

OTDK-dolgozat

2025

Budapesti Corvinus Egyetem



Pénzügy Intézet

**Statisztikai modellek alkalmazása sportfogadó stratégiák kidolgozásához és
a sportfogadópiacok hatékonyságának vizsgálatához**

**The use of statistical models to develop sports betting strategies and assess
the efficiency of the sports betting markets**



NEMZETI KUTATÁSI, FEJLESZTÉSI
ÉS INNOVÁCIÓS HIVATAL

A Kulturális és Innovációs Minisztérium ÚNKP-23-2-II-CORVINUS88
kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési
és Innovációs Alapból finanszírozott szakmai támogatásával készült.

Készítette: Babos András

Finance MSc

I. évfolyam

Témavezető: Madari Zoltán

Adatelemzés és Informatika Intézet

Statisztika Tanszék

Kézirat lezárásának dátuma: 2024. április 17.

Absztrakt

A technikai fejlődés és a globalizáció napjaink egyik vezető szórakoztató iparágává tette a sportot, és különösen a labdarúgást. A sportág gazdasági fejlődésével párhuzamosan a kapcsolódó piacok, így a sportfogadópiacok is dinamikus növekedésen mentek keresztül az elmúlt években, nagyban köszönhetően a fogadási lehetőségek bővülésének és a mérkőzésekről gyűjtött adatok szélesebb körének. A közelmúlt trendjének megfelelően a mérkőzések kimenetelének előrejelzése és a piacok hatékonysága kiemelten foglalkoztatja a fogadókat és a kutatókat, az említett témák pedig jelen dolgozat céljával is szolgáltak. Kutatásom kérdése tehát az volt, hogy a találkozók kimenetelének előrejelzése által azonosítható-e statisztikai arbitrázs-lehetőség az angol első osztályú labdarúgó bajnokság (Premier League) sportfogadópiacán. A meccsek eredményét osztályozási (klasszifikációs) és regressziós feladatként is értelmezve gépi tanuláson és neurális hálón alapuló modellekkel jeleztem előre, melyek magyarázó változói a csapatok közelmúltbeli meccsein nyújtott teljesítményét, valamint a játékoskeret pillanatnyi jellemzőit ragadták meg. A korábbi kutatásokhoz képest a kezdőcsapatok piaci értéke, a sérült és eltiltott játékosok száma, valamint a csapatok aktuális formája is bevonásra került a modellekbe. A becslések alapján különböző kockázati paraméterekkel rendelkező stratégiákat definiáltam, melyek közül a legjobb 76% feletti pontosságot, valamint 2,13%-os meccsenkénti hozamot ért el. A hozam pozitivitása azonban statisztikailag nem bizonyult szignifikánsnak, megerősítve ezáltal a labdarúgómérkőzések kimeneteléhez kapcsolódó magas bizonytalanságot és a fogadási piacok többnyire hatékony működését. Összességében bár jelen kutatás nem tudott szignifikánsan pozitív hozamú stratégiát felmutatni, az előrejelzés módszertanának bővítésével és az optimális fogadási stratégiák kockázati jellemzőinek meghatározásával utat mutathat a jövőbeli elemzések számára.

Kulcsszavak: Futballmérkőzések előrejelzése, machine learning, sportfogadópiac, hatékonyság, hozam-kockázati paraméterek

Abstract

Due to the technological development and globalisation in recent decades, sport and football has become one of the leading entertainment industries of our time. In parallel with the economic development of the sport, sports betting as a related market have experienced dynamic growth in recent years, largely thanks to the expansion of betting opportunities and the wider range of data collected on matches. In line with the recent trend, the prediction of match outcomes and the efficiency of these markets have been a major concern for bettors and researchers, and these topics have been the focus of this thesis. The question of my research was whether statistical arbitrage opportunities can be identified in the English Premier League sports betting market by predicting the outcome of matches. I predicted match outcomes both via classification and regression using machine learning and neural network models, whose explanatory variables captured the teams' performance in recent matches and the characteristics of the actual game. In addition to previous research, the market value of starting teams, the number of injured and suspended players, and the current form of teams were also included in the models. Based on these estimates, strategies with different risk parameters were defined, the best of which achieved an accuracy of over 76% and a return per match of 2.13%. However, the positivity of the return was not statistically significant, confirming the high uncertainty associated with the outcome of football matches and the mostly efficient functioning of betting markets. Overall, although the present study could not show a significantly positive return strategy, it may point the way to future analyses by extending the forecasting methodology and identifying risk characteristics of optimal betting strategies.

Keywords: football match prediction, machine learning, sports betting market, efficiency, return-risk parameters

Tartalomjegyzék

I. Bevezetés	1
II. Szakirodalmi áttekintés.....	4
II.1. A sportfogadópiacok működése és hatékonysága.....	4
II.2. A labdarúgómérkőzések kimenetelének előrejelzése matematikai modellekkel.....	9
II.3. A futballmeccsek kimenetelének adatalapú előrejelzése és a statisztikai arbitrázs jelensége	10
III. Felhasznált adatok és a kutatás módszertana	14
III.1. Felhasznált adatok és leíró statisztika	14
III.2. A modellezés során használt magyarázó- és eredményváltozók, illetve az oddsok bemutatása	15
III.3. Az előrejelző modellek kiválasztása, tanítása és tesztelése.....	19
III.4. Az előrejelző modellek felépítése	22
III.5. A sportfogadási stratégiák kidolgozásának és tesztelésének módszertana.....	24
IV. A kutatás eredményei.....	27
IV.1. Az előrejelzések és a fogadási stratégiák statisztikai elemzése	27
IV.2. A fogadási stratégiák értékelése pénzügyi szempontból.....	33
IV.3. A fogadási stratégiák összegzése, értékelése	40
V. Összegzés	43
VI. Hivatkozások jegyzéke	46

Ábrajegyzék

1. Ábra: A 2020-2021-es futbalszezon során a topligák mérkőzéseire elhelyezett tétek összege bajnokságokra lebontva. Forrás: Statista.com, 2023.	6
2. Ábra: A gólkülönbség eloszlása. Forrás: Saját szerkesztés.	17
3. Ábra: A Premier League mérkőzéseinek kimenetele (2018-2023). Forrás: Saját szerkesztés.	18
4. Ábra: A klasszifikációs modellek előrejelzéseinek megoszlása. Forrás: Saját szerkesztés.	27
5. Ábra: A sportfogadó stratégiákhoz tartozó konfúziós mátrixok. Forrás: Saját szerkesztés.	32
6. Ábra: A fogadási stratégiák fogadási egységben mért kumulált hozamai az idő függvényében (1. rész). Forrás: Saját szerkesztés.	35
7. Ábra: A fogadási stratégiák fogadási egységben mért kumulált hozamai az idő függvényében (2. rész). Forrás: Saját szerkesztés.	36
8. Ábra: A fogadási stratégiák által kiválasztott oddsok hisztogramja a 2019-2023-as tesztidőszak mérkőzései alapján. Forrás: Saját szerkesztés.	37
9. Ábra: A gólkülönbségbecslés küszöbértékének érzékenység vizsgálata: mérkőzésenkénti hozam és fogadásszám a küszöbérték függvényében. Forrás: Saját szerkesztés.	38
10. Ábra: A mérkőzések kimenetlénének megoszlása az átlagosan becsült gólkülönbség függvényében. Forrás: Saját szerkesztés.	40
11. Ábra: A mérkőzések kimenetlénének megoszlása az átlagosan becsült gólkülönbség függvényében. Forrás: Saját szerkesztés.	41

Táblázatok jegyzéke

1. Táblázat: A futballmérkőzések kimenetelét statisztikai módszertanok segítségével előrejelző releváns kutatások összefoglalása. Forrás: Saját szerkesztés.	11
2. Táblázat: A modellezés során használt változók leíró statisztikai adatai. Forrás: Saját szerkesztés.	15
3. Táblázat: A vizsgált mérkőzések oddsainak fő statisztikai jellemzői. Forrás: Saját szerkesztés.	18
4. Táblázat: A modellezés során használt hiperparaméterek optimális értékei. Forrás: Saját szerkesztés.	22
5. Táblázat: A fogadási stratégiák teszteléséhez használt mérkőzések kimenetelének megoszlása, illetve a kapcsolódó oddsok főbb statisztikai jellemzői. Forrás: Saját szerkesztés.	26
6. Táblázat: A regressziós modellek által becsült gólkülönbségek főbb statisztikai mutatói. Forrás: Saját szerkesztés.	28
7. Táblázat: A sportfogadóstratégiák pontossága és alapvető jellemzői. Forrás: Saját szerkesztés.	29
8. Táblázat: A vizsgált fogadási stratégiák főbb sajátosságai hozam és kockázati szempontból a 2019-2023-as időszak mérkőzései alapján. Forrás: Saját szerkesztés.	33

I. Bevezetés

Napjainkra a labdarúgás a világ egyik legnagyobb szórakoztató iparágává vált. A legnagyobb téttel bíró mérkőzéseket emberek milliói követik világszerte (a 2022-es FIFA Labdarúgó Világbajnokság döntőjét például a Nemzetközi Labdarúgó Szövetség becslése szerint közel 1,5 milliárd ember követte világszerte – FIFA.com, 2023), a tradicionális futballklubok többmilliós nemzetközi szurkolói táborral rendelkeznek, a sportág gazdasági súlya pedig nagyobb, mint valaha. A Statista 2018-as becslése alapján a sporthoz kapcsolódó piacok mérete a világ GDP-jének 0,5%-át tette ki (471 Milliárd dollár). Bár a labdarúgás gazdasági növekedése 2020-ban a Covid-19 járvánnyal járó általános globális recesszió következtében hirtelen megtorpant, a járvány elmúlásával hamar visszaállt a korábbi dinamikus növekedési pályára. Az elmúlt évtizedek technológiai fejlődése felgyorsította a sportág globalizálódását is, mely a korábbinál sokkal jobban integrált piacokat (gondolhatunk itt a játékospiacra, a sportturizmusra, de akár a merchandising termékek, vagy közvetítési jogok piacára is), szélesebb szurkolói bázist és a mérkőzésekhez, mint termékhez/szolgáltatáshoz való sokkal könnyebb hozzáférést eredményezett.

A labdarúgás népszerűségének egyik legfontosabb eleme a játék nehézségéből eredő kiszámíthatatlanság. A többi csapatsportághoz képest drasztikusan alacsony gólok/pontok száma sok esetben megjósolhatatlanná teszi a mérkőzések kimenetelét, melyre egy-egy véletlen esemény akár döntő befolyással is lehet. A labdarúgás ezen sajátossága számos esetben eredményez a csapatok teljesítményével (melyet manapság már komplex statisztikai mutatókkal mérnek már) ellentétes végeredményt, mely a sportklubok szakmai-gazdasági igazgatására és a szurkolók magatartására is jelentős hatást gyakorol. Ezen kiszámíthatatlanságra épül a sportág egyik legnagyobb piaca, a sportfogadás is, amelyet a bukmékerek és a fogadók mérkőzésről alkotott eltérő véleménye tart működésben.

Dolgozatom fókuszában a labdarúgás egyik kapcsolódó piaca, a sportfogadópiac áll. A futballt a kezdetektől kísérő sportfogadás szintén komoly átalakuláson ment keresztül az 1990-es évektől kezdődően. Az internet elterjedése megágyazott az online sportfogadásnak, amely idővel kiegészült az élő fogadás lehetőségével, valamint számos egzotikus eseménnyel is. A piac bővülésével párhuzamos egyre nagyobb lett az igény a meccsek kimenetelének, mint legjellemzőbb fogadási formának a hatékony előrejelzésére. A probléma komplexitása a szurkolók mellett számos kutatót is inspirált már, akik kvalitatív (szakértői vélemények) és kvantitatív adatok (például statisztikák, a meccsek külső körülményei) alapján próbálták megbecsülni a mérkőzések végeredményét, majd eredményeikre sportfogadó stratégiát építeni.

A korábbi tanulmányokkal összhangban kutatásom célja az angol első osztályú labdarúgó bajnokság (Premier League) futballmeccseinek előrejelzése gépi tanulást (machine learning) és neurális hálókat alkalmazó kvantitatív modellek segítségével. Modelljeimet a csapatok közelmúltban lejátszott mérkőzéseinek teljesítménymutatóira alapoztam, de a releváns kutatásokhoz képest több fontos külső változóval (kezdőcsapatok piaci értéke, sérültek és eltiltottak száma, illetve a csapatok formája) is kiegészítettem. A futballmeccsekről elérhető információ növekedését kihasználva a pontosabb becslés érdekében adatbázisomba bevontam számos, csak 2018 óta elérhető új statisztikai adatot (mint például a várható gólok - xG, várható gólpasszok - xAG, vagy részletes passz-statisztikák, ...) is. Az előrejelzések alapján különféle sportfogadó stratégiákat definiáltam, majd őket historikus oddsokon tesztelve megpróbáltam statisztikai arbitrázslehetőségeket¹ azonosítani a fogadópiacon.

A fogadási stratégiák statisztikai elemzésével arra a kérdésre kerestem a választ, hogy a legújabb statisztikai adatok, illetve a modellezésbe bevont külső változók magasabb becslési pontosságot eredményeznek-e. Az ezt követő pénzügyi elemzés (mely a hozam és a kockázat értékelését foglalta magába) során pedig a piac hatékonyságát, illetve a sikeres fogadási stratégiák főbb karakterisztikáit vizsgáltam. Ez alapján kutatási kérdéseim a következőképp formalizálhatók:

1. A legújabb statisztikai adatok felhasználása, illetve lényeges külső változók előrejelző modellekbe történő bevonása képes emelni a stratégiák előrejelző pontosságát?
2. A mérkőzések kimeneteléhez kapcsolódóan azonosítható statisztikai arbitrázs az angol Premier League sportfogadópiacán?

Az elemzés zárásaként részletes vizsgálatnak vettem alá az egyes fogadási módszerek kockázattvállalási hajlandóságát, a becslések korlátait pedig érzékenységvizsgálattal határoztam meg.

A dolgozat további felépítése a következő: a második fejezet a feldolgozott szakirodalom alapján összefoglalja a sportfogadópiacok, illetve a futballmeccs-kimenetek előrejelzésének korábban próbált módjait, míg a harmadik fejezet bemutatja az összegyűjtött adatbázis statisztikai jellemzőit, valamint az előrejelzéshez alkalmazott módszertant. A kutatás

¹ Alexander & Dimitriu (2005) definíciója alapján a statisztikai arbitrázs olyan pénzügyekben elterjedt replikáló stratégia, amely komplex statisztikai modellek segítségével automatikus (emberi beavatkozás nélküli) kereskedésen keresztül képes kihasználni a piacok pillanatnyi félrearázásait és statisztikailag szignifikáns hozamot generálni.

eredményei a negyedik fejezetben olvashatók, mely után a dolgozat az ötödik, összegző fejezettel zárul.

II. Szakirodalmi áttekintés

II.1. A sportfogadópiacok működése és hatékonysága

A futballmérkőzésekhez már a kezdetektől kapcsolódó sportfogadás ugyan a sportággal együtt komoly változásokon esett át az elmúlt évtizedekben, de alapvető jellemzői és működési logikája változatlanul megmaradt. A mérkőzéshez kapcsolódó események oddsait a piacot irányító sportfogadó irodák és –társaságok határozzák meg, a jellemzően magánszemély fogadók pedig számukra optimális módon választanak mérkőzéseket és teszik meg tétjeiket saját megérzéseik és információik alapján. A pénzügyi piacoktól eltérően a sportfogadó társaságok nem árjegyzőként tevékenykednek (azaz hangolják össze a keresletet a kínálattal), hanem az oddsok meghatározásával jellemzően komoly kockázatot vállalnak a fogadók tömegével szemben (Levitt, 2004). Ezen okból kifolyólag a fogadóirodák érdeke a fogadásra kínált események bekövetkezési valószínűségét tükröző oddsok olyan beállítása, mely hosszú távon profitábilis működést tesz számukra lehetővé. Pontatlanul meghatározott oddsok ugyanis nemcsak alkalmanként, hanem akár konzisztensen is képesek lehetnek aláásni a társaság pénzügyi teljesítményét.

Levitt (2004) tanulmánya alapján a fogadóirodák két módon, illetve ezek kombinációjával tudják hosszútávú profitabilitásukat biztosítani. Az egyik módszer arra épül, hogy a bukméker az oddsok beállításával körülbelül kiegyenlíti a két oldalra érkező téték volumenét. Ezen viselkedésével a fogadóiroda képes lehet mérsékelni, vagy teljes egészében lefedezni a kockázatot, így a szolgáltatásnyújtásért felszámolt teljes jutalékot profitként zsebre teheti. A másik módszer a fogadható események bekövetkezési valószínűségének, azaz a fogadás fair árának lehető legpontosabb meghatározása. Ilyenkor bár a bukméker által felvett pozíció nyitott marad és számos alkalommal veszteséghez vezet, a nagy számok törvénye alapján hosszú távon mégis profitábilis marad a fogadóiroda – köszönhetően megint a szolgáltatásnyújtásért felszámított jutaléknak. Ez utóbbi módszer azonban csak akkor fenntartható, ha a fogadásokat kiíró szervezet képes konzisztensen pontosabb becsléssel előállni, mint a fogadók. A feltétel biztosítása érdekében napjaink sportfogadó irodái komplex modellek és szakértők sokaságát használják előrejelzéseik során, melyeket időről időre korrigálnak a birtokukba kerülő újabb és újabb információk alapján².

