosításában.

# A RÉSZLETES SPECIFIKÁCIÓ LEÍRÁSA

## 5.1. Vázlatos rendszerterv

Az 5.1 ábrán látható a rendszer vázlatos terve.



* 1. ábra – Vázlatos rendszerterv

A felhasználók a saját böngészőjükből érhetik el a webalkalmazást, ami kiszolgálja a kéréseket. A hozzákapcsolat adatbázis szerverről kéri le az adatokat, amelyeket továbbít a kliensek felé. Ha a felhasználó szűrő értékeket állít be, vagy rendezni akarja a játékosokat, akkor ezek az adatbázis lekérdezésekben fognak érvényesülni és webalkalmazásnak csak továbbítania kell az adatokat.

A gépi tanuló komponens egy vagy több előre létrehozott modellt jelent, amely a belső számítások elvégzése után játékosajánlásokat képes végezni. Csatlakozhat az adatbázishoz közvetlenül, vagy a webalkalmazáson keresztül is lekérheti az adatokat. Ezek közül tesztelések útján tudom eldönteni, melyik a hatékonyabb. Ha a felhasználó játékos ajánlást szeretne kérni megadott szempontok szerint, akkor a webalkalmazás a gépi tanuló komponenssel együttműködve néhány játékost fog visszaadni eredményként.



* 1. ábra - Adatmodell

Az 5.2 ábrán látható az adatmodell.

3 táblát fog tartalmazni az adatbázis:

* Score:
  + A játékosok pontszámait tartalmazza. Ez alapján lehet őket rangsorolni a felhasználó vagy a Gépi Tanuló Komponens által. Az összes mező nem fért volna rá az ábrára, ezeket az 1. Melléklet-ben tüntettem fel.
* Player:
  + A játékosok személyes adatait tartalmazó tábla. Ezek szerint is lehet szűrni, de főként informáló funkciójuk van.
* Club:
  + A Csapatok találhatók meg ebben a táblában. A nevük, az ország és a liga, amelyben szerepelnek

A táblák elsődleges kulcsára indexeket is definiáltam a gyorsabb műveletek érdekében.

## 5.2. Vázlatos időterv

|  |  |  |
| --- | --- | --- |
|  | **Munkafázisok** | Tervezett befejezés  (szorgalmi időszak) |
| 1. | Rendszerterv pontosítása, szükséges eszközök beszerzése, telepítése | Első hét |
| 2. | Adatbázis környezet kialakítása, forrásfájlok importálása | Második hét |
| 3. | Próba lekérdezések futtatása, optimalizálás, tesztelés | Harmadik hét |
| 4. | Webalkalmazás konfigurálása, tesztelés | Negyedik hét |
| 5. | Tervezett webes felület kialakítása | Hatodik hét |
| 6. | Eddigi komponensek együttműködésének tesztelése | Nyolcadik hét |
| 7. | Gépi tanuló komponens létrehozása, integrálása | Tizedik hét |
| 8. | Rendszer tesztelése, hibák javítása, dokumentáció befejezése | Tizenharmadik hét |
|  |  |  |
|  | A munkafázisok alatt folyamatos dokumentálás\* |  |

## 5.3. Adatbázis

Ahogy a vázlatos rendszertervben is látható csak egy tábla lesz az adatbázisban. Ez a tábla tartalmazni fogja a játékosok adatait és pontszámait. Egy olyan relációs tábla lesz, ami nagyon sok (több mint 50) oszloppal rendelkezik. Mivel az adatbázist csak az adatok elemzésére és megjelenítésre fogom használni, ezért csak olvasni fogok az adatbázisból, írni nem.

### 5.3.1. Lehetőségek áttekintése

Ezek alapján egy relációs adatbáziskezelőt kell választanom, az alábbi lehetőségeket gondoltam végig:

* **MSSQL**
  + Előnyei:
    - Sok fajta kiadás, különböző árazással, különböző célokra.
    - Teljeskörű üzleti megoldások. Számos üzleti kiegészítőt tartalmaz, mint például az ETL megoldások vagy adat tisztító folyamatok. Ezenkívül biztosít eszközöket teljeskörű adminisztrációhoz, online analitikák számításához, adatbányászathoz és adatvizualizációhoz.
    - Gazdag dokumentáció áll rendelkezésre.
    - Felhőalapú adatbázis támogatás. Az elérhető megoldások: Microsoft Cloud, Azure SQL Database, SQL Server on Azure Virtual Machines. [2]
  + Hátrányai:
    - Az ingyenes verziója, az Express, csak belépő szintű eszközöket kínál. Ha komolyabb projektet kell támogatni vele, akkor másik kiadást kell választanunk, ami magas költségekkel jár.
    - Gyakran változó és nem tiszta licenszfeltételek.
    - Aki nem régóta foglalkozik ezzel a környezettel, nehézkes lehet számára a lekérdezés optimalizálás és a hatékonyság növelése. [2]
* **Oracle**
  + Előnyei:
    - Folyamatos innováció a napi munkafolyamatok megkönnyítésére.
    - Erős technikai támogatás és dokumentáció, közösségi támogatás. Ha valami probléma adódik, majdnem biztosan lehet rá találni megoldást.
    - Az Oracle multimodell megoldása hatalmas mennyiségű adat befogadását és feldolgozását teszi lehetővé.
    - Erős szinkron adatfeldolgozás. [2]
  + Hátrányai:
    - Az ingyenes kiadások funkciói korlátozottak. Az ennél magasabb szintű kiadásaik pedig nagyon magas költségekkel járnak.
    - Erőforrás-igényes technológia. Sok lemezterületet igényel, hatékony infrastruktúrára van szüksége.
    - Megfelelő szakértelem kell a kezeléséhez. [2]
* **PostgreSQL**
  + Előnyei:
    - Nagy vertikális skálázhatóság.
    - Egyéni adattípusok támogatása. Alapértelmezés szerint nagyszámú adattípust támogat, de lehetővé teszi a felhasználók számára, hogy saját adattípusokat hozzanak létre. Ez a NoSQL megközelítés támogatását jelenti.
    - Könnyen integrálható külső eszközök.
    - Nyílt forráskódú a projekt.
    - Erős fejlesztői közösség, ahol segítséget nyújtanak a problémák megoldásában.
    - A bonyolult lekérdezések és az előre definiált funkciókkal megvalósított egyedi interfészek széles választéka miatt a PostgreSQL tökéletesen illeszkedik az adatelemzéshez és az adattárházakhoz.
    - Jelentősen felgyorsítja a hatalmas mennyiségű adat feldolgozását. [2]
  + Hátrányai:
    - Inkonzisztens dokumentáció. Mivel a PostgreSQL közösség meglehetősen elosztott, ezért nem mindenhol egységes a dokumentáció.
    - Nincsenek az adatbázis állapotára vonatkozó ellenőrzési eszközök. [2]
* **MySQL**
  + Előnyei:
    - Ingyenesen használható a személyi használatra, de vannak más lehetőségei is vállalati célokra.
    - Egyszerű a felépítése és a kezelése.
    - Támogatja a legnagyobb felhőszolgáltatókat, olyan platformokon érhető el, mint az Amazon Web Services, Microsoft Azure. Ez teret adhat a növekedésnek. [2]
  + Hátrányai:
    - Sok mérnöki megoldást kíván a skálázhatóság. Ha előre látható, hogy az adatbázis jelentősen fog növekedni, nem ez a megfelelő adatbázis megoldás.
    - Mivel az Oracle tulajdonában áll, nem teljesen nyílt forráskódú, ami gátolja a közösségi fejlesztéseket. Ha bármilyen módosítást szeretnénk a környezetben, az komoly költségekkel jár az Oracle felé.
    - Nem követi teljesen az SQL szabványokat, nem tartalmaz néhány szabványos szolgáltatást. [2]

### 5.3.2. Választás

A fenti megoldások közül a PostgreSQL-t találtam a legalkalmasabbnak, szerintem az illik legjobban a projekthez.

* A skálázhatóság fontos, mivel folyamatosan fejlődik ez a terület és egyre növekednek a játékos adatait nyilvántartó adatbankok. Elképzelhető, hogy a jövőben nagyságrendekkel több játékos adatai is rögzítve lesznek.
* Szeretném kihasználni az egyéni adattípus lehetőségét és számomra kézenfekvő adatstruktúrákat létrehozni.
* A bonyolult lekérdezések és az adatelemzés támogatása nagyon fontos lesz a munka folyamán.
* A munkám során szeretnék törekedni a nyílt forráskódú megoldások használatára.