² Feddersen és szerzőtársai (2023) kutatásukban azt vizsgálták, hogy a fogadóirodák vajon számottevően korrigálják-e előrejelzéseiket (és oddsaikat) a csapatok által várhatóan kifejtett erőfeszítésnek megfelelően. A kutatók eredményeikben igazolták a vizsgált jelenség létezését, melyet azzal indokoltak, hogy a komolyabb hosszú távú célok (bajnoki cím, nemzetközi kupaszereplés kiharcolása, vagy a kiesés

A sportfogadópiacok működésének központi fogalma az odds, amely a fogadásra kínált események bekövetkezési valószínűségét hivatott tükrözni. Az oddsok és a valószínűség ekvivalenciáját a következő formula biztosítja:

$$Odds(A) = \frac{1}{P(A)}$$

ahol $P(A)$ a fogadásra kínált esemény valószínűségét jelöli. Azaz például, ha egy mérkőzésen a hazai győzelem, a döntetlen, és a vendégsiker valószínűsége 50%, 25% és 25%, akkor az ehhez kapcsolódó elméleti oddsok 2,00, 4,00 és 4,00 lennének. Mivel a fair oddsok használata hosszú távon 0 vállalati profitot eredményezne, az elméleti értékekből a fogadóirodák jellemzően levonnak valamennyit szolgáltatásnyújtási díjként³. A jutalék felszámítása miatt tehát a fogadó minden kimenetel esetén a várható hozamnál alacsonyabb kifizetésben részesül, a különbség pedig a szolgáltatást kínáló szervezet zsebében marad profitként.

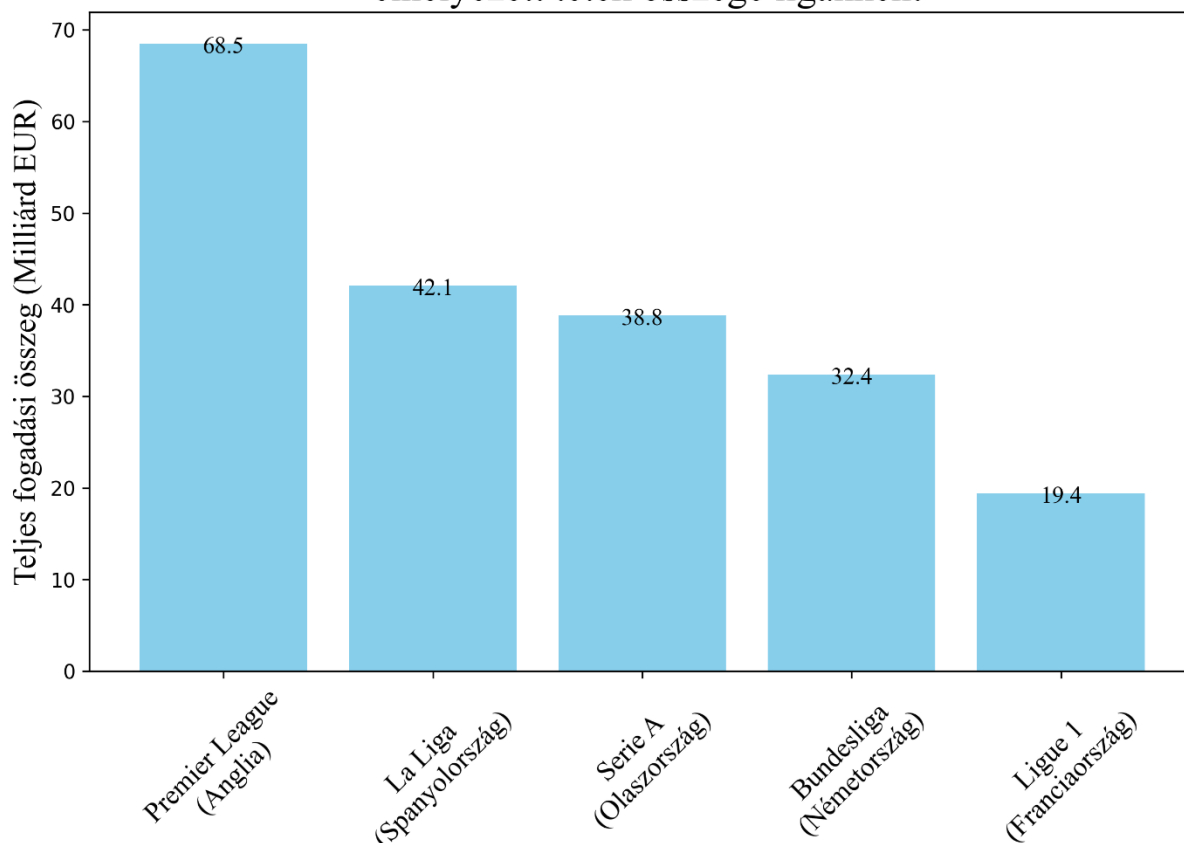
A téték elhelyezésének logikája a következőképp néz ki: a fogadó kiválasztja a számára szimpatikus mérkőzést és a kapcsolódó eseményt, majd egy általa meghatározott tétet (jelöljük F -fel) helyez a választott kimenetelre. Sikeres fogadás esetén a nyeremény összege az elhelyezett tét és az odds szorzata ($F * odds$) lesz, míg helytelen tipp esetén a teljes feltett tét elveszik.

Bár az alapvető mechanizmusok időben változatlanok maradtak, az online fogadás elterjedése, valamint a hirdetett események körének kiszélesedése jelentősen megváltoztatta a piac szerkezetét és elmúlt évtizedes trendjeit (Parke & Parke, 2019). A sportág globalizálódásából a digitalizáción maga is áteső fogadóipar is rengeteget profitált, melyet jól igazol a globális online fogadópiac méretének megnégyszereződése a 2017 (13,69 Milliárd USD) és 2023 (43,55 Milliárd USD) közti időszakban (Statista.com, 2024a). A dinamikus növekedés folytatását illetően egyetért a piackutató társaságok többsége, akik az európai piacon 5-10%-os éves bővülést jósoltak a következő szűk tíz éves időszakra (DBMR, 2022; GII, 2022, GVR, 2024 és Statista.com, 2024b). A sportfogadás 21. századi trendjeit megfigyelve egyáltalán nem meglepetés, hogy napjaink vezető sportfogadó irodái online lehetőségeket kínáló weboldalak, mint például a Bet365, a William Hill, a Betfair, a Unibet vagy a Betway (completesports.com, 2023).

elkerülése) nélkül pályára lépő csapatok hajlamosak csökkenteni erőfeszítéseiket, mely gyakran a győzelmi valószínűségük leértékeléséhez vezetett.

³ Ezen üzleti magatartás azt eredményezi, hogy a piacon megfigyelhető oddsokból számított valószínűségek összege meghaladja az egyet.

A 2020-2021-es szezonban a topbajnokságok mérkőzéseire elhelyezett tétek összege ligánként



1. Ábra: A 2020-2021-es futbalszezon során a topligák mérkőzéseire elhelyezett tétek összege bajnokságokra lebontva. Forrás: Statista.com, 2023.

Az európai labdarúgáshoz kapcsolódó fogadások nagy részben az öt topligához⁴, illetve a nemzetközi versenysorozatokhoz (UEFA Bajnokok Ligája, UEFA Európa Liga, UEFA Konferencia Liga) kapcsolódnak, de a célpontok közül magasan kiemelkedik a világ leggazdagabb és legszórakoztatóbb bajnoksága, az angol Premier League. Az angol első osztályú bajnokság dominanciája tehát nemcsak szakmai⁵ és gazdasági⁶ területen, hanem a sportfogadópia kereskedési volumenében is megfigyelhető. Ez utóbbit támasztja alá a fentebb látható 1. ábra, amely a 2020-2021-es szezonban a topbajnokságok mérkőzéseire elhelyezett

⁴ Európa top öt labdarúgó ligája közé az angol Premier League, a spanyol La Liga, az olasz Serie A, a német Bundesliga, illetve a francia Ligue 1 tartoznak.

⁵ A legjobb képességű futballisták magas koncentrációját támasztja alá a transfermarkt.com (2024) értékelése is, mely szerint a Premier League-ben játszó játékosok összértéke meghaladja a sorban második legértékesebb játékosokkal rendelkező bajnokságban, a La Ligában élő szerződéssel rendelkező sportolók összértékének kétszeresét.

⁶ A Deloitte 2023-as Annual Review of Football Finance elemzése alapján a Premier League-klubok 2021-2022-es szezon során realizált bevétele szintén kétszerese a rangsorban második La Liga egyesületei által generált forgalomnak.

tétek összegét vizualizálja bajnokságokra lebontva. Bár 2020-2021-ben a Premier League mérkőzéseire több, mint másfélszer annyi tét érkezett be, mint a sorban második La Liga meccseire, összességében mindegyik bajnokság sportfogadópiaca likvidnek mondható, mely jellemzően magas fokú hatékonyságot is biztosít.

A fogadási piacok általánosan magas szisztematikus kockázati szintje (mely a mérkőzések előrejelzéséhez kapcsolódó bizonytalanságból fakad), az iparág globalizálódása, valamint a fogadóirodák közti verseny erősödése gyakran megkérdőjelezi a hatékonyság kérdését, mely ezért már évek óta a piacot érintő kutatások egyik központi kérdése. A pénzügyi piacokkal kapcsolatban Fama vetette fel először a hatékony piacok elméletét (1970), melynek gyenge formája azt jelenti, hogy a pillanatnyi árak tükrözik minden múltban megfigyelt releváns információt (Angelini & De Angelis, 2019) és a jövőbeli árak lehető legpontosabb előrejelzéséül szolgálnak.

A sportfogadópiacok hatékonyságának gyenge formáját kezdetben OLS-regresszióval, majd rendezett probit regressziós modellekkel vizsgálták. Ezt a módszertant választotta Goddard és Asimakopoulou (2004) is, akik az angol Premier League 1999-2001-es időszakának mérkőzései alapján cáfolták a hatékonyság gyenge formáját. Elemzésük során ugyanis képesek voltak olyan fogadási stratégiát azonosítani⁷, amely statisztikailag szignifikáns hozamot eredményezett a tesztek során. Vlastakis és szerzőtársai (2009) több európai bajnokság mérkőzéseit tartalmazó adatbázison vizsgálták a kérdést, és szintén arra a következtetésre jutottak, hogy lehetséges a fogadóirodákat konzisztensen megverő fogadási stratégiát létrehozni. Poisson-eloszlásra építő multinomiális logit regressziós modelljével a szerzőpáros két anomáliát, a hosszútáv torzító hatását (long-shot bias), illetve a vendégcsapatok teljesítményének felülbecslő jelenséget is azonosították. Angelini és De Angelis (2019) 11 európai bajnokság találkozóit és a kapcsolódó online sportfogadópiacokat elemezve eltérő mértékű hatékonyságot tapasztaltak. A kutatók a 41 bukméker által meghatározott oddsok átlagát használva megerősítették az online piacok gyenge formájú hatékonyságát, a legmagasabb oddsok alkalmazása esetén – amelyeket a fogadók a piaci verseny miatt szabadon választhatnak – ugyanakkor anomáliákat is azonosítottak. Eredményeikkel három bajnokság mérkőzéseire is kapcsolódóan is megerősítették a hosszútáv torzító hatását, mely a fogadói jutalék megfizetése mellett is abnormális hozamhoz vezetett.

⁷ A szerzők szerint a piaci hatékonyság elsősorban a szezon végén (áprilisban és májusban) eltérő téttel játszott mérkőzésekhez kapcsolódott, így fogadási stratégiájuk is ezen időszakban rendezett találkozók előrejelzéséhez kapcsolódott.

Elaad és szerzőtársai (2020) több, mint 16 000 angol labdarúgó mérkőzés eredményeit és 51 online bukméker ajánlatát OLS-regressziókkal elemezve megállapították, hogy az oddsok általánosságban fairek, és mentesek mind hosszútávú, mind pedig a kimenetek előrejelzéséből adódó torzító hatástól. Mindazonáltal ők is igazolták, hogy az egyes bukmékerek által beállított oddsok önmagukban nem hatékonyak, mivel gyakran nem használták fel a versenytársak oddsai által tükrözött információkat. A szerzők eredményeikben mindezek mellett a fogadói jutalék és az irodák profitabilitásának csökkenéséről is beszámoltak.

A korábbi kutatásoktól eltérően (Richard & Vecer, 2021) a hatékony piacok martingál-tulajdonsága⁸ alapján közelítette meg a kérdést. Valószínűségbecslő modelljükkel a szerzők azt vizsgálták, hogy múltbeli információk hatást gyakorolnak-e a jelenbeli oddsra. Bayes-i faktorokra és a fogadók logaritmikus hasznosságfüggvényének maximalizálására alapozott kutatásuk során azonban nem találtak bizonyítékot piaci anomáliák létezésére. Végezetül a legfrissebb kutatás, Winkelmann és társai munkája (2024) abból a feltevésből indult ki, hogy a korábban azonosított torzító hatások jelentős része a vizsgált minta korlátozott méretének tudható be. Szimulációs elemzésükkel arra a következtetésre jutottak, hogy bár az elmúlt években a sportfogadó irodák jelentősen fejlesztették előrejelző módszereiket, időszakosan előfordulhatnak torzító hatások a piacon. Ezután az öt európai topliga 2005-2019 között lejátszott mérkőzéseit empirikus elemzésnek vetették alá és azt tapasztalták, hogy bár egyes szezonokban előfordultak piaci anomáliák, ezek nem voltak sem tartósak, sem szisztematikusak a ligákon keresztül. Mindezek mellett a szimulációs elemzés azt is alátámasztotta, hogy teljes piaci hatékonyság esetén is előfordulhatnak statisztikailag szignifikáns torzító hatások bizonyos időszakokban.

Összességében megállapítható, hogy a sportfogadópiacok hatékonysága az elmúlt években és napjainkban is számos kutatást inspirál, az eddigi modellek és elemzések azonban nem vezettek egyértelmű eredményre.

⁸ A pénzügyi piacokon megszokott martingál-tulajdonság az árfolyamok várható értékére vonatkozik, és azt mondja ki, hogy a jövőbeli várható értékre a legjobb becslés az aktuális árfolyam.

II.2. A labdarúgómérkőzések kimenetelének előrejelzése matematikai modellekkel

A sportfogadópiacok elemzésének kulcsfontosságú eleme a futballmérkőzések kimenetelének előrejelzése, melyhez a kutatók általában komplex matematikai, vagy statisztikai modelleket használnak. A problémát kezdetben többnyire az első módon közelítették meg, melyet a matematikai formulák kisebb adatigénye (gyakran egyedül a gólok eloszlására volt szükség az elemzéshez), egyszerűsége és könnyebb interpretálhatósága indokolt.

A releváns kutatások egy része kiemelt sportesemények (például Labdarúgó Világbajnokság vagy Európa Bajnokság) meccseinek előrejelzéséhez kapcsolódott. Az első próbálkozók között volt Stefani (1980), aki a legkisebb négyzetekre (OLS) épülő minősítő modelljével tett kísérletet az 1978-os világbajnokságon játszott találkozók eredményeinek becslésére. A szerző módszertanát professzionális és amatőr amerikai futballmeccsekre, illetve kosárlabdameccsekre is alkalmazta. Zeileis és szerzőtársai (2016 és 2018) számos online fogadóiroda véleményének konszenzusa alapján szimulálta le a 2016-os Európa Bajnokság, valamint 2018-as Labdarúgó Világbajnokság mérkőzéseit, és bár előrejelzéseik nem voltak minden esetben pontosak⁹ (a győztes válogatottat például egyik esetben sem tudták eltalálni), az esélyes csapatok teljesítményét pár kivételtől eltekintve helyesen modellezték. Groll és szerzőtársai (2018) a 2002-2014 közti világbajnokságok adatai alapján alkalmaztak Poisson-regressziós, illetve Random Forest (véletlen erdő) - alapú modelleket a 2018-as torna mérkőzéseinek szimulálására.

Természetesen előrejelző modellek és fogadási stratégiák a különböző bajnokságok mérkőzéseivel kapcsolatban is születtek. A matematikai alapokon nyugvó előrejelzés egyik alapvető munkája Maher (1982) nevéhez kötődik, aki a csapatok támadó- és védekező-képességét alapul véve modellezte a meccsek lehetséges kimenetelét. A szerző a gólok független, illetve kétváltozós Poisson-eloszlását feltételezve is szignifikáns eredményeket ért el. Rue és Salvesen (2000) az angol Premier League 1993-1994-es szezonjától az 1997-1998-as idényéig játszott mérkőzések adatait feldolgozva építettek fel előrejelző modellt, melyet a gólok Poisson-eloszlására és a csapatteljesítmény idő függvényében történő változására alapoztak. Dinamikusan változó általánosított lineáris modelljükkel a szerzők képesek voltak emelni a kimenetek meghatározásának pontosságát, és szignifikáns hozamot generálni az angol és a spanyol első osztályú mérkőzésekhez kapcsolódó fogadópiacon. Koopman és Lit

⁹ A meghirdetett oddsok természetéből adódóan a kutatók csak valószínűségi becslést tudtak adni, amely természetesen mindig magában hordozza a tévedés lehetőségét.

2015) a Poisson-eloszlás egy komplexebb változatát, egy sztohasztikusan változó intenzitás-értékekkel rendelkező kétváltozós modellt építettek fel, az erre alapozott stratégiájuk pedig képes volt megverni a fogadópiacokat. Diplomamunkájában Csikai (2016) az élő fogadások arbitrázsmentességét szintén a gólok Poisson-eloszlásának feltételezésével és dinamikusán változó intenzitásfüggvények kalibrálásával vizsgálta. Hasonló keretrendszerből kiindulva Mattera (2021) munkájában bináris fogadási kimenetek bekövetkezési valószínűségét modellezte az általánosított autoregresszív score (GAS – Generalized Autoregressive Score) módszerének segítségével.

A fogadási események előrejelzése mellett magát az optimális fogadási körülményeket tekintve is születtek érdekes kutatások. Archontakis és Osborne (2007) például dolgozatukban a Fibonacci-stratégiát alkalmazva megmutatták, hogy relatíve magas kockázatvállalással lehetséges pozitív hozamokat elérni. Matej és szerzőtársai (2019) a pénzügyekben ismert Kelly-kritériumot¹⁰ elemezték a sportfogadás kontextusában, és arra jutottak, hogy a sikeres befektetés szükséges feltétele a kritérium, mint kockázatvállalást kontrolláló mechanizmus alkalmazása. A szerzők stratégiáikat lóversenyeken, kosárlabdameérközéseken és futballmeccseken is tesztelték, mely alapján úgy vélték, a fogadópiacokon a részleges Kelly-kritérium képes optimális módon kezelni a felmerülő kockázatokat.

II.3. A futballmeccsek kimenetelének adatalapú előrejelzése és a statisztikai arbitrázs jelensége

Profitabilitásuk fenntartása érdekében a sportfogadóirodák igyekeznek a lehető legpontosabb becslést adni a fogadásra kínált események fair árára (azaz bekövetkezési valószínűségére), melyhez jellemzően nagy mennyiségű információt és komplex modelleket használnak fel. Ezzel párhuzamosan a fogadók is igyekeznek minél több információt minél hatékonyabban hasznosítani, hogy a piacon időszakosan fellelhető arbitrázslehetőségeket ki tudják használni. Más szóval a piac dinamikáját az irodák és a fogadók pontosabb előrejelzéseikért folyó állandó versengése határozza meg. Bár a fogadóirodák szakértelmüknek köszönhetően előnyösebb helyzetben vannak, mint a fogadók, a labdarúgást övező adatmennyiség növekedése, illetve a módszertani keretrendszer fogadók által vezetett folyamatos, gyors fejlődése napról napra kihívás elé állítja őket.

¹⁰ A kritérium formalizált alakja megmutatja, hogy a kedvező kimenetel bekövetkezésének adott valószínűsége mellett mekkora az optimális tét nagysága.

A piac ezen sajátossága alapján jelent meg a pénzügyekben már ismert statisztikai arbitrázs kifejezése is, amely automatikusan, nyilvános információkra épített modellekkel kihasználható arbitrázslehetőséget jelent. A sportfogadópiacok hatékonyságát komplex, nagy adatigényű módszerekkel tesztelő kutatók ilyen statisztikai arbitrázslehetőségeket próbáltak azonosítani, melyek mértékét és tartósságát a ráépülő fogadási stratégia pénzügyi teljesítménye - azaz hozam – kockázati mutatói – alapján határozták meg.

A labdarúgó mérkőzések jövőbeli kimeneteleit sokáig a kevésbé adatigényes matematikai modellekkel becsülték a kutatók, a nyilvánosságra kerülő adatok ugrásszerű növekedése azonban egy idő után lehetővé tette a sokkal pontosabb statisztikai modellek alkalmazását. Éppen ezért napjainkban a sportfogadóirodák becsléseit legjobban kihívás elé állító modellek a big data, a machine learning (gépi tanulás), illetve a neurális hálók módszertanaira épülnek. A futballmérkőzések statisztikai előrejelzésével foglalkozó releváns cikkeket és tanulmányokat az 1. táblázat foglalja össze.

1. Táblázat: A futballmérkőzések kimenetelét statisztikai módszertanok segítségével előrejelző releváns kutatások összefoglalása. Forrás: Saját szerkesztés.

Hivatkozás	Alkalmazott modell(ek)	Magyarázó változók	Elért becslési pontosság
Knoll & Stübinger (2020)	RFO, SVM, XGB, OLS Regression	Korábbi mérkőzések csapat-statisztikái.	66 - 75%
Stübinger és mtsai. (2019)	RFO, SVM, XGB, OLS Regression	A játékosok egyéni teljesítménymutatói korábbi meccseken.	69 - 81%
Raju és mtsai. (2020)	SVM, LR, NB Classifier, DT Classifier, AB Classifier	Korábbi mérkőzések eredményei és gól-statisztikái.	69 - 76%
Tax & Yme Joustra (2015)	NB Classifier, MLP Classifier, RFO Classifier, ...	Korábbi mérkőzések eredményei és játékoskeretre, valamint a mérkőzésre vonatkozó külső változók (fáradtság, utazás, edzőváltás, ...).	52 - 55%
Chinwe Peace és mtsai. (2014)	ANN, LR	Korábbi mérkőzések eredményei és gól-statisztikái, csapatok támadó- és védekező képessége.	75 - 85%
Carloni és mtsai. (2021)	LR, SVM, NB Classifier, RFO Classifier, ANN, KNN	A fogadóirodák által kínált oddsok.	-

Pipatchatchawal & Phimoltares (2021)	SVM, GNB, KNN, RFO, GBO, MLP (mind Classifier)	Korábbi mérkőzések csapat- és játékos-statisztikái, illetve a FIFA-videojáték játékosértékei.	56,5 - 56,8%
Baboota & Kaur (2019)	SVB, GNB, GBO, RFO (mind Classifier)	Korábbi mérkőzések csapatstatisztikái, a mérkőzés körülményeit leíró külső változók.	55 - 60%
Beal és mtsai. (2021)	RFO Classifier	Korábbi mérkőzések statisztikái és szakértők nyilvánosan megjelent cikkei.	63,19%
Vaidya és mtsai. (2016)	LR, RFO Classifier, NB Classifier	A csapatok aktuális formája, támadó- és védekező-képessége.	47 - 50%
Rudrapal és mtsai. (2020)	SVM, RFO Classifier, GNB, MLP Classifier	A csapatok és a játékosok közelmúltbeli teljesítményének statisztikai mutatói, illetve a meccs körülményeit leíró külső változók.	59 - 74%
Bunker & Thabtah (2019)	NB Classifier, LogitBoost, ANN, RFO Classifier, ...	Nincs, mivel a tanulmány a futballmeccsek előrejelzésének elméleti keretrendszerét foglalja össze.	-
Khan & Kirubanand V. B (2019)	SVM, XGB Classifier	A csapatok formáját leíró fontosabb statisztikai mutatók, illetve a fogadóirodák által kínált oddsok.	65 és 80%

Jelmagyarázat: RFO: Random Forest, SVM: Support Vector Machine, XGB: XGBoost, OLS: Ordinary Least Squares (legkisebb négyzetek módszere), LR: Logisztikus regresszió, NBC: Naiv Bayes Classifier, ABC: AdaBoost Classifier, DTC: Decision Tree (döntési fa) Classifier, MLP: Multilayer Perceptron, ANN: Artificial Neural Networks (mesterséges neurális hálók), KNN: K-nearest Neighbors (k-legközelebbi szomszéd algoritmus), GNB: Gaussian Naiv Bayes, GBO: Gradient Boosting.

Az 1. táblázatban összegyűjtött tanulmányok nagy része (mely alól az egyetlen direkt kivételt Tax és Yme Justra (2015) holland bajnokságra fókuszáló cikke jelenti) az angol Premier League mérkőzéseinek kimenetelét becsülte különféle módszertanok segítségével. Mivel az angol első osztályú bajnokságot övezi a legnagyobb nemzetközi érdeklődés, a leglikvidebb sportfogadópia, és talán a legtöbb nyilvános információ, kutatásomban én is erre a ligára fókuszálok.

A felsorolt tanulmányok bár módszertanukat tekintve hasonlítanak egymáshoz, a kiválasztott magyarázó változók miatt gyakran mégis teljesen különböző struktúrájúak. A modellezés során felhasznált változók leggyakrabban a csapatok vagy a játékosok korábbi meccseinek statisztikái, a becsülni kívánt mérkőzés körülményei, a megfigyelhető oddsok, illetve egyénileg definiált faktorok közül kerültek ki. A predikciós problémát a szerzők általában vagy osztályozási (klasszifikációs) vagy regressziós feladatként definiálták, melyhez

aztán az alkalmazott módszertant is igazították¹¹. Folytonos jellegéből fakadóan a regressziós becslések tárgya általában a csapatok által várhatóan szerzett gólok száma vagy azok különbsége volt, míg az osztályozási modellek közvetlenül a kategorikus kimenetelt (hazai győzelem, döntetlen, vendégsiker) próbálták megragadni.

Bár a kutatók többféle módon is értékelték módszereik teljesítményét, a becslés pontosságát – melyet a sikeresen prediktált és az összes vizsgált mérkőzés számának hányadosaként definiáltak - általában mindenki kiszámította. Az eredményeket összehasonlítva megfigyelhető, hogy a klasszifikációs becslések pontossága elmarad a regressziós előrejelzésektől, melyet a lehetséges kimenetek nagyobb száma¹² mellett a döntetlenek megjósolásának nehézsége is magyarázhat.

Összességében megállapítható, hogy a sportfogadopiachoz kapcsolódó statisztikai arbitrázs jelensége, illetve a mérkőzések kimenetelének komplex statisztikai modellekkel történő előrejelzése számos kutatást eredményezett, a felhasználható információk minőségi és mennyiségi növekedése, a más sportágakhoz képest alacsony becslési pontosság (Beal és mtsai., 2021) illetve probléma komplexitása indokolja az előrejelzések további finomítását, valamint a piaci hatékonyság vizsgálatát.

¹¹ Itt fontos megjegyezni, hogy a machine learning módszertanok többsége klasszifikációs és regressziós probléma megoldására is használható.

¹² A regressziós előrejelzésekre épülő fogadási stratégiák csak kellően nagy abszolútértékű becsült gólkülönbség esetén javasoltak fogadást, mely így definíciójából adódóan kizárta a döntetlen kimenetelre helyezett tét lehetőségét.

III. Felhasznált adatok és a kutatás módszertana

III.1. Felhasznált adatok és leíró statisztika

Empirikus elemzésem alapjául az angol Premier League mérkőzései szolgáltak a 2018-2019-es szezon kezdetétől egészen az adatgyűjtés lezárultáig (2023. december 10.). Az időtáv kiválasztását első sorban az adatok elérhetősége határozta meg, a csapatok teljesítményét legjobban leíró mutatók (várható gólok, gólpasszok, haladó támadó- és védekező statisztikák) ugyanis csak 2018 augusztusától voltak nyilvánosan elérhetőek. A Premier League mérkőzéseire épülő likvid sportfogadópiac egy részről jó lehetőséget teremt a különféle stratégiák tesztelésére, más részről viszont a nagy számú tranzakció és koncentrált tudás nagy mértékben megnehezíti a statisztikai arbitrázslehetőségek azonosítását.

Az adatgyűjtés során összesen 206 forduló 2060 futballmeccsének statisztikáit gyűjtöttem össze mindenki számára elérhető webhelyekről. A csapatok teljesítménymutatóinak forrásául a www.fbref.com, a kezdőcsapatok piaci értékének forrásául a www.transfermarkt.com, míg a sérült és eltiltott játékosok forrásául a www.soccerway.com weboldal szolgált. Az adattisztítást követően a feldolgozott szakirodalom és a statisztikai mutatók közti korrelációs kapcsolatok feltérképezésének segítségével elvégeztem a változószelekciót. Knoll és Stübinger (2020), illetve Vaidya és szerzőtársai (2016) alapján a mérkőzések legfontosabb karakterisztikáit a támadások, a kapura lövések, a passzjáték, és a védekezés mennyiségi és minőségi mutatóival igyekeztem leképezni. A teljesítménymutatókat a hazai- és vendégcsapatokra vonatkozóan külön szerepeltettem az adatbázisban, leszámítva a hazai csapat által megnyert párharcok arányát, mely egyben meghatározta a vendégcsapat által megnyert párharcok hányadát is. Ezek után az adatbázist kiegészítettem még három külső magyarázó változóval, a csapatok elmúlt öt hazai/vendég mérkőzésen szerzett pontjainak átlagával (mely az együttesek aktuális formáját hivatott jelezni), a kezdőcsapatok piaci értékével, illetve a mérkőzésről eltiltás, vagy sérülés miatt hiányzó játékosok számával (Tax és Yme Joustra, 2015). Ez utóbbi két változó azért lehet hasznos a modellezés során, mert a kezdőcsapat jellemzésén keresztül beépíti az előrejelzésbe a pályára lépő csapatok aktuális állapotát. Egy esetleges sérüléshullám, rotálás vagy kulcsjátékosok hiánya ugyanis jelentősen képes lehet visszavetni egy csapat győzelmi esélyeit, de a fenti változók figyelembevételével igyekeztem ezt a torzítást enyhíteni.

A változószelekció elvégzése után az adatbázisban összesen 65, egymással gyengén, vagy közepesen korreláló magyarázó változó maradt. A mutatók közül egyedül a kezdőcsapatok

piaci értékén kellett monoton transzformációt (logaritmus) alkalmaznom, mivel csak ezen változó eloszlása volt jelentősen balra ferde, jobbra elnyúló (melyet az átlag és a medián hányadosának 1,2-t meghaladó értékével azonosítottam). A passzpontosság, a szerelési-, védelési- és cselezési hatékonyság, valamint a megnyert párharcok mértéke arányszám-jellegű mutató (0 minimum és 1 maximum értékkel), míg a többi statisztikai adat valós értékű volt.

III.2. A modellezés során használt magyarázó- és eredményváltozók, illetve az oddsok bemutatása

A változók főbb statisztikai mutatóit a 2. táblázat foglalja össze.

2. Táblázat: A modellezés során használt változók leíró statisztikai adatai.

Forrás: Saját szerkesztés.