## 5.4. Webalkalmazás

A Webalkalmazás feladatai:

* Meg kell valósítania egy egyszerű kezelőfelületet a felhasználók számára.
* Kezelni a Klienstől bejövő kéréseket és SQL lekérdezések formájában továbbítani az Adatbázis felé. Az onnan visszajött adatokat, vissza kell küldenie a Kliensnek.
* Szükség esetén kommunikálnia kell a Gépi Tanulás komponenssel, ami adatot fog neki szolgáltatni. Ezeket fel kell dolgoznia és továbbítani a kliensnek.

### 5.4.1 Lehetőségek áttekintése

Az alábbi elterjedt szerveroldali keretrendszerek használatát gondoltam végig:

* **C# ASP.NET**
  + Előnyei:
    - MVC architektúrát követ, ami lehetővé teszi az elválasztását az adatok beolvasásának, feldolgozásának és továbbításának.
    - A Visual Studio (a keretrendszer fejlesztőkörnyezete) nagy segítséget jelent a kódolási idő csökkentésében. Segíti a kód áttekintését, ami a minőségét is javíthatja.
    - A fejlesztőkörnyezetnek köszönhetően nagyon gazdag eszköztárral rendelkezik, amik nagyban könnyítik a fejlesztési munkákat.
    - Könnyen kezelhető alkalmazást lehet vele létrehozni, mivel a konfigurációs állományok változtatásával a szerver újraindítása nélkül is tudjuk változtatni az alkalmazást. [1]
  + Hátrányai:
    - A karbantartásakor figyelni kell, hogy több webszerver-erőforrást igényel, mint a hasonló megoldások.
    - Nincs minden teljeskörűen dokumentálva.
    - Magas hozzáértés kell a használatához, illetve többféle technológiának az ismerete szükséges. [1]
* **Python Django Framework**
  + Előnyei:
    - ORM (Objektum-Relációs leképezés) támogatás, ami megkönnyíti az adatbázis kezelését.
    - Nagyon jó biztonsági megoldásokat biztosít, mint például megakadályozni a kód futtatást a megjelenített weblapon.
    - Nagyon sok kész megoldás van, amit hozzá lehet csatlakoztatni és nem a fejlesztőnek kell megírni. Néhány példa ezekre:

Autentikáció kezelése

Adminisztrátor kezelőfelület

Munkamenet kezelése

Külön Postgre csomag, plusz funkciókkal

* + - Nyílt forráskódú rendszer nagy aktív felhasználói közösséggel.
    - Megkönnyíti az alapból komplex, adatbázisra támaszkodó weboldalak írását. [15]
  + Hátrányai
    - Kis projektekhez, ahol nincsenek kihasználva az előnyei, feleslegesen nagy szoftver.
    - Ha túl sok adattáblát használunk, bonyolult lehet a kezelése, mert a keretrendszer által kezelt adatmodellek egyetlen fájlban tárolódnak.
    - Bonyolult problémákba lehet belefutni, ha eltérünk a keretrendszer által javasolt megoldásoktól. [15]
* **Java Spring Framework**
  + Előnyei:
    - Ez egy könnyű konténer megoldás, ami azt jelenti, hogy nem kell a fejlesztőnek osztályokat örököltetni és interfészeket implementálni a kódolás során. Nagy teljesítményű, könnyen skálázható alkalmazásokat tudunk vele létrehozni.
    - Előre megírt megoldásokat tartalmaz, amelyeket a fejlesztők egyértelműen megbízhatónak tartanak. Ezeket saját paraméterek szerint lehet konfigurálni.
    - A keretrendszer fejlesztői nagy hangsúlyt fektetnek a gyorsaságra.
    - Több biztonsági megoldást is tartalmaz, mint például, hogy figyeli a külső forrásból származó függőségeket. A Java Spring alkalmazásával teljesen biztonságos webalkalmazást tudunk létrehozni. [8]
  + Hátrányai:
    - Komplex a fejlesztése, nagy szakértelmet igényel. Aki még nem használta ezt a keretrendszert, először tapasztalatot kell benne szereznie, mielőtt ténylegesen tudná használni.
    - Nem sok útmutató és irányelv tartozik a fejlesztéséhez. Sokszor magukra vannak utalva a fejlesztők. [8]
* **Ruby on Rails**
  + Előnyei:
    - Könnyű megtanulni ennek a keretrendszernek az alapjait, de időbe fog telni amíg valaki megtanulja kihasználni a tényleges előnyeit.
    - Sok előre megírt eszköz található benne, amit használhatnak a fejlesztők. Ezek nagy segítséget adnak kódolási hibák megtalálásához, az optimalizáláshoz és a teszteléshez.
    - Nagy és aktív fejlesztői közösség, akik segítséget tudnak nyújtani a problémák megoldásakor.
    - A bevett szoftvertervezési minták követésével nagyon hatékony és jól működő megoldásokat lehet létrehozni. [14]
  + Hátrányai:
    - A rugalmasság hiánya. Ami az alapvető feladatokat és funkciókat illeti, a Ruby on Rails szinte kifogástalan. Mivel azonban sok az alapértelmezett, előre beállított objektum, nincs sok választási lehetőség egy probléma megoldása során. Tehát a keretrendszer választás előtt át kell gondolni az igényeket és hogy ennek megfelelnek-e a beépített modulok.
    - Folyamatosan fejlődik és alakul át a nyelv, ezért naprakész tudás szükséges hozzá.
    - Ha a fejlesztés során nem megfelelő megoldásokat alkalmazunk, az erősen kihatással lehet a teljesítményidőre, illetve problémákat okozhat a rendszer különböző elemei közötti kapcsolatban. [14]

### 5.4.2 Választás

A fenti megoldások közül a Python Django keretrendszer használatát gondolom a legalkalmasabbnak.

* Jó adatbázis feldolgozás támogatása van, ami fontos lesz a projektem során.
* A beépített modulok felhasználása mellett rugalmas alkalmazást tudok írni.
* A munkám során szeretnék törekedni a nyílt forráskódú megoldások használatára.
* Van már tapasztalatom a Python nyelv használatában és mivel a Gépi Tanuló komponenst is ezen a nyelven fogom elkészíteni, jól fog illeszkedni hozzá.

## 5.5. Gépi Tanulás Komponens

A Gépi Tanulás Komponens képes lesz az adatbázis elérése, onnan az adatok lekérésére. Ezek alapján gépi tanulási eljárások segítségével modelleket hoz létre, amelyek segítségével, a Webalkalmazás kérésére játékos ajánlásokat tud majd nyújtani.

Az egyik ilyen megoldás a Klaszterezés (osztályozás) lesz. Ennek a lényege, hogy a legjobb és legismertebb játékosokat felügyeletlen gépi tanulás segítségével osztályokba soroljuk. Ezeket az ismert játékosok képességeivel mindenki tisztában van. A felhasználó kiválaszthat egy ilyen játékost és az algoritmus megpróbál ugyanabba a csoportba tartozó, hasonló profilú játékosokat keresni (kevésbé ismerteket is), akiket érdemes lehet a csapatnak jobban megfigyelni.

### 5.5.1. Választás

Az adatelemzés és a gépi tanulás során két meghatározó programozási nyelv terjedt el, az R és a Python. Mivel eddig nem használtam még egyáltalán az R nyelvet, viszont a Pythonban már van tapasztalatom a gépi tanulás algoritmusok terén, ezért az utóbbit fogom használni.

## 5.6 Adatbetöltő script [[2]](#footnote-2)

Azért, hogy ne csak statikusan, az előre beégetett adatokon lehessen dolgozni és lekérdezéseket futtatni, a rendszer kiegészítésre kerül egy python adatbetöltő scripttel.

Ez a script SQL utasítások segítségével tölti be az adatokat az adatbázisba.

# A TERVEZÉS SORÁN VÉGZETT MUNKAFÁZISOK ÉS TAPASZTALATAIK LEÍRÁSA

## Adatbázis tervezése

Először fel kellett mérnem, hogy milyen igényeket szeretnék kiszolgálni és milyen lehetőségeim vannak:

* Az eredeti, általam talált adattábla a játékosok személyét tekintve csak a nevüket, életkorukat és születési dátumokat tartalmazta.
* Mivel szerettem volna ennél több információt megjeleníteni róluk, keresni kezdtem és megtaláltam ugyanennek az adatbázisnak egy frissebb változatát. Ez már több adatot tartalmazott.
* Az új adatokkal kiegészülve frissítettem az adatmodellt, így a korábbi 1 tábla helyett, már 3 táblából áll az adatbázis. (Eszerint frissítettem a dokumentációt a Vázlatos rendszerterv résznél)

Meg kellett vizsgálnom még az adattisztítás szükségességét. Ez alapján a következő két feladat várt rám:

* A rendelkezésre álló új adattáblában több játékosnak a neve (a különleges karaktereket tartalmazók) helytelen karakterekkel jelent meg, mivel a fájl rossz karakterkódolással lett mentve.
* Sok (~10.000) olyan játékos van, akik a korábbi adatbázisban szerepeltek, de a frissítettben nem. Ennek az oka, hogy ezek a játékosok felhagytak időközben a profi pályafutásukkal (a két adathalmaz aktualitása között 3 év van).