	Hazai csapat				Vendégcsapat			
	Min	Medián	Átlag	Max	Min	Medián	Átlag	Max
Támadójáték								
Kapura lövések	0	13	13,6	36	1	11	11,3	30
Kaput eltalált lövések	0	4	4,5	15	0	4	3,8	15
xG büntetők nélkül	0,0	1,3	1,4	5,1	0,0	1,0	1,2	4,9
xAG	0,0	0,9	1,1	4,4	0,0	0,8	0,9	4,4
Lövést előkészítő akciók	0	23	24,0	59	0	19	20,0	58
Gólhelyzetet előkészítő akciók száma	0	2	2,5	15	0	2	2,2	16
Szögletek száma	0	5	5,7	17	0	4	4,7	19
Labdaérintések az ellen-fél büntetőterületén belül	2	23	24,4	80	0	19	20,4	68
Sikeres cselek aránya	0,0%	54,5%	54,3%	100,0%	0,0%	53,8%	53,7%	100,0%
Passzjáték								
Sikeres passzok száma	84	393	407,2	902	93	373,5	386,8	893
Passzpontosság	45,9%	78,3%	77,3%	92,3%	50,4%	77,9%	76,7%	92,3%
Progresszív passzok száma	5	38	40,4	118	5	34	35,9	113
Progresszív labdavezetések száma	2	18	19,2	68	1	16	17,1	53
xA	0	0,8	0,9	4	0	0,7	0,8	3,5
Kulcspasszok száma	0	10	10,1	26	0	8	8,5	26
Passzok a büntetőterületen belülre	0	8	8,7	32	0	7	7,3	28
Beívelések száma	0	5	5,1	22	0	4	4,5	19
Pontos beívelések száma	0	2	2,1	12	0	1	1,6	10
Kiugratások	0	1	1,4	15	0	1	1,3	12

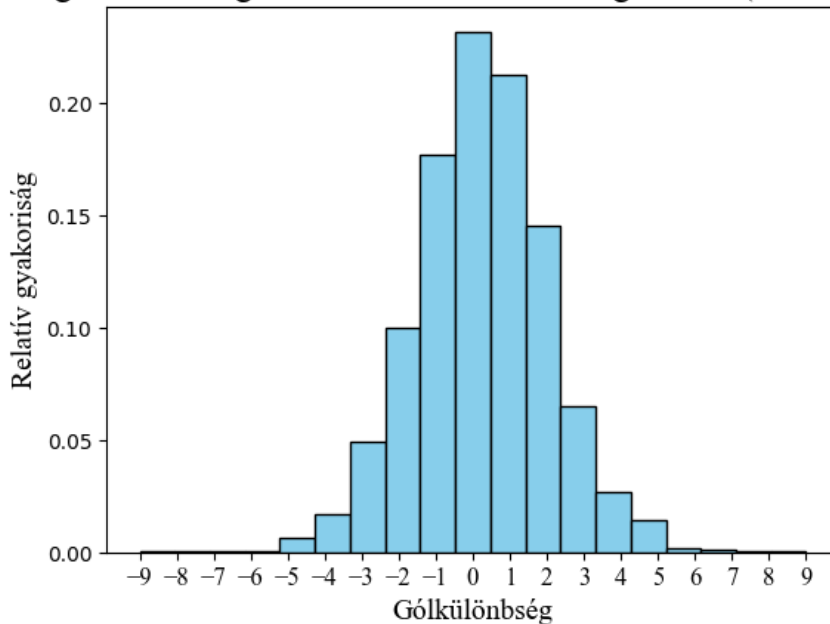
Védekező akciók								
Szerelések száma	3	16	16,4	36	3	16	16,6	46
Szerelési hatékonyság	0,0%	44,4%	44,6%	100,0%	0,0%	43,8%	43,9%	100,0%
Közbeavatkozások száma	1	10	9,9	31	0	10	10,1	33
Blokkolt lövések száma	0	3	3,1	15	0	3	3,8	15
Tisztázások száma	1	17	18,9	59	1	21	22,4	80
Szabálytalanságok száma	0	10	10,5	23	1	10	10,7	25
Sárga lapok	0	2	1,7	7	0	2	1,8	7
Piros lapok	0	0	0,1	2	0	0	0,1	2
Labdaszerzések száma	23	54	53,7	87	24	52	52,4	98
Védési hatékonyság	0,0%	66,7%	67,5%	100,0%	0,0%	90,0%	88,1%	100,0%
Megnyert párharcok aránya	11,0%	50,0%	50,7%	92,0%	8,0%	50,0%	49,3%	89,0%
Általános statisztikák								
Kezdőcsapat piaci értékének logaritmus	3,5	5,3	5,3	6,9	3,1	5,3	5,3	6,9
Sérült és eltűnt játékosok száma	0	4	3,9	11	0	4	4,0	12
Forma	0	1,5	1,6	3	0	1,2	1,2	3

A táblázat adatai megerősítik a korábban már számos empirikus kutatásban is igazolt hazai pálya előnye¹³ - jelenséget, hiszen a hazai pályán játszó csapatok átlagos teljesítménymutatói minden támadó- és passzstatisztikában felülmúlják az idegenben szereplőket. A védekezést jellemző mutatóknál természetesen ellentétes dinamika figyelhető meg, ami szintén alátámasztja azt a vélekedést, hogy egy csapat hazai pályán dominánsabban és támadóbb szellemben lép pályára, mint idegenben. A jelenség legnyilvánvalóbb bizonyítéka mégis az általános statisztikai adatok közt található: az elmúlt öt hazai pályán, illetve vendégségben lejátszott mérkőzés pontátlagában a 0,4 pontnyi különbség minden szignifikancia szinten jelentősnek mondható.

Empirikus elemzésem központjában két mutatószám, a mérkőzés kimenetele (mely lehet hazai győzelem, vendég siker vagy döntetlen), illetve a gólkülönbség állt, melynek eloszlását a 2. ábra mutatja be.

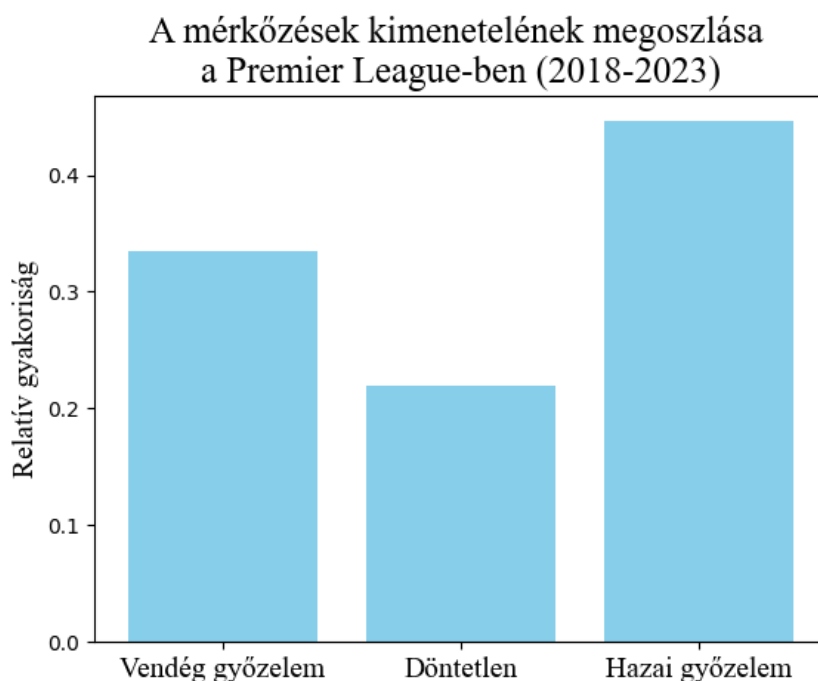
¹³ A hazai pálya előnyének jelensége – melyet dolgozatában Kovács (2022) részletesen vizsgált – a saját pályáján játszó csapatok ceteris paribus nagyobb győzelmi esélyére, illetve a vendégként játszott meccseknél konzisztensen jobb teljesítményre utal. A közönség támogatása, kevesebb utazás, esetleges játéktekintői részrehajlás, illetve a megszokott körülmények mellett számos további faktor magyarázhatja a jelenséget.

A gólkülönbség eloszlása a Premier League-ben (2018-2023)



2. Ábra: A gólkülönbség eloszlása. Forrás: Saját szerkesztés.

A gólkülönbség hisztogramja szintén megerősíti a hazai pálya előnyét, hiszen az eloszlás jobbra elnyúló, + 0,54-es ferdeségi mutatóval. Ezzel teljesen összhangban van az egyes kimenetek relatív gyakorisága is: a 2060 mérkőzést tartalmazó mintában összesen 919 hazai (44,61%) és 689 vendég győzelmet (33,44%) figyeltem meg 452 döntetlen mellett (21,94%) (lásd 3. ábra). Mivel kutatásom fő célja a jövőbeli futballmérkőzések kimenetelének előrejelzése volt, a modellezés során kezelni kellett a gólkülönbség aszimmetrikus eloszlásának, illetve a hazai pálya előnyének kérdését. Az eredmények szempontjából fontos megjegyezni, hogy munkám során kezeltem célváltozóként a gólkülönbséget, illetve a kimenetelt is, melyek azonban egymással teljesen ekvivalenssé tehető információt közvetítenek, így a kétféle típusú előrejelző modellek becsléseit a fogadási stratégiák definiálásakor lehetséges volt kombinálni egymással.



3. Ábra: A Premier League mérkőzéseinek kimenetele (2018-2023). Forrás: Saját szerkesztés.

A fogadási stratégiák kidolgozásához és idősoros teszteléséhez a Bet365 online fogadásközvetítő társaság kimenetelekre (hazai győzelem, döntetlen, vendégsiker) vonatkozó hivatalos oddsait használtam, melynek historikus adatait a www.football-data.co.uk oldalról töltöttem le. Az oddsok leíró statisztikáit az alább látható 3. táblázat foglalja össze. A hazai- és vendéggyőzelmekhez kapcsolódó oddsok közti eltérést a hazai pálya torzító hatása okozza, míg a döntetlenek magas oddsait az alacsony relatív frekvenciájuk (az európai topbajnokságokban az eldöntetlen mérkőzések aránya 20 és 30% körül mozog – [sofascore.com](https://www.sofascore.com), 2024) okozza. Az oddsok korábban vázolt értelmezése alapján a döntetleneknél előforduló 4,20-as átlagos érték nagyjából 23,4%-os valószínűségnek felel meg, ami nem is áll messze a tapasztalt historikus gyakoriságtól (az általam vett mintában a döntetlenek aránya 21,9% volt).

3. Táblázat: A vizsgált mérkőzések oddsainak fő statisztikai jellemzői.

Forrás: Saját szerkesztés.

	Min	Medián	Átlag	Max
Hazai győzelem	1,06	2,25	3,03	23,00
Döntetlen	2,80	3,75	4,28	17,00
Vendéggyőzelem	1,12	3,25	4,70	41,00

A www.bet365.com 2024-ben a világ második legnagyobb forgalmú online sportfogadást kínáló weboldalaként (similarweb.com, 2024) számtalan sportágban és kategóriában kínál fogadásokat, így az általuk meghatározott oddsok vélhetően pontosan tükrözik a piac várakozásait. A letöltött adatok a mérkőzés kezdete előtti utolsó hivatalos oddsokat tartalmazzák, így tükröztek már minden kezdés előtt nyilvánossá váló információt (kezdőcsapat, sérültek, eltiltottak, a csapatok formája, időjárási körülmények, ...).

III.3. Az előrejelző modellek kiválasztása, tanítása és tesztelése

Kutatásom célja az angol első osztályú labdarúgó mérkőzések kimenetelének előrejelzése volt machine learning modellek alkalmazásával, illetve sportfogadási stratégiák kidolgozása az eredmények alapján. Az előrejelzést kétféle módszerrel végeztem el: a gólkülönbséget regressziókkal modelleztem, míg a kimeneteket közvetlenül úgynevezett klasszifikációs eljárással vizsgáltam. Mivel a magyarázó változók között nem tapasztaltam erős korrelációs kapcsolatot, nem alkalmaztam semmilyen dimenziócsökkentő eljárást. Ugyan a faktorelemzés, vagy a főkomponenselemzés elő szokott fordulni hasonló modellezési feladatoknál, megfelelő változószelekcióval sikerült elkerülnöm a multikollinearitást a modellben.

A manuális adatelőkészítés után a modellezést a Python-programnyelvet alkalmazva¹⁴ végeztem el. Az adatok előkészítését követően úgy döntöttem, hogy a releváns dolgozatokban olvasottak alapján a feladatot osztályozási (klasszifikációs) és regressziós problémaként felfogva mindkét módon megpróbálom előrejelezni a mérkőzések kimenetelét, majd az ezekre épített fogadási stratégiák tesztelésével összehasonlítom a modellek teljesítményét. A két módszer közti fő különbséget az becslt eredményváltozó jelentette, hiszen míg klasszifikációs eljárás esetén a három lehetséges kimenetel (hazai győzelem, döntetlen, vendégsiker) egyikébe próbáltam besorolni a különböző karakterisztikájú mérkőzéseket, addig a regressziós becslés függő változója a mérkőzés várható előjeles gólkülönbsége volt.

A II.3. fejezetben bemutatott szakirodalom alapján a becsléshez két – a futballmérkőzések előrejelzéséhez jól illeszkedő – gépi tanulásra (machine learning), illetve egy neurális hálókra alapozó módszert választottam: a Random Forest (Véletlen erdő, a későbbiekben gyakran RFO), az Extreme Gradient Boosting (a későbbiekben gyakran XGB),

¹⁴ A folyamat során a pandas, a numpy, a skopt, az sklearn.preprocessing, az sklearn.model_selection, az sklearn.ensemble, az sklearn.neural_network és az xgboost, a vizualizáláshoz pedig a matplotlib.pyplot modulokat használtam.

illetve a Multilayer Perceptron (a későbbiekben gyakran MLP) algoritmust. Alkalmasságuk miatt mindhárom modellt alkalmaztam a klasszifikációs és a regressziós becslés során is.

A Random Forest algoritmus a döntési fák módszerére épül, azok átlagos előrejelzésének használatával pedig képes volt optimalizálni a pontosságot, de mégis elkerülni a túltanulásból eredő előrejelzési hibákat. A döntési fák kialakításához szükséges részhalmazok kiválasztása véletlenszerűen történik, míg a becslések összefésülése átlagszámítással vagy a leggyakoribb érték kiválasztásával. Az algoritmus legfontosabb paraméterei a döntési fák maximális mélysége, a minimális mintaszám, amely a döntési csomópontok további szétosztását eredményezi, a tovább nem bontott részhalmazok minimális mintaszáma, illetve a modellezés során létrehozott döntési fák száma (Rigatti, 2017).

Az Extreme Gradient Boosting (XGBoost) módszer kezdetleges előrejelzésekből kiindulva iteratív módon csökkenti a számított becslési hibát. Az algoritmus az egyes lépésekben döntési fák hozzáadásával csökkenti a maradék hibát, a túltanulás elkerülése érdekében pedig különböző regularizációs technikákat használ. A fő paraméterek közé a tanulási ráta (learning rate), a döntési fák maximális mélysége, az egyes csomópontokban lévő elemek minimális száma, illetve az iterációs lépésekben az adatok felhasználását befolyásoló mintavételi paraméter tartozik (James és mtsai., 2023).

A neurális hálókra alapuló Multilayer Perceptron algoritmus egy több rétegű, bonyolult struktúrát feltételez a magyarázó- és eredményváltozók között. A modellek minden esetben tartalmaznak egy bemeneti- és egy kimeneti réteget, köztük azonban egy, vagy akár több rejtett réteg is meghúzódhat. Az egyes rétegek különböző számú neuronból állnak, melyek nemlineáris aktivációs függvényekkel vannak összekapcsolva. A neuronok rendelkeznek egy meghatározott súllyal, illetve torzítás-értékkel, bemeneti adataikat pedig a megelőző réteg neuronjainak kimeneti információi alapján nyerik. A modell optimalizálása úgy zajlik, hogy az input adatok végighaladnak a hálózattól a bemeneti változóktól egészen az outputig, majd az előrejelzések kiértékelése után az algoritmus frissíti a paraméterbeállításokat. Az így leírt iterációs folyamattal a modell egy előre definiált hibafüggvény minimumértékét igyekszik megtalálni, mely a struktúra komplexitása miatt azonban csak néha sikerül. Túl lassú konvergencia esetén a rendszer kimeneti változó előrejelzése során a paraméterként megadott maximális iterációs szám elérésekor érvényes beállításokat alkalmazza. Az algoritmus legfontosabb paraméterei a hálózat szerkezetéhez (rétegek és neuronok száma), a beállítások frissítéséhez (tanulási ráta, regularizációs technikák), illetve a neuronok közti

információáramlást biztosító aktivációs függvényhez kapcsolódnak (Popescu és szerzőtársai, 2009).

Empirikus elemzésem során épített modellek célja a futballmérkőzések kimenetelének előrejelzése volt főként mérkőzésen belüli statisztikák, illetve három külső változó segítségével. A modellezés klasszifikáció és regresszió esetén is három fázisból állt: a machine learning- és neurális háló-modellek hiperparamétereinek optimalizálásából, az előrejelzések elvégzéséből és végül a fogadási stratégiák kidolgozásából. A hiperparaméter-kereséshez először két egyenlő méretű részre (1030-1030 mérkőzés), tanító- és validáló halmazra osztottam az adatbázist. Erre azért volt szükség, hogy elkerüljem a modell túltanulását és biztosítsam, hogy a végső hiperparaméterek nemcsak a tanító-, hanem a validáló halmazon is jó teljesítményt érjenek el. A hiperparaméter-optimalizálást a Random Forest- és XGBoost-modellek esetében a Knoll & Stübinger (2020) által beállított értékekkel kezdtem, majd először Grid Search-öt, végül utolsó lépésben Bayes-i keresést alkalmaztam. A két keresőalgoritmus a hiperparaméterek előre megadott értékeinek összes kombinációját kiválasztotta, majd illesztette a tanító adatbázisra, végül kiértékelte validáló adatbázison. A folyamat végén a program azt a paraméter-kombinációt javasolta, amely a legmagasabb pontosságot érte el a validációs halmazon. A keresést először minden esetben a diszkrét értékekkel dolgozó Grid Search-csel kezdtem, majd miután kellően leszűkítettem a paraméterek optimális értékeit, utolsó lépésben az előre megadott eloszlás alapján modelleket illesztő Bayes-i algoritmust alkalmaztam. Utóbbi eljárást természetesen csak a folytonos jellegű paraméterek esetében alkalmaztam, melyek közé a Learning_rate, a Subsample és a Colsample_bytree tartoztak. Az algoritmusok négy-ötöszi futtatása minden modell esetében optimális klasszifikációs teljesítményt/Root Mean Squared Error-t eredményezett a túltanulás elkerülése mellett. Az egyes módszerek optimális hiperparaméter-kombinációit a következő oldalon látható 4. táblázat foglalja össze.