Az adatok szempontjából ez azt jelenti, hogy vannak olyan játékosok, akiknek vannak pontszámaik, de személyes adatok nincsenek róluk.

Mivel az alkalmazás a pontszámokra fókuszál, ezért benne hagytam őket az adathalmazban, de valamilyen karaktert kellett helyettesítenem az üres értékek helyére.

## Django webalkalmazás tervezése

A webalkalmazás tervezésénél összeszedtem az általam előre kitalált funkciókat, amelyeket teljesítenie kell. Próbáltam felhasználói fejjel megvizsgálni, hogy hogyan lehet egyszerűsíteni a munkafolyamatokat, illetve ergonomikusan kialakítani a felhasználói felületet.

Az előre meghatározott funkciók a következők:

* Az adatbázisban tárolt adatok megjelenítése
  + A játékosokat megjelenítése táblázatban
  + A pontszámokat megjelenítése táblázatban
  + A csapatokat megjelenítése táblázatban
  + Mindegyik megjelenítésnél az oszlopokat egyenként el lehessen rejteni jelölőnégyzetek segítségével
* Szűrők elhelyezése
  + A játékosokat és a csapatokat lehessen szűrni a tulajdonságaik alapján
  + A pontszámoknál egyenként megadható legyen egy érték, amelyeknél csak nagyobb pontszámokat listáz ki
* Pontszám szerinti rendezés
  + A játékosoknál meg lehessen határozni egyszerre legalább 5 olyan tulajdonságot, amely szerint rendezve lesznek (prioritás szerint, az első megadott érték lesz a rendezésben az elsődleges szempont, az utolsó érték lesz az utolsó szempont)
* Legyen lehetőség a Gépi Tanuló komponens használatára
  + Egy külön képernyőn lehessen felparaméterezni és az általa adott ajánlásokat megjeleníteni

## Gépi tanuló komponens tervezése

# 

Ez a rész még tervezés alatt van. Egyelőre kettő funkciót szeretnék biztosan megvalósítani:

* A felhasználó kiválaszt egy játékos, akire a profilja szerint szüksége lenne a csapatában. Gépi tanulás segítségével a program felméri a kiválasztott játékos képességeit, majd hasonló profilú játékosokat fog kilistázni neki.

(Ez akkor lehet hasznos, ha van egy kiszemelt játékos, de nem lehetséges a szerződtetése. Ilyenkor alternatívákat kaphat a felhasználó)

* A felhasználó kiválaszt egy pozíciót. A gépi tanulás segítségével a program felméri, hogy az adott pozícióra milyen tulajdonságok jellemzőek és eszerint a pozícióban játszó legjobb játékosokat kilistázza neki.

(Ez azért fontos, mert sok tulajdonság van a játékosokhoz rendelve. A felhasználónak nehéz feladat felmérni, hogy melyik pozícióban játszó játékosok miben erősek.)

## Adatbetöltő script tervezése

A scriptnek kezelnie kell az új adatok betöltését amellett, hogy a korábbi adatok ne vesszenek el és a későbbiekben a historizáció alapján riportokat lehessen rájuk ültetni.

Emiatt létrejön kettő új tábla:

* A játékosok korábbi személyes adatainak tartalmazó tábla
* A játékosok korábbi pontszámait tartalmazó tábla

Mindkét új tábla az eredetihez képest kiegészül egy *ValidTo* mezővel, amely jelzi, hogy meddig volt érvényes az adat. Így pontosan lehet látni, hogy hogyan változtak a játékosok képességei és személyes tulajdonságai (például magasság, testsúly) az idők során.

A scriptnek egy .csv kiterjesztésű fájlt kell megadni paraméterként, amiből be tudja importálni a sorokat az adatbázisba.

# MEGVALÓSÍTÁS LEÍRÁSA

## Adatbázis kialakítása

A forrásadatok két *.csv* kiterjesztésű fájlban álltak rendelkezésre. A két fájl (közel) ugyanannak az adathalmaznak a néhány évvel későbbi változata, ezért az összefésülésük nem okozott problémát.

Importáltam mindkét adathalmazt a PostgreSQL szerverre, ahol a mezők típusait már a bennük szereplő adatok szerint határoztam meg.

Terv szerint létrehoztam egy nagy ideiglenes táblát, amibe beletettem mindkét adathalmaz adatait (játékosnév szerint megfeleltetve egymásnak). Ezt az ideiglenes táblát fogom darabolni és a végső táblákká alakítani.

Ezt akadályozta, hogy az egyik fájl rossz karakterkódolással volt mentve, ezért a nevekben különleges karakterek helytelenül jelentek meg. Ez egyértelműen akadályozta, hogy a név szerint kapcsoljam össze a táblákat. Ez nagyságrendileg 10.000 adatpont vesztését jelentette volna.



7.1 ábra – Hibás kódolású karakterek a nevekben

Megoldás:

1. Exportáltam az új adatforrásból érkező neveket egy csv fájlba.
2. Írtam egy rövid python scriptet, ami beolvassa ezt az adathalmazt és kiválogatja belőle a hibás neveket.
3. Ahol hibás karaktert talált, ott a karakter helyére egy \_ karaktert helyettesített.

A képen szöveg látható

Automatikusan generált leírás

7.2 ábra – A hibás karakterek alulvonásra cserélve

1. Ezt az új adathalmazt visszatöltöttem a PostgreSQL szerverre.
2. Az régebbi (első) adathalmaz helyesen tartalmazza ezeket a neveket. A helyes neveket összevetettem a helytelen nevekkel és ahol csak egy karakter eltérés volt két név között, azt megtalálta.

(Az SQL-ben a string összehasonlításoknál a \_ karakter azt jelenti, hogy annak a helyén bármilyen karakter szerepelhet. Ezért ezt helyettesítettem be a helytelen karakterek helyére).

1. A hibás nevek helyére behelyettesítettem a javított változatokat.

Ezzel a megoldással az összes hibás nevet sikerült kijavítanom és utána el tudtam végezni velük a táblák összekapcsolását.

Voltak olyan játékosok, akik nem szerepeltek mindkét adatbázisban (valószínűleg ezek befejezték az aktív pályafutásukat), de mivel a pontszámaik megvoltak, úgy döntöttem, hogy megtartom őket. Ezeknél a tulajdonságokat egységesen ’Unknown’ felirattal töltöttem fel.

A kialakult ideglenes táblát már szét bírtam bontani a szükséges mezők szerint.

## Webalkalmazás kialakítása

### Környezet kialakítása

A webalkalmazást Django keretrendszerben írtam. A Django egy Python-alapú ingyenes és nyílt forráskódú keretrendszer, emiatt a környezet kialakításaként a Python legfrissebb verzióját telepítettem.

A következő Python csomagokat használtam fel a fejlesztés során, ezeket külön kell telepíteni: *pandas, sqlalchemy, sklearn, psycopg2*.

Ezekután telepítettem a Django keretrendszert. A Djangoval együtt automatikusan települ egy úgynevezett lightweight webszerver, amit a fejlesztés során lehet használni, de a tényleges publikáláshoz egy másik webszerver telepítésére van szükség. [17]

A Django alkalmazást a projektkönyvtárban a „python manage.py runserver” paranccsal tudjuk elindítani.

Ezután bármilyen böngészőben a *127.0.0.1:8000* címen érjük el az alkalmazást.

### Django projekt felépítése

A Django egy úgynevezett MVT (Model View Template) tervezési mintát használ. A minta három fontos rétegből tevődik össze:

* + Model:

Ez a réteg felelős az adatbázissal való kapcsolatért, az adatok megszerzéséért és az adatmodell reprezentálásáért a Django projekt számára.

* + View:

Ez a réteg valósítja meg az üzleti logikát, kezeli a Model által biztosított adatokat, feldolgozza a kéréseket és kommunikál a Template-tel.

* + Template:

Ez a réteg felelős teljes egészében a felhasználóval való kommunikációért és a felhasználói felület megjelenítéséért.



* 1. ábra – MVT működése [16]

A 7.3 ábra szemlélteti az MVT minta működését.

### 7.2.3 Tervezés

A megvalósítást azzal kezdtem, hogy leírtam a felhasználói igényeket, amiket ki kell szolgálnia a webalkalmazásnak. Ezek a következők:

* Játékosok megjelenítése
  + A játékosok személyes adatainak megjelenítése
  + A játékosok pontszámainak megjelenítése
  + A megjelenítendő mezők láthatóságát lehessen be-/kikapcsolni

Túl sok oszlop az összes személyes adat és pontszám ahhoz, hogy egyszerre lehessen szemlélni őket, ezért fontos, hogy a nem szükséges mezők láthatóságát ki lehessen kapcsolni.

* + Lehessen szűrni a játékosokat a személyes tulajdonságaik szerint
  + Lehessen sorba rendezni a játékosokat a személyes adataik és pontszámaik alapján, növekvő és csökkenő sorrendben is
* A csapatokat lehessen egy külön oldalon megjeleníteni
  + Legyen lehetőség a szűrésükre is
* Legyen lehetőség a Gépi Tanuló komponens funkcióinak igénybevételére
  + A gépi tanuló komponens által visszaadott értékek befolyásolják a megjelenített játékosokat

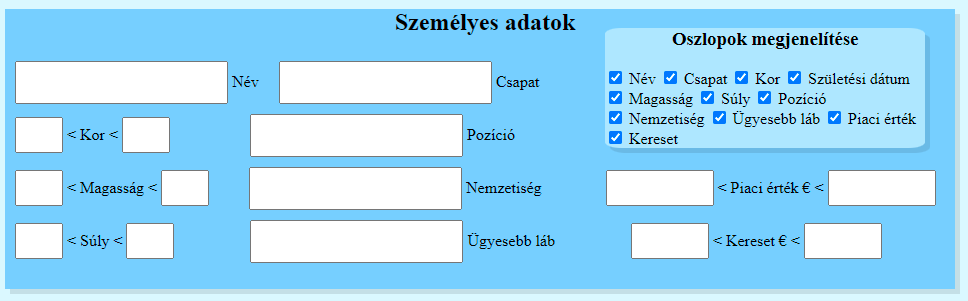
### 7.2.4 Felhasználói felület

#### Játékosok

A szükséges funkciók meghatározása után kialakítottam a felhasználói felületet, először a tényleges funkciók nélkül.



7.4 ábra – Teljes felhasználói felület

A 7.4 ábrán látható Személyes adatok panel lehetőséget biztosít arra, hogy a szűrő értékeket adjunk meg és szabályozzuk a játékosok tulajdonságainak megjelenítését. 

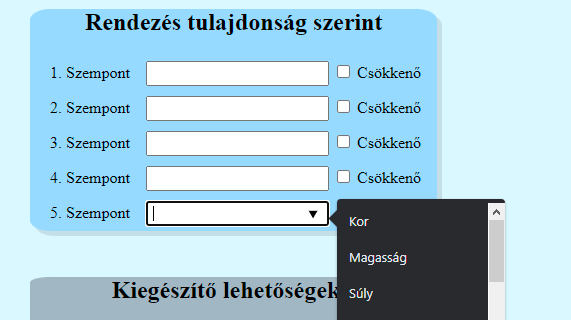
* 1. ábra – Személyes adatok panel

A 7.5 ábrán látható Pontszámok megjelenítése panel lehetőséget biztosít a játékospontszámok megjelenítésének szabályozására. Mivel túl sok tulajdonság van ahhoz, hogy mindegyik saját jelölőnégyzetet kapjon, csoportonként lehet állítani. A csoportban található tulajdonságok leírása az 1. Mellékletben található.



* 1. ábra – Pontszámok megjelenítése panel

A 7.6 ábrán látható Rendezés tulajdonság szerint panel lehetőséget biztosít a rendezési szempontok meghatározására úgy, hogy az 1. szempont lesz az elsődleges. Minden szemponthoz be lehet állítani, hogy csökkenő sorrendben legyenek rendezve az adatok az adott szempont szerint. Az alapértelmezett minden szempontnál a növekvő sorrend.  
A mezőbe kattintva egy lenyíló listából lehetőség nyílik a látható mezők kiválasztására.



7.6 ábra – Rendezés tulajdonság szerint panel

A 7.7 ábrán látható Kiegészítő lehetőségek panel lehetőséget biztosít a gépi tanuló komponens használatára.   
A *Hasonló játékosok keresése* gomb megnyomása után a felhasználó a kiválasztott játékos nevére kattintva indíthatja el a folyamatot. Néhány másodperccel később a gépi tanuló komponens által visszaadott játékosok jelennek meg a lenti táblázatban.   
Az Alulértékelt játékosok keresése gombra kattintva a gomb melletti mezőben megadott

A felhasználói felület kialakítása után megvalósítottam a Model rétegben az adatkapcsolatot és az adatmodellezést. Ez később kifejtésre kerül.

Az adatok rendelkezésre állása után megjelenítettem az adatokat is a felső panelek alatt.

Így, mikor az adatok és a funkció nélküli felhasználói elemek is rendelkezésre álltak, elkezdtem megvalósítani a funkciókat.

Mivel a szűrő és sorba rendező funkciók mind a megjelenítendő játékosok halmazát változtatják, az SQL lekérést kell módosítani a felhasználó cselekedeteinek megfelelően.

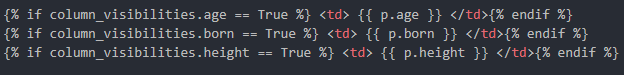
* A személyes adatok szűrése a lekérdezés WHERE része után szúr be felételeket

Ha a felhasználó helytelen karaktert, érvénytelen értéket ad meg, a lekérdezés nem fog találni egyező értéket, de hibát nem fog kapni.

* A megadott rendezési szempontok pedig a lekérdezés ORDER BY része után szúródnak be, annak megfelelően, hogy a *csökkenő* jelölőnégyzetbe van-e pipálva.

Ha a felhasználó helytelen karaktert, érvénytelen értéket ad meg, a lekérdezés figyelmen kívül fogja hagyni a megadott rendezési értéket.

A *pontszámok és oszlopok megjelenítése* panelek az SQL lekérdezést nem módosítják, hanem a Django template-ben figyelmen kívül hagyják az adott oszlop megjelenítését. Erre a 7.5 ábrán látható egy példa a kódból. Ha az *age, born* és *height* mezők láthatósága nem *True*, akkor egyszerűen nem fognak megjelenni. Ezek a láthatóságok a **View** rétegben lévő változókban tárolódnak.



7.5 ábra – Django példakód a láthatóságok megoldására

Mivel a pontszámok láthatóságát is kell állítani, de azokból túl sok van hozzá, hogy mindegyik külön gombot kapjon (több, mint 60 darab), ezért ezeknek a láthatóságát csoportonként kezelem. Az **1. Mellékletben** megtalálhatók a pontok leírása és csoportosítása, ennek megfelelően egy-egy csoporthoz rendeltem egy gombot.

A csoporthoz tartozó gomb a csoportban lévő összes mező láthatóságát azonosan befolyásolja.

#### Csapatok

A csapatok a játékos neve mellett a csapat nevére kattintva érhetők el. A *csapatok felületen* kezdetben csak a kiválasztott csapat található, de a szűrőből az értékét kitörölve megkapjuk az összes csapatot, a ligát, amiben játszanak és a csapat székhelyét. Ez egy vizsgált játékos esetében plusz információval szolgálhat. (Mennyire erős az a bajnokság, mennyire kicsi/nagy csapatnál hozza az adott teljesítményt stb.)

KÉP Beszúrása

### 7.2.5 Logika

A felhasználói interakciók feldolgozása és a megfelelő adatok megjelenítése mögött bonyolult logikák futnak le. Ezek a **View** rétegben hajtódnak végre, de kiszerveztem őket egy külön **Logic** programozási osztályba. Az osztály példányosítása után a megfelelő **View** metódusban már meg tudom hívni az üzleti logika függvényeit, amik így nem közvetlenül a **View**-ban vannak.

Ennek az átláthatóság miatt az is az előnye, hogy a később tárgyalt üzleti logikai tesztelés teljesen független az azt használó alkalmazástól.

### 7.2.6 Működés szerint:

1. **View** megkapja a Web-kérést, benne az összes szükséges paraméterrel és beállítással, amiket a felhasználói interakció generált.
2. A **View** átadja a paramétereket a tőle különálló üzleti logikának és visszakapja a logika által előállított SQL lekérdezést.
3. Az SQL lekérdezést a **Model** segítségével lefuttatja, megkapva így a megjelenítendő játékosok halmazát.
4. A megkapott játékosokat átadja a **Template**-nek, ami a beágyazott dinamikus kódokkal előállítja a weblapot, amit a **View** a Web-kérésre válaszolva visszaküld a felhasználónak.

## 7.3 Gépi tanuló komponens

A Gépi tanuló komponens funkciói a belső logikájukban különböznek, de azonos elven segítik a játékosok megjelenítését:

1. A felhasználó valamilyen interakciójának hatására létrejön az az adat, ami az algoritmusok alapjául fog szolgálni (játékos azonosító vagy pozíció, a kiválasztott funkciótól függően)
2. A **View** a Web-kérésből kiolvassa ezt a paramétert és meghívja a megfelelő külső Gépi Tanuló programot
3. A Gépi Tanuló program a belső logikája szerint műveleteket végez és visszaadja a megjelenítendő játékosok azonosítóját
4. Miután a **View** megkapta ezt a listát, beágyazza egy SQL lekérdezésbe és a **Model** réteg segítségével beolvassa az megjelenítendő játékosokat.
5. A megkapott játékosokat átadja a **Template**-nek, ami a beágyazott dinamikus kódokkal előállítja a weblapot, amit a **View** a Web-kérésre válaszolva visszaküld a felhasználónak.