A paraméterek optimalizálása során figyelembe kellett venni a futballmérkőzések előrejelzésének komplexitását (rengeteg lehetséges magyarázó változó, kevés egyértelmű indikátor, magas véletlen-faktor), valamint a túltanulás kockázatát, így végül többnyire átlagosnak mondható értékek mellett döntöttem, néhány kivétellel. A kivételek közé tartozik az XGBoost Regressor magas Gamma-értéke (amely a döntési fák növekedését szabályozza), illetve a Random Forest Classifier magas Min_samples_split paramétere (amely döntési csomópontok szétválasztásához szükséges minimális mintaszámot határozza meg). Ugyan a fenti kivételek növelték a túltanulás kockázatát, úgy vélem, a validáló halmaz magas elemszáma képes volt ezen hatásokat kompenzálni.

4. Táblázat: A modellezés során használt hiperparaméterek optimális értékei.

Forrás: Saját szerkesztés.

	RFO Regressor	XGB Regressor	RFO Classifier	XGB Classifier
Max_depth	8	6	6	4
Max_features	Sqrt	-	Sqrt	-
Min_samples_leaf	2	-	3	-
Min_samples_split	2	-	9	-
N_estimators	400	-	225	-
Gamma	-	3,4	-	0,36
Learning_rate	-	0,05	-	0,05
Min_child_weight	-	5	-	2
Subsample	-	0,91	-	0,89
Colsample_bytree	-	0,65	-	0,75
	Multilayer Perceptron Regressor		Multilayer Perceptron Classifier	
Activation function	Tahn		Tahn	
Alpha	0,0001		0,0001	
Learning_rate	Adaptive		Adaptive	
Solver	Stochastic Gradient Descent		Stochastic Gradient Descent	
Hidden_layer_size	(200,)		(200,)	
Batch_size	-		256	
Max_iteartions	200		200	

A Multilayer Perceptronok esetében egyértelmű volt a grádiensereszkedő algoritmus, a tangens hiperbolikus aktivációs függvény és a dinamikusan alkalmazkodó learning rate használata (alternatív megoldások jelentősen rosszabb teljesítményt eredményeztek), míg a regularizáció alpha értékét, valamint a rejtett réteg(ek) méretét iterációs közelítéssel optimalizáltam.

Összességében kijelenthető, hogy a beállított hiperparaméterek a túltanulás elkerülése mellett optimális modell-teljesítményhez vezettek a validáló adatbázisokon, valamint illeszkednek a korábbi releváns kutatások során használt beállításokhoz is (Knoll & Stübinger, 2020).

III.4. Az előrejelző modellek felépítése

Az empirikus elemzés során alkalmazott klasszifikációs és regressziós modellek magyarázó változói a csapatok védekező-, támadó- és passzstatisztikái, valamint a formájukat és kezdőcsapatukat meghatározó külső változókból tevődtek össze. A két előrejelző módszer között különbség egyedül az eredményváltozóban mutatkozott, hiszen míg a regressziók tárgya

előjeles gólkülönbség volt, addig klasszifikáció során direktben a kimeneteleket próbáltam meghatározni¹⁵.

A modellezés során az egyik fő kihívást az aktuális mérkőzés előtt ismeretlen statisztikai mutatók becslése jelentette. Knoll & Stübinger (2020) alapján a mérkőzés kezdete előtt nem ismert magyarázó változókat az érintett csapatok elmúlt három mérkőzésének számtani átlagaként határoztam meg. Mivel a csapatok formáját, a kezdőjátékosok piaci értékét és a hiányzók számát a kezdőrúgás előtt megismerheti a fogadó, ezen információkat én is adottnak vettem a modellekben. Figyelembe véve a csapatok eltérő teljesítményét hazai pályán és idegenben, a statisztikai mutatókat a pályaválasztói jognak megfelelően becsültem meg (azaz például egy vendégként játszó csapat statisztikáit a megelőző 3 vendégként lejátszott meccs teljesítményeinek átlagaként jeleztem előre). Ennek megfelelően a vizsgált szezonokban az 1-6. forduló mérkőzéseinek statisztikáit csak tanításra használtam, előrejelzést és fogadási stratégiákat a későbbi fordulók találkozóiira alapoztam. Ezen elv alól kivételt képzett a 2020-2021-es szezon, ahol csak a 8. fordulótól kezdtem az előrejelzést köszönhetően az első fordulóból elhalasztott több mérkőzésnek. Az adatbázis ily módon történő felosztásával sikerült lekövetni a csapatok keretének szezonok közti esetleges átalakulását, illetve elegendő információt biztosítani a modelleknek a frissen feljutó csapatok mérkőzéseinek előrejelzéséhez. Ezen logika alkalmazása végül 1372 mérkőzés kimenetelének becslését tette lehetővé.

Hogy a modellek tanítása és alkalmazása során is azonos mértékű bizonytalanságot biztosítsak (mely főként a magyarázó változók becslésének nehézségéből fakadt), mindkét esetben az előrejelzéshez összeállított – korábbi meccsek átlagos statisztikai adatait tartalmazó – adatbázist használtam. A becslések robusztusságának érdekében az első - 2018-2019-es – szezon 380 mérkőzését csak tanító célra vettem igénybe, a fogadási stratégiák kidolgozására a későbbi szezonok 7-38. fordulóiban lejátszott mérkőzések becsült kimenetelei alapján került sor. A modellek pontosságának növelése érdekében a tanító adatbázist mérkőzésről mérkőzésre bővítettem, hogy ezzel is növeljem az elérhető információ mennyiségét. A tanítás és a kimenetek prediktálása az n . mérkőzés esetében a következő lépésekben zajlott:

1. Optimálisan parametrizált modell újraillesztése az első $(n-1)$ mérkőzésre.
2. Az n . mérkőzés jövőbeli kimenetelének előrejelzése a modell alapján.
3. Az előrejelzett kimenetel/gólkülönbség elmentése egy erre létrehozott változóban.

¹⁵ Természetesen a fogadási stratégiák kidolgozásakor a gólkülönbség könnyen transzformálható volt kimenetellé, ráadásul a becslés mértéke a kimenetel valószínűségéről is árult el információt.

A fentiek alapján végrehajtott becslési stratégiával sikeresen tudtam egyeztetni az oddsok és a modellek által tükrözött információ mennyiségét, figyelembe venni a hazai pálya előnyét, valamint elkerültem a jövőbe tekintésből származó esetleges torzító hatásokat (look ahead bias). Összességében tehát a Random Forest, az XGBoost és a Multilayer Perceptron módszertanok klasszifikációs és regressziós előrejelzésre történő alkalmazásával hat különböző becsült kimenetelt generáltam, melyek kombinációiból a következő fejezetben ismertetésre kerülő fogadási stratégiák felépültek.

III.5. A sportfogadási stratégiák kidolgozásának és tesztelésének módszertana

A III.2. fejezetben bemutatott modellekkel összesen hat különböző előrejelzést sikerült adnom az angol Premier League 1372 mérkőzésére, melyek kombinálásával utolsó lépésben különböző sportfogadási stratégiákat állítottam össze. A modellezés és az eredményekre épített stratégiák fő célja a sportfogadópiacon megfigyelhető esetleges statisztikai arbitrázslehetőségek azonosítása és kihasználása volt. A stratégiák kidolgozását azok statisztikai és pénzügyi értékelése követte, mely alapján lehetőség nyílt következtetések levonására a piac hatékonyságával, illetve a jövőbeli kimenetekhez kapcsolódó általános bizonytalanság korlátaival kapcsolatban. Hogy a mérkőzések és a piacok természetéről minél pontosabb képet kapjak, számos fogadási stratégiát kidolgoztam és teszteltem, melyek az alapvető referencia-stratégiákat leszámítva a klasszifikációs előrejelzések kombinálásával, illetve a regressziós modellek becsléseinek küszöbértékek alapján történő kategorizálása alapján álltak elő.

Az elemzett sportfogadó stratégiákat a következő módon definiáltam:

- „*Hazai*”: Referenciastratégia, minden mérkőzésen 1 egységnyi tétet tesz a hazai pályán pályára lépő csapatra.
- „*Legalacsonyabb odds*”: Referenciastratégia, minden mérkőzésen 1 egységnyi tétet helyez a legalacsonyabb odds-szal rendelkező kimenetelre (amely az oddsok természete alapján vagy a hazai-, vagy a vendéggyőzelem volt).
- „*RFO & XGB & MLP*”: A stratégia minden olyan mérkőzésen, amelyen azonos kimenetelt jelzett előre a három klasszifikációs modell, 1 egységnyi tétet tesz a közös kimenetelre.
- „ $|GK| > 1$ ”: A stratégia a gólkülönbség regressziós előrejelzésére épít és minden olyan esetben, amikor a három modell (RFO, XGB és MLP) által becsült gólkülönbség abszolút értékének átlaga meghaladja az 1-et, 1 egységnyi tétet helyez a becsült gólkülönbség

előjelének megfelelő kimenetelre. Azaz negatív prediktált gólkülönbség esetében vendégsikerre, míg pozitív becslés esetén hazai sikerre.

- „ $|GK| > 1,5$ ”: A „ $|GK| > 1$ ” stratégia szigorúbb változata. Amennyiben a három modell által becsült gólkülönbség abszolút értékének átlaga meghaladja az 1,5-öt, 1 egységnyi tétet tesz a becsült gólkülönbség előjelének megfelelő kimenetelre¹⁶.
- „*MLP Regressor*: $|GK| > 1,5$ ”: Egyedi regressziós modell előrejelzésére épülő stratégia, amely minden olyan esetben, amikor a Multilayer Perceptron által becsült gólkülönbség átlaga meghaladja az 1,5-öt, a stratégia 1 egységnyi tétet helyez az előrejelzett kimenetelre¹⁷.
- „ $|GK| > 1$ & 3 *Classifiers*”: A legkomplexebb stratégiák egyike, amely kombinálja a regressziós modellek átlagos gólkülönbség-becslését és a klasszifikációs modellek előrejelzését. Minden olyan esetben, amikor az összes klasszifikációs előrejelzés egybeesik az átlagosan becsült gólkülönbség előjelével, és az átlagosan becsült gólkülönbség meghaladja az 1-et, a stratégia 1 egységnyi tétet helyez a prediktált kimenetre¹⁸.
- „ $|GK| > 1,5$ & 3 *Classifiers*”: A „ $|GK| > 1$ & 3 *Classifiers*” módosított változata. Minden olyan esetben, amikor az összes klasszifikációs előrejelzés egybeesik az átlagosan becsült gólkülönbség előjelével, és az átlagosan becsült gólkülönbség meghaladja az 1,5-öt, a stratégia 1 egységnyi tétet helyez a prediktált kimenetre.

A stratégiák definíciójából adódóan a legnagyobb számú tétet a két referenciastratégia helyezte el, hiszen mind a „Hazai”, mind a „Legalacsonyabb odds” aktívan fogadott az összes (1372) mérkőzésre, míg a machine learning modellek előrejelzésére épülő stratégiák igyekeztek azokat a mérkőzéseket kiválasztani, ahol relatíve komoly esély van a prediktált kimenetel bekövetkezésére. Mivel a hasonló releváns kutatásokkal összhangban (Tax & Yme Joustra, 2015, Baboota & Kaur, 2019 és Pipatchatchawal & Phimoltares 2021) a klasszifikációs céllal felépített modellek csak relatíve alacsony - 55-60%-os - pontosságot voltak képesek elérni, előrejelzéseikre nem építettem külön stratégiákat.

¹⁶ Természetesen az 1 és 1,5-ös küszöbértékek mellett számos alsó határral rendelkező stratégiát is elemzésnek vettem alá, de gyengébb eredményeik miatt nem részletezem őket jelen dolgozatban.

¹⁷ Az átlagos gólkülönbség-becslést természetesen Random Forest és XGBoost regressziókkal is elvégeztem, de hasonlóságuk miatt csak a legjobban teljesítő MPL - variációt mutatom be jelen dolgozatban.

¹⁸ A minimum 1,5 gólos átlagos gólkülönbség-becslést kombináltam legalább két klasszifikációs modell azonos irányú előrejelzésével is, de az erre épülő stratégiát nem mutatom be külön, mivel teljesen megegyezik a „ $|GK| > 1,5$ & 3 *Classifiers*” stratégiával.

A stratégiák tesztelésére a 2019-2023-as időszak – korábban említett kivételtől eltekintve - 7-38. fordulókban játszott összesen 1372 mérkőzését használtam. A vizsgált mérkőzések kimenetelének eloszlása és a hozzájuk tartozó oddsok főbb leíró statisztikai mutatói alább, az 5. táblázatban láthatóak.

5. Táblázat: A fogadási stratégiák teszteléséhez használt mérkőzések kimenetelének megoszlása, illetve a kapcsolódó oddsok főbb statisztikai jellemzői. Forrás: Saját szerkesztés.

Kimenetel	Relatív gyakoriság
Hazai győzelem	44,46%
Döntetlen	22,30%
Vendég győzelem	33,24%

Oddsokhoz tartozó kimenetel	Min	Medián	Átlag	Max
Hazai győzelem	1,07	2,25	2,97	19
Döntetlen	2,80	3,75	4,20	13
Vendéggyőzelem	1,14	3,20	4,45	26

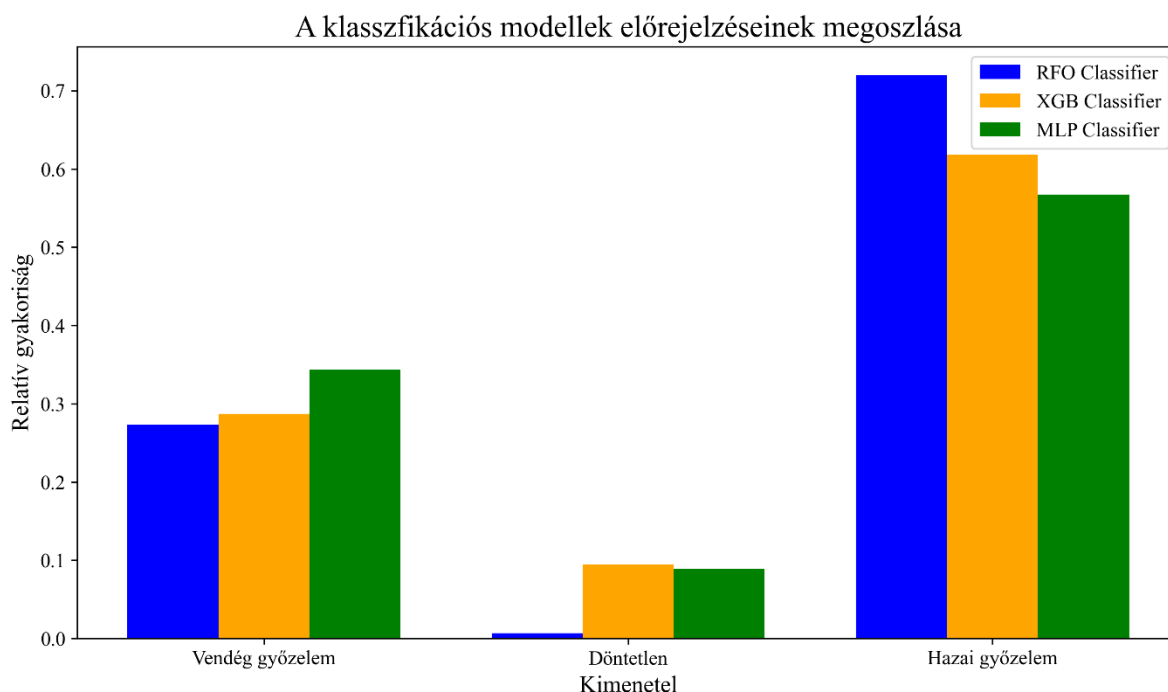
Az adatbázis szűkítése szinte semmilyen mértékben nem befolyásolta a kimenetek megoszlását, az oddsok azonban szűkebb intervallumokban szóródnak, mint a kiindulási halmaz mérkőzéseire képest. A kevésbé extrém szélsőértékeket magyarázhatja a szezonkezdéshez kapcsolódó extra bizonytalanság elmaradása (a tesztelésre összeállított adatbázis sosem tartalmazott mérkőzést az első hat fordulóból), valamint a meccsek kiegyenlítettebbé válása is a 2018-2019-es szezonhoz képest, melynek mérkőzései már nem szerepelnek az új adatbázisban.

IV. A kutatás eredményei

IV.1. Az előrejelzések és a fogadási stratégiák statisztikai elemzése

A mérkőzések kimenetelét előrejelző modellekkel arra a kérdésre kerestem a választ, hogy létezik-e statisztikai arbitrázs az angol labdarúgó Premier League-re épülő sportfogadópiacokon. A kimenetek becslése során abból a feltevésből indultam ki, hogy nemcsak a csapatok általános (pl. keretérték) vagy aktuális állapota (hiányzók száma, közelmúlt mérkőzéseinek eredménye), hanem a korábbi meccsek statisztikai adatai is komoly hatással vannak a jövőbeli mérkőzések kimenetelére. A közelmúlt mérkőzéseire épülő modellezési stratégia mindezek mellett azt is igényli, hogy a csapatok teljesítménymutatói rövid távon stabilak legyenek, hiszen enélkül a statisztikai mutatók előrejelzése nem lenne lehetséges historikus adatok alapján¹⁹. Jelen fejezetben a mérkőzések előrejelzett kimeneteleire épített sportfogadási stratégiák elemzését mutatom be statisztikai és pénzügyi szempontból.