Ezáltal a gépi tanuló komponens a folyamat üzleti logikai részének megvalósítását veszi át, a **View** réteg pedig egy központi irányítóként szerepel.

### Játékos ajánlás

A lényege, hogy a felhasználó kiválaszthat egy játékost, aki illene a csapatába, de valamilyen ok miatt nem lehetséges a megszerzése. A program keres olyan játékosokat, akik hasonló tulajdonságokban erősek, így alternatívát kínál az eredeti játékos helyett.

Az eljárás logikája:

1. Megkeressük az eredeti játékos 20 legjobb tulajdonságát.
2. A többi játékos közül olyanokat keresünk, akiknek a legjobb 20 tulajdonsága közül legalább 13 (tapasztalat által meghatározott érték) szerepel az eredeti játékos 20 legjobb tulajdonságai között.
3. A leszűrt játékosokból a tulajdonságaik alapján egy klaszterizáció kerül végrehajtásra, ami a játékosok halmazokba rendezését jelenti. A klaszterizációt K-Means algoritmussal valósítottam meg.
4. Az eredeti játékossal megegyező halmazban található játékosok az eredetivel hasonló stílusúak lesznek, őket adja vissza az algoritmus.

Mivel sok adatsor (játékos) van az adatbázisban nagyon sok tulajdonsággal, ezért a modellt kezelhetetlen ideig kellene futtatni. A futási idő javítása érdekében leszűrtem a szóba jöhető játékosok halmazát:

* + Meghatároztam az eredeti játékos 20 legjobb tulajdonságát.
  + Sok olyan játékos van, akit alapból ki lehet zárni, mert más poszton játszik vagy annyira gyengék ezek a képességei, hogy nem érdemes ajánlani őket. Emiatt leszűrtem a játékosokat úgy, hogy a figyelembe vett 20 tulajdonságban az 1-20 terjedelmű skálán legalább a 7 értéket érjék el.
  + Ezután végrehajtottam az eljárás logikájában leírt 2. lépést
  + Az így kapott néhányan száz nagyságrendű adatra már könnyen lehetett gépi tanuló algoritmust alkalmazni.

A közös értékek alsó határának meghatározására elvégeztem egy 100 véletlenszerű játékosra vonatkoztatott felmérést. A 7.7 ábrán látszik, hogy a megkövetelt közös tulajdonságok növelésével hogyan csökken a talált játékosok száma.

A 13-as elfogadási határnál még közel 250 játékos felel meg átlagosan. Ezek, a képességeik alapján, már hasonlítani fognak az eredeti játékoshoz, de a klaszterizáció tovább szűkíti a lehetőségeket.



7.7 ábra – Hasonló játékosok a közös tulajdonságok függvényében

#### Klaszterek számának meghatározása – Sziluett Módszer

A halmazokba rendezett pontok Sziluett értéke megmutatja, hogy a pontok mennyire illenek a saját halmazukba a többi halmazhoz képest. Ez egy -1 és +1 közötti valós szám. A +1-hez közelítő érték azt jelenti, hogy a pontok jól illenek a saját halmazukba.

Az ideális klaszterszám megállapításához létre kell hozni több K-Means modellt, különböző klaszterszámokkal (> 2). Az a klaszterszámú modell a legoptimálisabb, ahol a legnagyobb a Sziluett érték.

### Alulértékelt játékosok keresése

A lényege, hogy a felhasználó meghatároz egy pozíciót és a Gépi tanuló algoritmus visszaadja számára a 100 legalulértékeltebb játékost a kiválasztott pozícióban.

Az alulértékelt játékos azt jelenti, hogy a többi játékos piaci árát figyelembe véve az alulértékelt játékosnak a jelenleginél magasabb piaci árral kellene rendelkeznie. Minél nagyobb a különbség a becsült és tényleges piaci ára között (a becsült javára), annál alulértékeltebb.

Az eljárás logikája:

1. A játékosok közül ki kell szűrni azokat, akik a felhasználó által megadott pozícióban játszanak.
   1. Ha a megadott pozíció „GK” (, tehát kapusok keresése a feladat), akkor a játékosoknak csak a kapusokra vonatkozó tulajdonságait kell figyelembe venni.
   2. Ha a megadott pozíció nem „GK” (, tehát úgynevezett mezőnyjátéksok keresése a feladat), akkor csak kapusokra vonatkozó tulajdonságok kivételével az összes tulajdonságot figyelembe kell venni. [[3]](#footnote-3)
2. A szűrés után kapott játékosokra egy Lineáris Regressziós modellt kell építeni, ahol a magyarázó változók a játékosok tulajdonságai (az 1. pontban leírtak szerint), a célváltozó pedig a játékosok piaci ára.
3. A modellnek meg kell adni ugyanazokat a játékosokat, akik alapján készült a modell, így megkapható, hogy a Regresszió szerint mennyinek kellene lennie az egyes játékosok piaci árának a többi játékos ára szerint.
4. A száz legalulértékeltebb játékos megtalálásához a játékosokat sorba kell rendezni a becsült és tényleges piaci áruk különbsége szerint csökkenő sorrendben és ezek közül az első százat kiválasztani.

## Komponensek közötti kapcsolatok kialakítása

### Adatbázis – Webalkalmazás

A Django keretrendszer egyik legnagyobb erőssége az ORM (Object-Relational Mapper). Az ORM egy programozási technika adatok konvertálására adatbázisok és objektumorientált programozási nyelvek között. Az általunk írt programban egy objektum leképezést hoz létre, aminek a felépítése megegyezik az adatbáziséval.

A használata jelentősen csökkenti a megírandó kód mennyiségét.

A Django Model rétegében létre kell hoznunk az modelleket. Ez valójában osztályok definiálását jelenti, amelyek tulajdonságai megegyeznek az osztályok által modellezett táblák mezőivel.

Ezt nem muszáj kézzel megírni, van lehetőség meglévő adatbázis táblákból Django modelleket generálni és fordítva is.

Mivel az adatbázisban már létre voltak hozva a táblák a megfelelő kulcsokkal, ezért ezekből generáltam a Django modelleket:

1. A Django projektben található *settings.py* fájlban be kell állítani az adatbázis elérését (cím, felhasználónév, jelszó, adatbázis)
2. A projektmappában a következő parancsokat kell végrehajtani:
   * + *python manage.py makemigrations*
     + *python manage.py sqlmigrate [Projekt neve]*

Ezek hatására a projektkönyvtárban a *migrations* mappán belül megjelenik egy fájl, ami tartalmazza az automatikusan generált modelleket. Ezt természetesen le kell ellenőrizni, különös tekintettel a mezők típusára.   
Ha megfelelő a modell, akkor a fájl tartalmát be kell másolni a *models.py* fájlba és innentől az adatbázis struktúra leképezhető a Django-n belül.

A Python biztosít eszközöket az adatbázis lekérdezések kezelésére, de én jobban szeretem a nyers SQL lekérdezést saját magam megírni és úgy lefuttatni. Ez a megoldás biztosította számomra a szabadságot, hogy tetszőleges lekérdezéseket generáljak különböző logikák alapján.

Például: *players = Player.objects.raw(’SELECT \* FROM player’)*

A fenti programkód eredményeként a *players* változó tartalmazni fogja az összes játékost a tulajdonságaikkal együtt.

### 7.4.2 Adatbázis – Gépi tanuló komponens

Itt az *sqlalchemy* csomag *create\_engine* metódusát használtam. Ezzel a metódussal létrehoztam egy engine objektumot, aminek paraméterként meg kellett adni a kapcsolódás adatait (cím, felhasználónév, jelszó, adatbázis). Az így létrejött engine objektumnak meghívtam a *.connect()* metódusát, ezáltal kaptam egy connection objektumot.

A Python *pandas* csomagjában található *read\_sql()* metódussal tudtam elérni az adatbázist. Ennek paraméterként egy SQL lekérést kellett megadni és az előbb létrehozott connection objektumot. Mivel az SQL lekérést itt is én állítottam elő, kihasználtam ennek az előnyét és a megadott logikák szerint általam írt metódusokkal generáltattam.

Itt a webalkalmazással ellentétben nem kellett Modelleket definiálni és leképezni az adatbázis tábláit osztályokba. Ebben az esetben a főként *pandas* csomagban található *DataFrame* és *List* adattípusokon hajtottam végre a műveleteket.