A fogadási metódusok elemzése előtt azonban érdemes rátekinteni a 4. ábrára és a 6. táblázatra, amely az egyes modellek által előrejelzett kimenetek megoszlását, illetve a becsült gólkülönbségek főbb statisztikai jellegzetességeit tartalmazzák.



4. Ábra: A klasszifikációs modellek előrejelzéseinek megoszlása. Forrás: Saját szerkesztés.

¹⁹ A csapatok jellemző játéktípusa indokolja az átlagos statisztikai mutatók stabilitásának feltételezését, mindazonáltal edzőváltás vagy a keret összetételét érintő változások könnyen tudnak törést eredményezni egy-egy csapat teljesítményében.

6. Táblázat: A regressziós modellek által becsült gólkülönbségek főbb statisztikai mutatói.

Forrás: Saját szerkesztés.

	RFO Regressor	XGB Regressor	MLP Regressor
Minimum	-2,03	-3,17	-3,26
1. Kvartilis	-0,11	-0,27	-0,41
Medián	0,23	0,28	0,30
Átlag	0,27	0,26	0,28
3. Kvartilis	0,67	0,83	0,94
Maximum	2,73	3,58	3,45
Szórás	0,70	0,92	1,04

A három klasszifikációs eljárás közül a Random Forest szinte semmilyen, de a másik két modell is csak nagyon alacsony arányban prediktált döntetlent, melyet jellemzően a hazai győzelmek becslésével kompenzáltak (a jósolt vendéggyőzelmek és a tényleges gyakoriság relatíve egybeesik). A valós gyakoriságokat a Multilayer Perceptron becslései közelítik meg legjobban, de még ez a módszer is valamelyest túlbecsülte a hazai győzelmek arányát. Összességében kijelenthető tehát, hogy a hazai pálya torzító hatása az előrejelző modellekben is erőteljesen megjelenik, és talán már túlzott mértékben is korrigálták becsléseiket a pályaválasztói jog alapján.

A három regressziós modell által becsült gólkülönbség mutatói közül az XGBoost és a Multilayer Perceptron paraméterei hasonlítanak egymásra, míg a Random Forest Regressor ennél sokkal konzervatívabb becsléseket adott. Ezt erősíti meg a becsült differencia szűkebb intervalluma, illetve nullához közelebbi középvértékei. A becsült gólkülönbségek szórása is ezt a tényt erősíti meg, amely a Multilayer Perceptron Regressor esetében majdnem eléri a Random Forest Regressor 150%-át. Folytonos természete miatt a becsült gólkülönbség önmagában nem alkalmas fogadási stratégiára (a folytonos gólkülönbség a döntetleneket képtelen bármilyen formában előrejelzni), de megfelelő küszöbértékek használatával képes a mérkőzések kimenetelét a klasszifikációs eljárásoknál sokkal pontosabban becsülni. A fejezet további része ezen sportfogadó stratégiák statisztikai sajátosságait mutatja be.

A sportfogadási stratégiák pontossága és főbb jellemzői a 7. táblázatban láthatóak.

7. Táblázat: A sportfogadóstratégiák pontossága és alapvető jellemzői.

Forrás: Saját szerkesztés.

	Hazai	Legalacsonyabb odds	RFO & XGB & MLP	GK > 1	GK > 1,5
Az előrejelzés pontossága	44,46%	55,10%	55,15%	67,54%	76,52%
A fogadás részletei					
Elhelyezett tétek száma	1372	1372	961	342	132
Hazai győzelemre tett tétek száma	1372	845	676	249	100
Vendég győzelemre tett tétek száma	0	527	281	93	32
Átlagos kifizetés	0,98	0,97	0,97	0,97	1,02
	MLP Regressor: GK > 1,5		 GK > 1 & 3 Classifiers	 GK > 1,5 & 3 Classifiers	
Az előrejelzés pontossága	69,30%		67,74%		76,52%
A fogadás részletei					
Elhelyezett tétek száma	228		341		132
Hazai győzelemre tett tétek száma	161		248		100
Vendég győzelemre tett tétek száma	67		93		32
Átlagos kifizetés	1,00		0,97		1,02

Amint az a táblázatban is jól látszik, hogy a referenciastratégiák és a klasszifikációs stratégia pontossága messze elmarad a regressziós modelleken alapuló fogadási módszerekétől. Mindenképpen érdemes részletesebben megvizsgálni a három klasszifikációs modellre épülő stratégia teljesítményét, hiszen pontosságban csak kevéssel teljesítette túl a mögöttes elemzést egyáltalán nem igénylő „Legalacsonyabb odds” - módszert. Az 55,15%-os pontosság hátterében a labdarúgásban jellemző magas bizonytalansági faktor állhat, amely kiegyenlített csapatok mérkőzésein gyakran játszik kulcsszerepet. A tény, hogy a klasszifikációs modellek csak négy esetben jósoltak egyhangúan döntetlent – ráadásul 0%-os pontossággal -, jól jelzi a pontosztokodás előrejelzésének nehézségét. Ettől függetlenül a klasszifikációs modellek teljesítménye nem tekinthető rossznak, hiszen a rá épülő stratégia pontosságban megközelítette

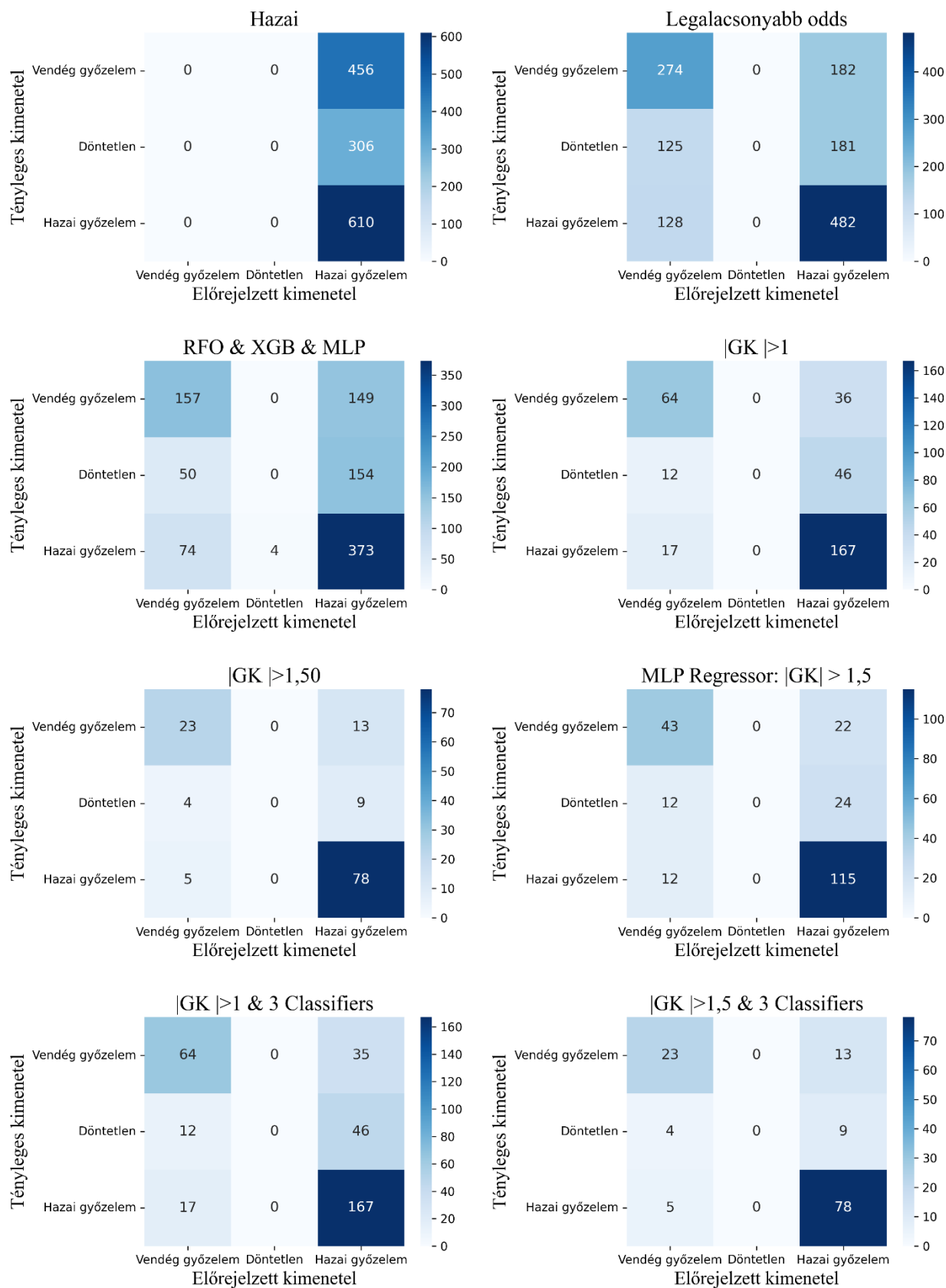
a hasonló kutatásokban épített legjobb modellek teljesítményét (lásd 1. táblázat). A kiegyenlített párharcok magas bizonytalansága indokolja a becsült gólkülönbségre épülő stratégiák definiálását is, melyek csak küszöbérték alatti/feletti előrejelzés esetén javasoltak fogadást, jelentő valószínűséget biztosítva ezáltal a kimenetel eltalálására. A döntetlennel kapcsolatos várakozásokat jól jelzi az oddsok természete is, hiszen a vizsgált időszakban a bukmékerek sosem tekintették a döntetlent a legvalószínűbb kimenetelnek (azaz az ehhez társított oddsnál mindig volt alacsonyabb), így végül a „Legalacsonyabb odds” - stratégia sem tett sosem tétet döntetlenre.

A fogadások mennyisége és megoszlása tovább árnyalják a stratégiák közti különbségeket a pontosság tekintetében. Ahogy az a táblázatból is látszik, egyértelmű negatív irányú kapcsolat fedezhető fel az elhelyezett tétek száma és az előrejelzési pontosság között. Ez a tény szintén erősítheti azt a vélekedést, mely szerint a nyereséges sportfogadóstratégiák egyik kiemelten fontos képessége a fogadás számára alkalmas mérkőzések kiválasztása. Definíciójából fakadóan a két referenciastratégia generálta a legtöbb fogadást, de az már korántsem volt egyértelmű, hogy a legpontosabbnak bizonyuló „ $|GK| > 1,5$ ” és „ $|GK| > 1,5 \& 3 \text{ Classifiers}$ ” stratégia helyezi el a legkevesebb tétet.

Szintén érdekes összefüggés érhető tetten az átlagos kifizetés és a pontosság közt: az elemzés alapján megfigyeltem, hogy a magasabb pontosság magasabb átlagos kifizetéssel is párosul. A legalacsonyabb átlagos kifizetéseket a két referenciastratégia, a klasszifikációs stratégia, illetve az 1-es küszöbértéket alkalmazó stratégiák produkálták, míg a legmagasabb, 1 feletti átlagos kifizetést a másfél gólos küszöbértéket alkalmazó stratégiák érték el.

A stratégiák találati arányáról alkotott képét tovább árnyalják a konfúziós mátrixok, melyek a helyesen, illetve helytelenül prediktált kimeneteket tartalmazzák a valós kimenetek függvényében (5. ábra). A helyesen kategorizált mérkőzéseket a mátrixok főátlói tartalmazzák, míg a többi cella a különböző típusú hibákat mutatja. A tény, hogy a rosszul becsült kimenetek között nagyjából 50% a döntetlenek aránya, jól jelzi az előrejelzést övező nagy bizonytalanságot. A hazai pálya torzító hatása a mátrixokban is megfigyelhető, mely egyrészt a stratégiákban jelentős túlsúlyban lévő hazai sikerre helyezett tétekben, másrészt a téves fogadások szerkezetében érhető tetten. Míg ugyanis a helytelenül hazai sikerre tett tétek esetében a valós kimenetel inkább döntetlen lett, addig rosszul kategorizált vendéggyőzelmek esetében már a hazai sikerek vannak túlsúlyban.

A fogadási stratégiák elemzése alapján összességében kijelenthető, hogy relatíve pontos előrejelzésre csak a regressziós becsléseket megfelelő küszöbértékkel alkalmazó stratégiák voltak képesek, a klasszifikációs eljárások nem tudták érdemben felülmúlni a hazai vagy az esélyesebb csapatokra tétet helyező naiv stratégiákat. Az egyedi regressziós metódusok közül a „RFO Regressor: $|GK| > 1,5$ ” érte el a legjobb előrejelzési pontosságot, de magas kiválasztási aránya miatt pénzügyileg mégis az „MLP Regressor: $|GK| > 1,5$ ” bizonyult a legjobbnak. Nem meglepően az átlagos gólkülönbségbecslésre, valamint a három klasszifikációs modellre épülő stratégiák teljesítettek a legjobban, mind pontosságban, mind pedig pénzügyi szempontból. A fogadási metódusok pontosságának elemzése egyértelműen igazolja a machine learning – és neurális háló – alapú modellezés hozzáadott értékét, mely leginkább a fogadásra alkalmas mérkőzések kiválasztásában nyilvánult meg. A találati arány azonban önmagában nem elégséges mérőszám a stratégiák értékeléséhez, így a IV.2. fejezetben részletes pénzügyi elemzésnek vettem alá őket.



5. Ábra: A sportfogadó stratégiákhoz tartozó konfúziós mátrixok. Forrás: Saját szerkesztés.

IV.2. A fogadási stratégiák értékelése pénzügyi szempontból

Empirikus elemzésem legfontosabb eleme a kidolgozott fogadási stratégiák pénzügyi értékelése, melynek fókuszában a hozam és a kockázat vizsgálata állt. Az alább látható 8. táblázat a vizsgált nyolc fogadási módszer legfontosabb nettó hozam- és kockázati adatait tartalmazza.

8. Táblázat: A vizsgált fogadási stratégiák főbb sajátosságai hozam és kockázati szempontból a 2019-2023-as időszak mérközései alapján. Forrás: Saját szerkesztés.

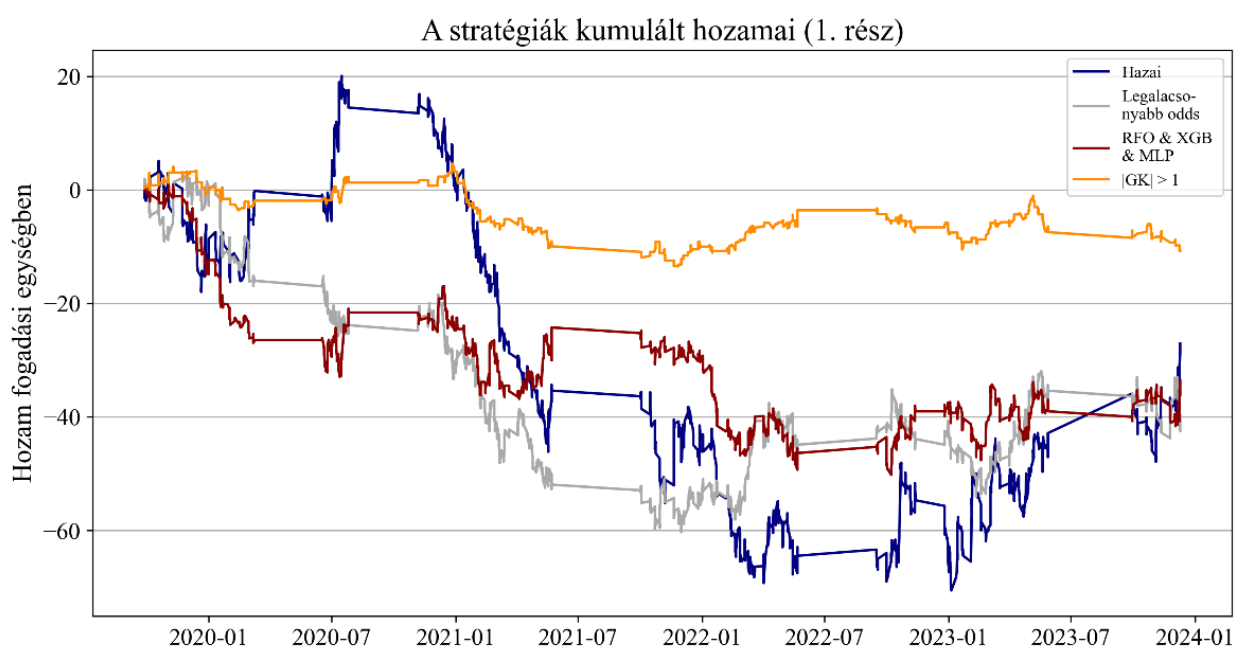
	Hazai	Legalacso- nyabb odds	RFO & XGB & MLP	GK > 1
Átlag	-0,0198	-0,0290	-0,0349	-0,0311
Minimum	-1,0000	-1,0000	-1,0000	-1,0000
1. Kvartilis	-1,0000	-1,0000	-1,0000	-1,0000
Medián	-1,0000	0,2200	0,1800	0,2200
3. Kvartilis	0,7200	0,7500	0,6600	0,4400
Maximum	10,0000	1,7500	6,0000	2,0000
Szórás	1,3683	0,9332	0,9814	0,7228
Csúcsosság	6,1382	-1,5636	0,7956	-1,0108
Ferdeség	1,9517	0,1487	0,6968	-0,2354
Nyertes fogadások aránya	0,4446	0,5510	0,5515	0,6754
	GK > 1,5	MLP Regressor: GK > 1,5	GK > 1 & 3 Classifiers	GK > 1,5 & 3 Classifiers
Átlag	0,0213	-0,0002	-0,0283	0,0213
Minimum	-1,0000	-1,0000	-1,0000	-1,0000
1. Kvartilis	0,0950	-1,0000	-1,0000	0,0950
Medián	0,2000	0,2000	0,2200	0,2000
3. Kvartilis	0,4000	0,4400	0,4400	0,4000
Maximum	1,2000	4,5000	2,0000	1,2000
Szórás	0,6009	0,7640	0,7219	0,6009
Csúcsosság	-0,5392	3,9418	-1,0012	-0,5392
Ferdeség	-0,8376	0,5967	-0,2417	-0,8376
Nyertes fogadások aránya	0,7652	0,6930	0,6774	0,7652

Összehasonlítva az egyes módszerek profitabilitását, világossá válik, hogy egyedül a machine learning modellekre épített regressziós előrejelzés eredményezett pozitív hozamú

fogadási stratégiát. A legalacsonyabb hozamot meglepő módon nem a referenciastratégiák, hanem a három klasszifikációs modell együttes előrejelzésére épülő fogadási módszer produkálta, mérközésenként -3,49%-kal, melyet sorrendben a „ $|GK| > 1$ ” és a „Legalacsonyabb odds” stratégiák követtek -3,11, illetve -2,90%-kal. A klasszifikációs modellek gyengébb teljesítményét elsősorban a kiegyenlített mérközések körüli jelentős bizonytalanság okozhatta, míg a „ $|GK| > 1$ ” esetében túl alacsony volt a fogadásra alkalmas mérközések kiválasztási küszöbértéke. Éppen ezért várakozásaimnak megfelelően pozitív hozamot egyedül a kellően magas küszöbértéket alkalmazó regressziós becslésre épülő stratégiák – „ $|GK| > 1,5$ ”, illetve „ $|GK| > 1,5$ & 3 Classifiers” tudtak generálni (2,13%). A két sikeres stratégia közül az utóbbit a módszertani teljesség kedvéért definiáltam, de mivel a klasszifikációs modellek bevonása semmivel sem szűkítette a kiválasztott mérközések halmazát, a két módszer tökéletesen azonos eredményre vezetett. Ennek megfelelően a továbbiakban a két módszert együttesen fogom értékelni. Az önálló regressziós becslésekre épülő metódusok hozama nagyjából 0 környékén alakult, de a fedezeti pontot egyedül a Multilayer Perceptron Regressor 1,5 gólos küszöbértéket alkalmazó becslése érte el a tesztadatbázison. A legjobban teljesítő stratégia hozamának szignifikanciáját Wilcoxon tesztjének segítségével ellenőriztem, a 0,3154-es p-érték alapján viszont arra a következtetésre jutottam, hogy az eredmény nem szignifikáns.

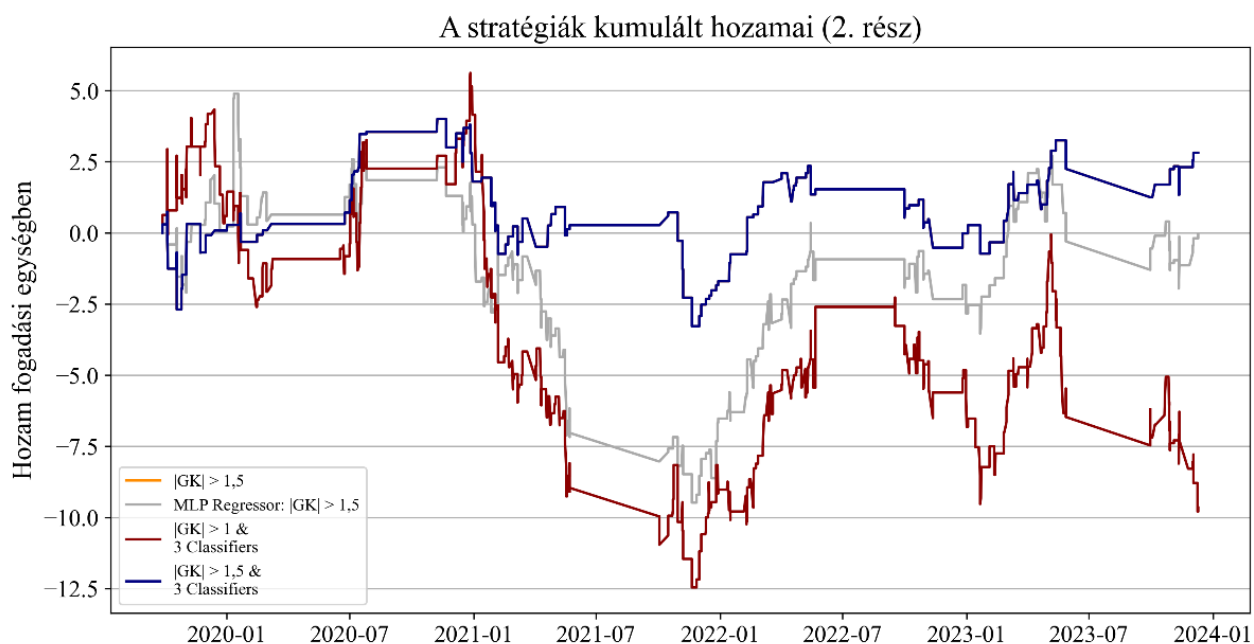
Természetes módon a fogadási módok minimum értéke -1 volt minden esetben, azaz a hibás fogadást semmilyen előrejelző módszerrel nem lehetett kiküszöbölni. A kvartilisek elemzése azonban már előre vetíti a „ $|GK| > 1,5$ ” stratégia sikerességét, mely az effektív meccs-kiválasztásban rejlik. Az első kvartilis ugyanis csak ebben az esetben vett fel pozitív értéket, minden más stratégia esetén -1 maradt, jelezve azok alacsonyabb találati arányát. A 44,46%-os pontosságú „Hazai” stratégiát leszámítva a medián hozamok egységesen 0,2 körül alakultak. A nettó hozamok maximum értékei jelentősen különböztek az egyes stratégiák esetében, mely jól mutatja azok kockázatvállalási hajlandóságát. Általánosságban a legalacsonyabb kockázatot a maximum 2,20-as oddsra sikeresen fogadó, ezzel 1,2 nettó hozamot generáló „ $|GK| > 1,5$ ” stratégia vállalta, míg a legnagyobbat a statisztikai előrejelzésre egyáltalán nem építő „Hazai” a maga 10-es értékű hozamával (11,00-s fogadott odds). A maximális generált hozamok, illetve az átlagos hozamok vizsgálata negatív összefüggést feltételez a hozam és kockázat dimenziójában, hiszen az MLP Regressort leszámítva az összes stratégia esetében az alacsonyabb maximális hozam magasabb átlaghozammal párosult. Ezen hipotézist erősíti meg az egyes metódusok szórás-mutatója is, mely a regressziós becslések esetén volt a legalacsonyabb (0,6 - 0,76), a referenciastratégiák esetében pedig a legmagasabb (0,93 – 1,37).

Stübinger és szerzőtársai (2019), illetve Knoll és Stübinger (2020) alapján az egyes stratégiák kumulált hozamait idősorként is elemeztem, illetve vizualizáltam. A fogadási egységben mért hozamok alakulását az 6. és 7. ábra mutatja be. A diagramon időszakosan látható megfigyelés nélküli egyenes vonalú szakaszok a szezonok első hat fordulójának mérkőzéseit mutatják, melyeket kizárólag a modellek tanítására használtam fel. A 7. ábrán a „ $|GK| > 1,5$ ”, illetve a „ $|GK| > 1,5$ & 3 Classifiers” stratégiák tökéletes egybeesése (csak a később vizualizált „ $|GK| > 1,5$ & 3 Classifiers” időszora látszik) egyértelműen igazolja a fogadási módszerek ekvivalenciájára vonatkozó korábbi megállapításomat.



6. Ábra: A fogadási stratégiák fogadási egységben mért kumulált hozamai az idő függvényében (1. rész). Forrás: Saját szerkesztés.

A pénzügyi teljesítmény idősoros diagramja megerősíti a stratégiák elemzésének azon korábbi konklúzióját is, mely szerint a statisztikai modellezést nélkülöző referenciastراتيجiák, illetve a klasszifikációs előrejelzések képtelenek voltak tartósan pozitív hozamot generálni. Természetesen ugyanez igaz a nem megfelelő kiválasztási küszöbértéket (mint kimeneteli valószínűséget) választó regressziós fogadási módszerekre is. A korábbi elemzésekkel összhangban a legmagasabb hozamot – és ezzel párhuzamosan a legkisebb volatilitást/ingadozást – a regressziós becslésekből másfél gólos küszöbértékkel választó fogadási stratégiák generálták.

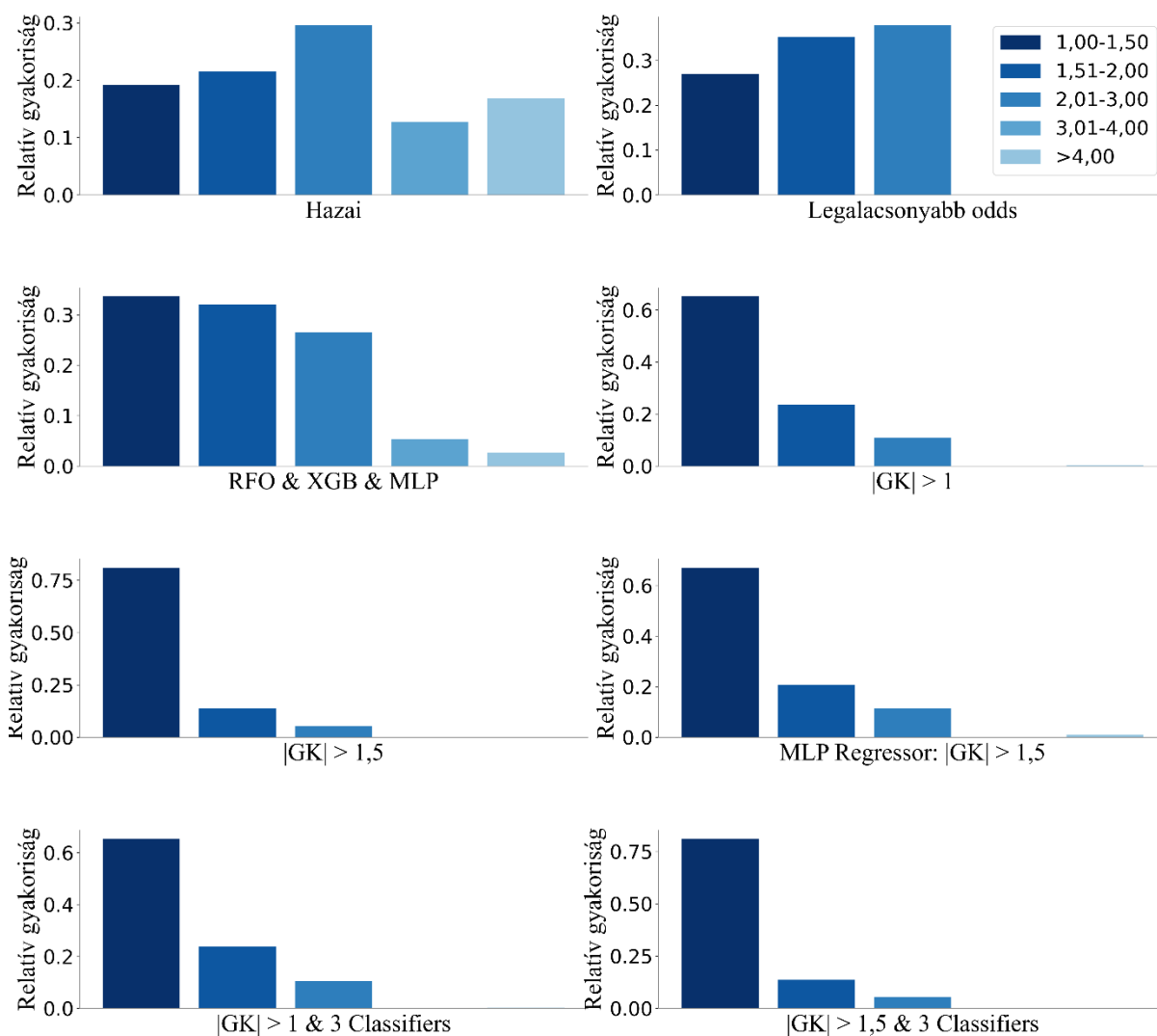


7. Ábra: A fogadási stratégiák fogadási egységben mért kumulált hozamai az idő függvényében (2. rész). Forrás: Saját szerkesztés.

Jól látható a 7. ábrán is, hogy a „ $|GK| > 1,5$ & 3 Classifiers” kumulált hozama mindössze két alkalommal, a tesztelési időablak kezdetén, illetve 2021 őszén süllyedt számottevően 0 alá, összességében pedig inkább pozitív tartományban mozgott. Az idősorokon megfigyelhető trendeknek megfelelően a legalacsonyabb kumulált értékkel a legtöbb fogadást eszközölő referenciastratégiák, a „Legalacsonyabb odds” és a „Hazai”, illetve a három klasszifikációs becslést alapul vevő „RFO & XGB & MLP” módszer zárt, sorrendben -39,84-gyel, -27,12-vel és -33,58-cal. Ezzel szemben a legjobban teljesítő metódusok közül a „ $|GK| > 1,5$ ”, illetve a „ $|GK| > 1,5$ & 3 Classifiers” +2,81-es, a „MLP Regressor: $|GK| > 1,5$ ” pedig -0,05-ös kumulált hozammal zárt. A regressziós becslésekből 1-es küszöbértékkel kiválasztó „ $|GK| > 1,5$ ”, illetve az ezt még klasszifikációs előrejelzéssel kombináló „ $|GK| > 1$ & 3 Classifiers” összesen 10,65, illetve 9,65 fogadási egységnyi veszteséget halmozott fel. A kumulált hozamok idősoros elemzése alapján levonható a következtetés, hogy a stratégiák teljesítményei közti különbségeket nem outlier értékek, hanem a hatékonyabb meccs-kiválasztás okozza. Azaz a „ $|GK| > 1$ & 3 Classifiers” azért eredményezte a legmagasabb hozamot, mert a legmagasabb pontossággal választotta ki azokat a mérkőzéseket, amelyeken az előrejelzett kimenetel kellően magas valószínűséggel bekövetkezett. A legalacsonyabb fogadásszám, a legkisebb szórás, illetve a legmagasabb találati arány is mind ezt támasztják alá.

Az egyes fogadási stratégiák kockázatvállalási hajlandósága a kiválasztott oddsok elemzésével írható leginkább, melyet az 8. ábra mutat be. A diagramokon a fogadásra kiválasztott oddsok hisztogramja látható stratégiánként.

A fogadott oddsok hisztogramja az egyes stratégiák esetében



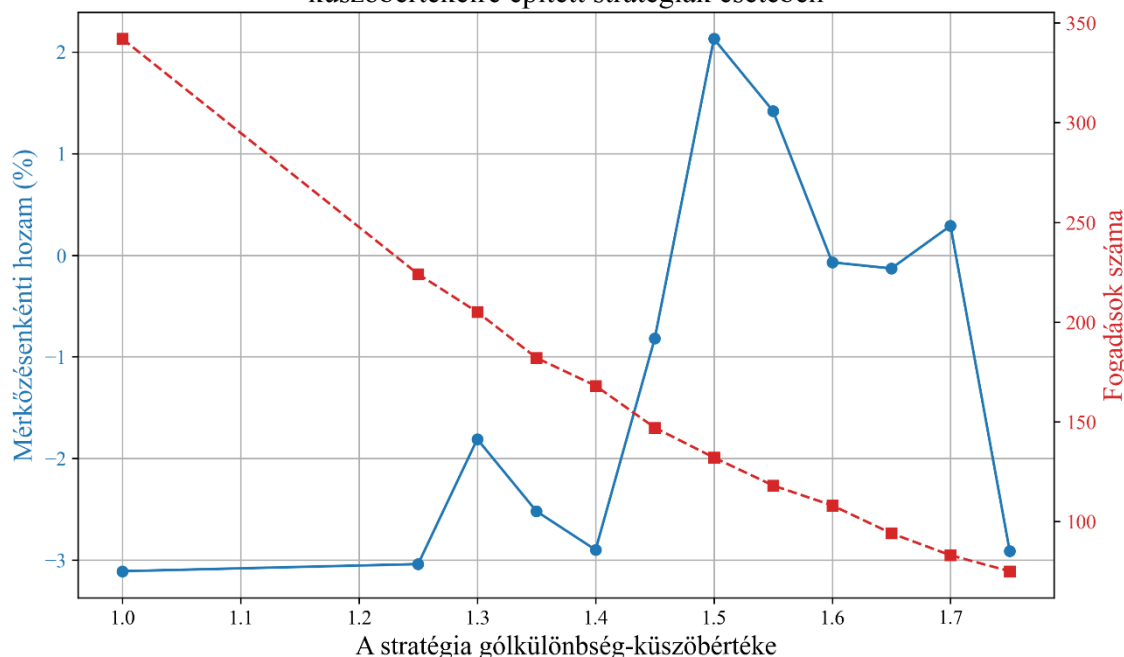
8. Ábra: A fogadási stratégiák által kiválasztott oddsok hisztogramja a 2019-2023-as tesztidőszak mérkőzései alapján. Forrás: Saját szerkesztés.