### 7.4.3 Webalkalmazás – Gépi tanuló komponens

A Webalkalmazás a megfelelő funkció meghívásakor a Gépi Tanuló komponenshez fordul. A Gépi Tanuló komponens különálló, *.py* kiterjesztésű Python programokat jelent.

Azért, hogy tényleg különálló legyen, alapból egy másik mappába helyeztem és nem behívtam a Webalkalmazásba és ott futtattam belőle a megfelelő függvényeket, hanem a Windows környezetben futtattam. Ez azért előnyös, mert a két alkalmazás így független egymástól. Egy másik nyelven íródott futtatható fájlt (, amelynek az input és output értékei megegyeznek) is be lehet rakni a helyére.

A Python *subprocess* könyvtárával van erre lehetőség. A Webalkalmazásban a *subprocess.run()* metódust kell futtatnunk. Több paramétert meg lehet adni, az én esetemben így nézett ki:



### 

A FindSimilarPlayer.py fájl megkapja a kiválasztott játékos azonosítóját argumentumként (playerid), feldolgozza, majd visszaadja a hasonló játékosok azonosítóját, amik a subprocess\_output változóban tárolódnak, majd épülnek be az SQL lekérdezésbe.  
  
A Gépi tanuló komponensben kezelni kell, hogy kapni fog argumentumként egy azonosítót. A megtalált játékosok azonosítóját nem tudja közvetlenül visszaadni az őt meghívó Webalkalmazásnak, mivel nem függvényhívás történt, hanem a Windows környezetben lett elindítva az alkalmazás. Ezáltal viszont a *print()* metódusok eredményei, amik Windows konzolból elindítva a konzolra íródnának ki, most ezeket kapja meg a Webalkalmazás.

Ebben az esetben a *print()* függvényt használtam a *return* megvalósításaként és azzal adtam vissza értékeket. Figyelni kellett, hogy pontosan egy darab *print()* legyen a kódban, mert ha több mindent „ír ki” a program, akkor a webalkalmazás nem fogja tudni feldolgozni.

## Adatbetöltő script

Működése:

1. A script a paraméterként megadott, .csv kiterjesztésű fájl tartalmát beolvassa egy Pandas DataFrame típusú objektummá.
2. A Players és Score tábla tartalmát kimenti egy ideiglenes táblába, majd üríti ezeket a táblákat.
3. Az mezők neve szerint megpróbálja betölteni az adatokat a megfelelő táblába.
4. Ha sikeres volt a betöltés, néhány tesz lekérdezést futtat, hogy ténylegesen a megfelelő mennyiségű és minőségű adat került e be.
5. Ha az ellenőrzés sikeres, akkor az ideiglenes táblákból áttölti az adatokat a Players\_hist és Score\_hist táblákba úgy, hogy a *ValidTo* oszlopot az aktuális dátummal tölti ki.
6. Ha az ellenőrzés vagy az adatbetöltés során hiba merül fel, akkor az ideiglenes tábából az adatokat visszatölti az eredeti Player és Score táblákba a felhasználónak visszaadja a hibaüzenetet.

Ezzel a megoldással a Player és a Score táblákban az aktuális adatok szerepelnek, amik az alkalmazás elsődleges felhasználási módjához szükségesek, a historizációs riportokhoz felhasználható adatok pedig a Player\_hist és Score\_hist táblákban.

# TESZTELÉS

## Tervezés

A három nagy komponenst külön kell kezelni tesztelési szempontból. Ha mindegyik maradéktalanul eleget tesz az előre meghatározott funkcióknak, akkor rendszer egészének működését az alkalmazás használatával lehet tesztelni.

Ha a komponensek külön-külön átmennek a teszteken, akkor az előre meghatározott funkció lista teljesülését fogom figyelembe venni. Ha minden funkció működik és a rendszer tesztelhető állapotba kerül, akkor elkezdem az aktív használatát, hogy minél több hibát felderítsek.

## 8.2 Adatbázis

Az adatbázisban csak lekérdezések fognak futni, adatmanipuláció nem. A lekérdezések tesztelését a lekérdezést generáló függvényekben végeztem úgy, hogy összevetettem a függvények által visszaadott adathalmazt a közvetlenül az adatbázison lefuttatott lekérdezések eredményével.

Adatbetöltés csak az erre a célra készített adatbetöltővel történhet, ennek a tesztelését a 8.4 fejezet tárgyalja.

## Webalkalmazás és Gépi Tanuló komponens tesztelése

A Webalkalmazás és a Gépi Tanuló komponens üzleti logikájának függvényeire úgynevezett egységteszteket írtam, amelyek az adott függvények bemenetekre adott kimenetét vizsgálják. Több fajta bemeneti paraméternél kell vizsgálni a kimenetet, attól függően, hogy milyen lehetséges értékeket kaphat az adott függvény.

Ezek az egységtesztek egy külön, python környezetben futtatható fájlban találhatók. A python biztosít egy Unittest nevű tesztelési csomagot, amely a Test előtaggal kezdődő függvényeket automatikusan felismeri, és lefuttatja a bennük megírt egységteszteket.

Az egységtesztek leírása és eredménye az 2. Mellékletben található.

## 8.4 Adatbetöltő tesztelése

Az adatbetöltő scriptet a tényleges működése és használata során tudtam tesztelni.   
A fejlesztés során is figyelembe vettem a hibalehetőségeket. Annak elkerülése érdekében, hogy hibás vagy hiányos adatok kerüljenek a rendszerbe, a script bármilyen ellenőrzés közben fellépő hiba esetén visszafordítja a folyamatot és jelez a felhasználónak a hibával kapcsolatban. Ha a fájlban található adatok a korábban említett adatforrásokkal megegyező struktúrájúak, akkor hiba nélkül történik az adatbetöltés.

# AZ EREDMÉNYEK BEMUTATÁSA, ÉRTÉKELÉSE

A munka során az elsődleges célom az volt, hogy a feladatkiírásban szereplő pontok teljesüljenek és egy használható rendszert hozzak létre. Ezeket sikerült teljesítenem.

## 9.1 Eredmények

Az rendszert a terveknek megfelelően valósítottam meg, minden előzetes elvárásnak megfelel és plusz funkcióként az adatbetöltés lehetősége is megjelent.

Sikerült olyan eszközt készíteni, amely segítségével a felhasználók hozzáférhetnek a Football Manager nevű játék adatbázisában rejlő értékes információkhoz. Programozási és adatbázis tudás nélkül is dinamikusan tudják változtatni a lekérdezéseket, ezáltal a megjelenített játékosokat.

### 9.1.1 Komponensek

Az adatbázisban kialakítottam az adatmodellnek megfelelő struktúrát és az adattisztítás után betöltöttem az adatokat.

A Webalkalmazást olyan módon írtam meg, hogy minden előre meghatározott igényt kielégít. A felhasználói interfészen kívül kommunikál az adatbázissal és a Gépi Tanuló komponenssel, és megvalósítja az üzleti logikát.

A Gépi Tanuló komponens a megkapott paraméterek alapján elvégzi az üzleti logikának megfelelő műveleteket, majd a Webalkalmazásnak visszaadja az eredményt.

Az adatbetöltő lehetőséget ad az új adatok betöltésére, a régiek megtartása mellett.

## 9.2 Értékelés

A játékosok összehasonlítása egy fontos szempont lehet, amely által nagyságrendekkel csökkenthetik az átigazoláskor szóba jöhető játékosokat. Ez az adatbázis és az alkalmazás erőssége, hogy gyorsan és könnyen ki lehet szűrni az elvárásoknak nem megfelelő játékosokat, illetve objektíven össze lehet hasonlítani őket a képességeik alapján.  
Amiatt, hogy ez egy webes alkalmazás, a felhasználónak nem kell semmilyen programot telepítenie, egyszerűen elérheti a böngészőjén keresztül. Sajnos mobil eszközön a megjelenítés nem kényelmes, mert a táblázatban sok oszlop van egymás mellett, ez pedig csak hordozható vagy asztali számítógépen teszi lehetővé az alkalmazás használatát.

### Felhasználási lehetőségek

A rendszer a meghatározott céloknak megfelelően alkalmas arra, hogy olyan játékosokat azonosítsanak a segítségével, akik tehetségesek, de még alulértékeltek. A játékosok képességeinek egymáshoz viszonyítása objektíven megmutatja a köztük lévő különbségeket. A gépi tanuló algoritmusok és a szűrő mezők segítségével a felhasználó olyan játékosokat tud saját maga felfedezni, akik megtalálásához hosszabb kutatómunkák lennének szükségesek.

Ezek alapján a rendszer jó lehetőséget nyújt azon edzők/játékosmegfigyelők számára, akiknek nincs lehetőségük több százas/ezres nagyságrendű játékosok megvizsgálására. A rendszer nyújtotta lehetőségekkel a lecsökkenthetik a szóba jöhető és megvizsgálandó játékosok számát néhány tízes nagyságrendűre. Láthatják számok formájában, hogy a játékosok milyen területen erősek és milyen területen gyengék. Ez nagy lehetőséget jelent a tényleges üzleti felhasználásban.

A valós csapatok edzőin kívül segítséget nyújthat azoknak a játékosoknak, akik valamilyen virtuális, foci tematikájú játékban keresnek új játékosokat vagy szeretnék összehasonlítani őket.

A 3.3 Pontban célként kitűzött felhasználási lehetőségeknek eleget tesz az elkészült alkalmazás.

# 10. MEGVALÓSÍTÁS ELEMZÉSE, TOVÁBBFEJLESZTÉSI LEHETŐSÉGEK

A megvalósítás hatékony, mivel a felhasználó egy intuitív, könnyen használható felületet kap, ahol az interakciói szerint hajtódnak végre a háttérfolyamatok.

Az adatok megfelelő, könnyen hozzáférhető struktúrában vannak tárolva és a Python széles támogatottságának köszönhetően nagyon egyszerűen képződnek le a Webalkalmazás és a Gépi tanuló komponens számára.

A megvalósítás során törekedtem a gyakorlatias megoldások keresésére. A felesleges adatmozgatások elkerülése végett az adatlekéréseknél már az adatbázisban megtörténnek a szűrések és csak azok az adatok mozognak, amelyre a felhasználónak ténylegesen szüksége van. A komponensek közötti kommunikáció során is csak a játékosok azonosítói mozognak, nem a teljes adatszerkezet.

A függvények írásánál igyekeztem az Egyetlen Felelősség elvét (Single Responsibility) betartani és így jól tesztelhető programrészleteket írni.

A fejlesztés során mindig az felhasználói igény kielégülése, a performancia és a tesztelhetőség voltak az elsődleges szempontok. Ezeknek az alkalmazás minden részénél sikerült megfelelni.

## Továbbfejlesztési lehetőségek

Továbbfejlesztési lehetőségként szeretném megemlíteni a historikus adatokra építhető riportok lehetőségét. Sok információ kinyerhető belőle, ha látszódik a játékosok fejlődése. Megfelelő számú historikus adat alapján egy predikciós rendszer építésére is lehetőség lenne, amely megmutatja, hogy a játékosok a jövőben milyen teljesítményt érhetnek el.

A felhasználói keresések eredményének mentése és a találatként kapott játékosok táblázatos formába történő exportálása egy fontos funkció lehet.

Továbbfejlesztésként lehet még említeni, hogy az alkalmazás rendszeres időközönként frissítse a játékosok fizetésére vonatkozó adatokat egy megbízható forrásból, hogy mindig naprakész legyen.

# A SZAKDOLGOZAT TARTALMI ÖSSZEFOGLALÓJA MAGYAR ÉS ANGOL NYELVEN

## Magyar nyelven

A szakdolgozatom céljául tűztem ki:

* az adatalapú játékosmegfigyelés ismertetését a football tekintetében,
* a Football Manager nevű játék adatainak megfelelő struktúrában történő eltárolását,
* ezen adatok elérhetővé tételét egy webalkalmazás segítségével,
* szűrő és rendező lehetőségek biztosítása az adatokra,
* a gépi tanulás használatának lehetősége a felhasználó számára, a megfelelő játékosok megtalálásának érdekében.

Ez idő közben kiegészült egy plusz funkcióval: új adatok betöltésének lehetősége a játékosadatok frissen tartása érdekében.

Ezek mellett a játékosok értékelése, piaci áruk kialakulása is kifejtésre került. A Django, a PostgreSQL és a külső Python programok együttműködésének és megoldásainak bemutatása is fontos része a dolgozatnak.

Kitérek arra is, hogy miért ezeket az adatokat használtam és milyen jellegű üzleti értéket lehet kinyerni belőlük.

## In english

The aim of my thesis was:

* to exposit data-driven scouting in football,
* to store the data of Football Manager in the fitting structure,
* to create a web application to make the data reachable,
* to provide data filtering and sorting,
* to provide machine learning usage to the user to find the adequate players.

A new feature has been added: the opportunity of import new data to keep up to date the player informations.

In addition player ratings, their price development were explained. The presentation of cooperation and solutions of Django, PostgreSQL and external Python programs is an important part of the thesis.

I have also explained why I used this data and what kind of business value can be extracted from it.

# IRODALOMJEGYZÉK

[1] Advantages and Disadvantages of ASP.NET, (<https://www.software-developer-india.com/advantages-and-disadvantages-of-asp-net/>), utoljára megtekintve: 2021.12.04

[2] Comparing Database Management Systems: MySQL, PostgreSQL, MSSQL Server, MongoDB, Elasticsearch, and others (<https://www.altexsoft.com/blog/business/comparing-database-management-systems-mysql-postgresql-mssql-server-mongodb-elasticsearch-and-others/>), utoljára megtekintve 2021.12.03

[3] Concept work „Data-driven Scouting in Football” (<https://millernton.de/en/2020/07/03/konzeptarbeit-datengetriebenes-scouting-im-fussball/>), utoljára megtekintve 2021.11.14

[4] Daniel Memmert & Dominik Raabe : Data Analytics in Football: Positional Data Collection, Modelling and Analysis, *Routledge*, 2018

[5] Dave Braines & Joe Pavitt & Richard Tomsett : Applied AI Letters, *John Wiley & Sons, Ltd.*, 2021

[6] Füredi Balázs: Legyőzheti-e egy kis csapat a Liverpoolt az adatok segítségével? (<https://blog.hiflylabs.hu/hu/2020/06/11/legyozheti-egy-kis-csapat-a-liverpoolt-az-adatok-segitsegevel/>), utoljára megtekintve: 2021.12.06

[7] Here’s why England star Jude Bellingham’s No.22 was retired at Birmingham City (<https://www.birminghammail.co.uk/sport/football/football-news/jude-bellingham-birmingham-city-dortmund-20776728>), utoljára megtekintve: 2021.11.13

[8] Javas Spring Pros and Cons (<https://www.javatpoint.com/java-spring-pros-and-cons>), utoljára megtekintve: 2021.12.03

[9] Justin Harper: Data experts are becoming football’s best signings, (<https://www.bbc.com/news/business-56164159>), utoljára megtekintve: 2021.12.06

[10] Luca Pappalardo & Paolo Cintia & Paolo Ferragina & Emanuele Massucco & Dino Pedreschi & Fosca Giannotti: PlayeRank : Data-driven Performance Evaluation and Player Ranking in Soccer via a Machine Learning Approach, 2019 (<https://dl.acm.org/doi/pdf/10.1145/3343172>)

[11] Martijn Ernest: The importance of scouting, (<https://ifbi.brussels/programme/the-importance-of-scouting/>), utoljára megtekintve: 2021.12.06

[12] Oliver Müller & Alexander Simons & Markus Weinmann : Beyond crowd judgements: Data-driven estimation of market value in association football (<https://www.sciencedirect.com/science/article/pii/S0377221717304332>), utoljára megtekintve: 2021.12.02

[13] Tom Hamilton & Tor-Kristian Karlsen : Football Manager 2021 could find next Messi, but would club scouts use it? (<https://www.espn.com/soccer/soccer-transfers/story/4242573/football-manager-2021-could-find-next-messibut-would-club-scouts-use-it>), utoljára megtekintve: 2021.11.08

[14] Victor Rak, Pros and Cons of Ruby on Rails (<https://sloboda-studio.com/blog/pros-and-cons-of-ruby-on-rails/>), utoljára megtekintve: 2021.12.03

[15] Youssef Nader: What is Django? Advantages and Disadvantages(<https://hackr.io/blog/what-is-django-advantages-and-disadvantages-of-using-django>), utoljára megtekintve: 2021.12.03

[16] Django MVT (<https://www.javatpoint.com/django-mvt>), utoljára megtekintve: 2022.03.03.

[17] Django Documentation, Writing your first Django app, part 1 (<https://docs.djangoproject.com/en/4.0/intro/tutorial01/>), utoljára megtekintve: 2022.03.03.

# MELLÉKLETEK

## Melléklet – Játékosok pontszámai

Mindegyik játékos rendelkezik az alábbi tulajdonságokkal. Ezek 1-20 között egész számú értékek lehetnek, arányosan a játékos képességével.

Az 1 a legrosszabb, a 20 a legjobb adható pontszám.

Kapusoknál lényeges tulajdonságok

|  |  |
| --- | --- |
| AerialAbility | Milyen messzire ér el a kapus ugró helyzetben |
| CommandOfArea | Mennyire valószínű, hogy egy levegőben ívelt labdát megszerez a kapus |
| Communication | Mennyire jól tudja a kapus szervezni a védelmet |
| Eccentricity | Mennyire valószínű, hogy a kapus jól reagál a váratlan helyzetekben |
| Handling | Mennyire biztosan tudja elkapni a labdát |
| Kicking | Milyen messzire tudja rúgni a labdát a kapus |
| OneOnOnes | Mennyire jól teljesít a kapus az egy-az-egy elleni szituációkban |
| Reflexes | Milyen gyorsan képes reagálni a lövésekre |
| RushingOut | Mennyire jól tudja felmérni a kapus, hogy kirohanjon-e a kapujából megszerezni a labdát |
| TendencyToPunch | Mennyire valószínű, hogy a kapus csak kiüt egy fogható labdát, ahelyett, hogy elkapná azt |
| Throwing | Mennyire pontosan tudja a kapus a csapattársaihoz dobni a labdát |

Technikai tulajdonságok

|  |  |
| --- | --- |
| Corners | Mennyire pontosan tudja elvégezni a játékos a szögletet |
| Crossing | Milyen pontosan tudja a játékos a pálya széléről beívelni a labdát |
| Dribbling | Mennyire hatékonyan tudja kezelni a játékos a labdát, amikor nála van |
| Finishing | Mennyire pontosak a játékos kapura tartó lövései |
| FirstTouch | Mennyire hatékonyan tudja átvenni és továbbítani a labdát a játékos |
| Freekicks | Mennyire pontosan tudja elvégezni a szabadrúgásokat a játékos |
| Heading | Mennyire pontosan tudja a kívánt helyre fejelni a labdát a játékos |
| LongShots | Milyen pontosan tud kapura lőni messziről a játékos |
| Longthrows | Mennyire messzire tudja bedobni a labdát a játékos |
| Marking | Mennyire jól tud helyezkedni a játékos, hogy könnyen eljuthasson hozzá a labda |
| Passing | Mennyire pontosan tudja passzolni a játékos a labdát |
| PenaltyTaking | Mennyire pontosan tud elvégezni a játékos egy büntető rúgást |
| Tackling | Mennyire hatékonyan tudja elvenni a játékos a labdát az ellenféltől, anélkül, hogy szabálytalankodna |
| Technique | Mennyire jól tud magasabb technikai szintű |

Szellemi tulajdonságok

|  |  |
| --- | --- |
| Aggression | Mennyire valószínű, hogy a játékos belemegy egy fizikai konfliktusba |
| Anticipation | Mennyire jól tudja a játékos felmérni a csapattársai és ellenfelei következő lépését a játékos |
| Bravery | Mennyire hajlamos a játékos belemenni olyan szituációba, ami fájdalommal vagy sérüléssel járhat |
| Composure | Mennyire jól tűri a játékos a mentális nyomást meccs közben |
| Concentration | Mennyire tud koncentrálni a játékos a mérkőzés egésze alatt |
| Vision | Mennyire jól látja meg a játékos a megjátszható csapattársait |
| Decisions | Mennyire jó döntéseket hoz a játékos a labdával |
| Determination | Mennyire ad bele mindent a játékos a siker érdekében |
| Flair | Mennyire tud a játékos váratlanokat húzni a labdával |
| Leadership | Mennyire inspiráló és motiváló a játékos személyisége |
| OffTheBall | Mennyire jól tud a játékos a mozgásával zavart kelteni az ellenfélnél |
| Positioning | Mennyire jól helyezkedik a játékos labda nélkül |
| Teamwork | Mennyire pontosan követi a játékos a taktikai utasításokat |
| Workrate | Mennyire jól teljesít fizikálisan a játékos a meccs alatt |

Fizikai tulajdongások

|  |  |
| --- | --- |
| Acceleration | Mennyire gyorsan éri el a játékos a maximális sebességét |
| Agility | Mennyire jól tud a játékos megállni, elindulni és irányt váltani |
| Balance | Mennyire jól tud állva maradni a játékos fizikai nyomás alatt is |
| Jumping | Milyen magasra tud ugrani egy játékos fejelésnél |
| LeftFoot | Mennyire ügyesen használja a bal lábát a játékos |
| NaturalFitness | Mennyire jól tudja fenntartani a játékos az erőnléti állapotát meccsek között |
| Pace | Mennyire gyorsan tud futni egy játékos, ha eléri a maximális sebességét |
| RightFoot | Mennyire jól használja a jobb lábát a játékos |
| Stamina | Mennyire nehezen fárad ki a meccs közben a játékos |
| Strength | Mennyire tud a játékos fizikai előnyt szerezni egy ellenféllel szemben |
| Consistency | Mennyire valószínű, hogy a játékos a ki tudja aknázni a képességeit a meccsen |
| Dirtiness | Mennyire gyakran szokta a játékos megszegni a szabályokat, hogy ezzel előnyt szerezzen |
| ImportantMatches | Mennyire jól teljesít a játékos a nagy, fontos meccseken |
| InjuryProness | Mennyire valószínű, hogy a játékos megsérül meccsen vagy edzésen |
| Versatility | Mennyire jól teljesít a játékos egy számára idegen pozícióban |
| Adaptability | Mennyire könnyen tud a játékos új országban letelepedni |
| Ambition | A játékos mennyire törekszik a siker érdekében |
| Loyalty | A játékos mennyire becsületes, hűséges a csapatával szemben |
| Pressure | A játékos mennyire jól teljesít nyomás alatt |
| Sportsmanship | Mennyire sportszerű a pályán a játékos |
| Temperament | Mennyire könnyen veszíti el a hidegvérét a játékos |
| Controversy | A játékos mennyire vállalja fel a konfliktusokat |

## Melléklet – Egységtesztek leírása, eredménye

### Webalkalmazás üzleti logikájának egységtesztjei

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Függvény neve | Leírás | Bemenet | Kimenet | Tesztek kimenete |
| CheckboxValueToBoolean | A http kérésből a megadott jelölőnégyzet paraméter értékét Igaz/Hamis értékké alakítja, attól függően, hogy a jelölőnégyzet be van-e pipálva | Http kérés,  Jelölőnégyzet neve | True/False | SIKERES |
| ColumnVisibilitiesInit | Az alkalmazás összes betöltésekor a mezők láthatóságát Igaz értékre állítja | Oszlopok láthatóságát tartalmazó lista | Oszlopok láthatóságát tartalmazó lista, Igazra állított értékekkel | SIKERES |
| CreateOrderByList | A http kérés alapján egy listát készít a sorba rendezési mezők értékeivel | Http kérés | A sorba rendező mezők értékeit tartalmazó lista | SIKERES |
| CreateOrderByString | A sorba rendező mezők értékeit tartalmazó listából egy SQL szintaktikát követő sorba rendezési (ORDER BY) feltételt állít elő | A sorba rendező mezők értékeit tartalmazó lista | SQL szintaktikájú sorba rendezési feltétel | SIKERES |
| CreateWhereString | A szűrő mezők értékei alapján egy SQL szintaktikájú (WHERE) lekérdezési felételt hoz létre | Http kérés,  szűrő mezők nevei | SQL szintaktikájú lekérdezési feltétel | SIKERES |

### Gépi Tanuló komponens üzleti logikájának egységtesztjei

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Függvény neve | Leírás | Bemenet | Kimenet | Tesztek kimenete |
| FindTargetAbilities | A megadott játékosnak megkeresi azt a 20 tulajdonságát, amiben a legjobb | Játékos azonosítója | Játékos 20 legjobb tulajdonságát tartalmazó lista | SIKERES |
| FindPlayersWithSimilarAbilities | Megkeresi azokat a játékosokat, akiknek 20 legjobb tulajdonsága közül legalább 13 szerepel a paraméterként kapott tulajdonság listában | 20 tulajdonságot tartalmazó lista | Játékosazonosítók listája | SIKERES |
| FilterPlayers | Megkeresi azokat a játékosokat, akik a megadott tulajdonságokban legalább 7 értékkel rendelkeznek | Tulajdonságokat tartalmazó lista | Játékosazonosítók listája | SIKERES |
| DetermineOptimalKValue | A kapott játékos entitások tulajdonságait figyelembe véve meghatározza, hogy hány halmazba lehet őket optimálisan besorolni | A játékos entitások (Pandas Dataframe adattípus) | Halmazok optimális  száma | SIKERES |
| CompareLists | Összehasonlít két megadott listát és meghatározza, hogy hány közös elemük van | Lista1,  Lista2 | Közös elemek száma | SIKERES |
| GetPlayersOnPosition | Kiválogatja a megadott pozícióban játszó játékosokat | Pozíció | Játékosazonosítók listája | SIKERES |

1. Sport tudomány [↑](#footnote-ref-1)
2. Rövid, egy részfeladat automatizálására szolgáló program [↑](#footnote-ref-2)
3. Az 1. pontban említett pontszámok leírása az 1. Mellékletben található meg. [↑](#footnote-ref-3)