Definíciójából adódóan a legdiverzebb oddsokat a „Hazai” stratégia választotta, amely majdnem az esetek 30%-ában javasolt extrém kimeneteleket (3,00 feletti oddsot). Szintén meghatározásából fakadóan a „Legalacsonyabb odds” csak 3,00 alatti oddsokat választott, melyek eloszlása azonban enyhén a maximum felé ferdült. A kalsszifikációra épülő módszer már igyekezett biztosabb kimenetellel kecsegtető mérkőzéseket választani, de a fogadási lista még így is közel 10%-ban extrém – 3,00 feletti odds-szal rendelkező - kimeneteleket

tartalmazott. Ezzel szemben a regressziós becslést használó fogadási stratégiák szinte kizárólag alacsony odds-szal járó, rendkívül valószínű kimeneteket választottak alacsonyabb elemszámú tranzakciós listával, amely extrém módon jobbra ferdült eloszlást eredményezett minden esetben. A legnagyobb arányban a „ $|GK| > 1,5$ ” és a „ $|GK| > 1$ & 3 Classifiers” stratégia választott biztosnak tűnő (az oddsok alapján legalább $P = 2/3$ valószínűséggel bekövetkező) kimenetelt, de még a leginkább kockázatvállalónak tekinthető „MLP Regressor: $|GK| > 1,5$ ” előrejelzésében is csak 1/3 arányban van jelen 1,50-nél magasabb odds-szal rendelkező eredmény.

Az egyes stratégiákat kockázatvállalási hajlandóság szempontjából összehasonlítva látszik, hogy a legmagasabb hozamot a leginkább kockázatkerülő stratégia érte el. A diagramok alapján úgy tűnik tehát, hogy a sportfogadópiacon a kockázat minimalizálása a legmagasabb hozamokhoz vezető út. Mivel a mérkőzések kimeneteléhez kapcsolódó kockázatot a regressziós modellek voltak képesek legjobban megragadni, a kérdést részletesebben az átlagos gólkülönbség-becslések elemzésén keresztül vizsgáltam meg. A három regressziós modell által becsült gólkülönbség átlagára 1,25 és 1,75 közt 0,05-os lépésközönkénti küszöbértékekkel fogadási stratégiákat illesztettem, majd megvizsgáltam ezen fogadási módok mérkőzésenkénti hozamát (ezúttal már %-ban kifejezve), illetve a hozzájuk kapcsolódó fogadási lista elemszámát. Az érzékenységvizsgálat eredményét a 9. ábra foglalja össze.

Mérkőzésenkénti hozam és fogadások száma a gólkülönbség küszöbértékeire épített stratégiák esetében

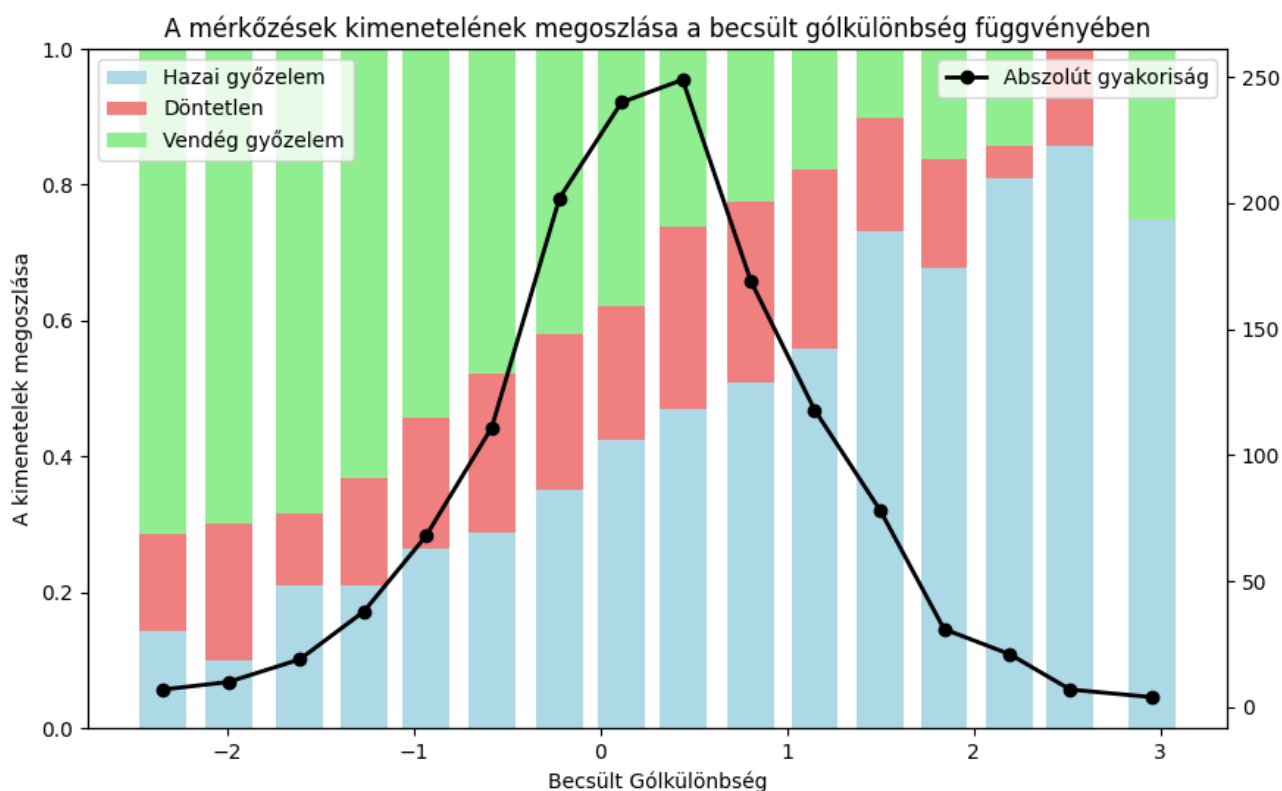


9. Ábra: A gólkülönbségbecslés küszöbértékének érzékenység vizsgálata: mérkőzésenkénti hozam és fogadásszám a küszöbérték függvényében. Forrás: Saját szerkesztés.

A diagram egyértelműen mutatja a fogadásszám monoton csökkenését a stratégiák szigorodásával (azaz magasabb belépési küszöb figyelembevételével). A jelenség matematikailag magyarázható, hiszen a szigorúbb küszöbérték alkalmazása nem más, mint a becsült gólkülönbség folyamatos értelmezési tartományán (halmazán) a metszetképzés alkalmazása. Az elért hozam és a küszöbérték között viszont egy sokkal kevésbé egyértelmű kapcsolat fedezhető fel: 1,5-es érték alatt a küszöbérték szigorítása (azaz a kockázat csökkentése) jellemzően növeli az elért hozamot, a másfeles határ felett azonban ez a trend hirtelen megfordul. Azaz a kockázat minimalizálása csak egy bizonyos pontig vezet eredményre, ezután a szigorítás már a fogadási bázis és az oddsok drasztikus lecsökkentéséhez vezet, amely végső soron visszaveti a stratégiák profitabilitását.

A jelenség hátterében az átlagosan egyre alacsonyabb odds-ok mellett a mérkőzések kimeneteléhez kapcsolódó nem azonosítható, szisztematikus kockázat rejlik. Wunderlich és szerzőtársai (2021) az angol Premier League hét szezonjának mérkőzéseit elemezve próbálták meghatározni a gólok születésében a véletlen események szerepét, és kutatásában arra jutottak, hogy a vizsgált 7263 gól majdnem 50%-ában szerepet játszott véletlen esemény. Az azonosított jelenség ahhoz vezet, hogy még az előzetesen teljesen biztosnak kimenetelűnek ítélt mérkőzésekhez is jelentős véletlen faktor kapcsolódik. Ezt erősíti meg a 10. ábra is, amely a becsült gólkülönbség függvényében hisztogramként ábrázolja a fogadási stratégiák tesztelésére felhasznált mérkőzések kimenetelét. Mindezek mellett a diagram az egyes becslési kategóriákba eső mérkőzések relatív gyakoriságát is tartalmazza. A kimenetelek megoszlása a becsült gólkülönbség függvényében a várakozásoknak megfelelően alakult, hiszen a gólkülönbség növekedésével emelkedett a hazai győzelmek aránya és csökkent a vendégeké. Az extrém értékek vizsgálata ugyanakkor egyértelműen mutatja, hogy lehet bármilyen magabiztos egy előrejelzés, nagyjából $P = 0,2 - 0,3$ valószínűséggel a várt eredmény mégsem fog bekövetkezni. A $(-1,5; 1,5)$ gólos átlagosan becsült különbség intervallumból kilépve megfigyelhető, hogy negatív különbség esetén a vendég-, pozitív különbség esetén a hazai győzelmek aránya 70-75% környékén stabilizálódik. Ezen ponton túl már a válogatási küszöbérték növelése nem eredményez jobb becslési pontosságot, illetve magasabb elért hozamot. Ezen modellkeretben tehát az optimális fogadási stratégia kidolgozása ekvivalens a tesztalmaz azon legbővebb részének elkülönítésével, amelyben a hazai/vendég győzelmek

aránya maximális és állandó. Kutatásom során használt tesztadatbázis esetében ez a 1,5 gólos küszöbérték alkalmazásával volt elérhető.



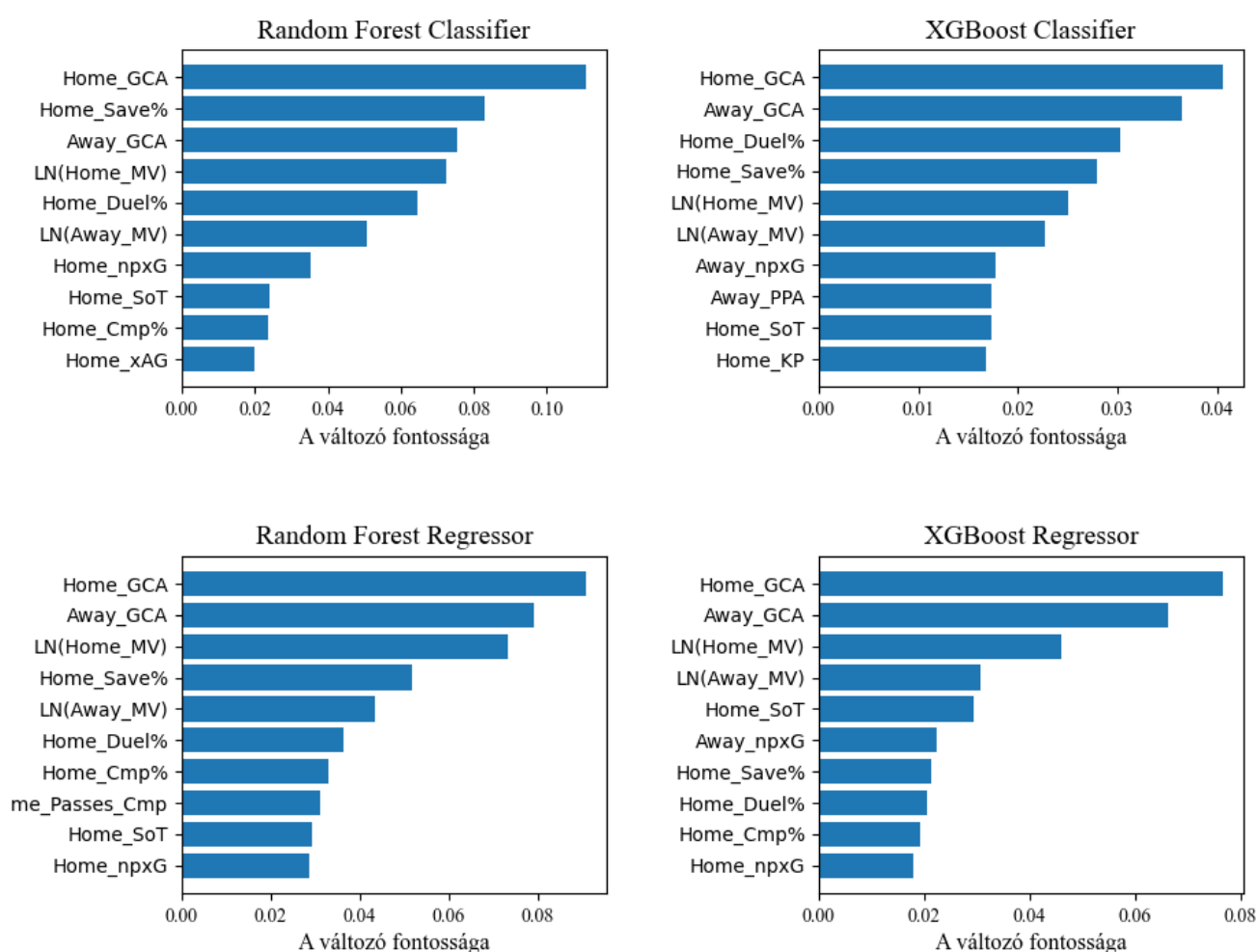
10. Ábra: A mérkőzések kimenetelének megoszlása az átlagosan becsült gólkülönbség függvényében. Forrás: Saját szerkesztés.

IV.3. A fogadási stratégiák összegzése, értékelése

Kutatásomban az angol Premier League mérkőzéseire kapcsolódó sportfogadópiacon igyekeztem statisztikai arbitrázs-lehetőségeket azonosítani és kihasználni machine learning modellek alkalmazásával. Összehasonlítva eredményeimet Knoll és Stübinger (2021) tanulmányával megállapítható, hogy bár a modellek előrejelzési pontossága emelkedett, a legjobb stratégia által realizált hozam elmaradt a korábbitól (5,42% vs. 2,13%). A magasabb pontosság ellenére alacsonyabb hozamokra magyarázatot adhat a kutatások eltérő fókusza, hiszen míg Knoll és Stübinger a Premier League mellett a spanyol, a német és az olasz első osztály mérkőzéseit is vizsgálta, addig jelen elemzés kizárólag az angol első osztály meccseire koncentrált. Az angol bajnokságra épülő sportfogadópiac a leglikvidebb, leghatékonyabb piacok közé tartozik, így kevésbé ért meglepetésként, hogy az előrejelző modellek csak elhanyagolható mértékben voltak képesek arbitrázslehetőséget azonosítani. Másrészt az idő

dimenzióját sem szabad elhanyagolni, hiszen míg Knoll és Stübinger a 2013-2018-as időszak adatait használta, addig jelen kutatás a 2018-2023-as időszak mérkőzéseit vette alapul. A két adatgyűjtés között bár bővült az elérhető statisztikai adatok köre, de ezzel párhuzamosan pontosabbá váltak a sportfogadóirodák előrejelzései is, amely hatékonyabbá tette a piacokat. Továbbá lehetséges okot szolgáltathat akár maga a pusztán véletlen is, hiszen a legjobb sportfogadó stratégiák által kiválasztott mérkőzések kimenetele is bizonytalan volt az esetek körülbelül 25%-ában. Két-három szerencsésen alakuló mérkőzés pedig akár 4-5%-ra is emelheti az egy mérkőzésre jutó realizált hozamot.

A Random Forest és XGBoost modellek legfontosabb magyarázó változói (feature importance)



11. Ábra: A mérkőzések kimenetlének megoszlása az átlagosan becsült gólkülönbség függvényében. Forrás: Saját szerkesztés.

Jelmagyarázat: GCA: Kialakított gólhelyzetek száma, Save%: védési hatékonyság, MV: A kezdőcsapat piaci értéke, Home_Duel%: A párharcok hazai csapat által megnyert aránya, npxG: Várható gólok a büntetők leszámításával, SoT: Kaput eltaláló lövések száma, Cmp%: Passzok pontossága, xAG: Várható gólpasszok, PPPA: Büntetőterületen belülrre érkező passzok száma, KP: Kulcspasszok száma, Passes_cmp: Sikeres passzok száma.

Knoll és Stübinger (2021) kutatásához képest jelentősen bővítettem a magyarázó változók körét, mely a változók fontossági mutatóján is egyből megmutatkozik (11. ábra). A komplexebb elérhető statisztikai adatok (pl.: xG, xAG, ...) alkalmazása mellett tartalmi újítként emeltem be a modellbe három külső változót: a kezdőcsapatok piaci értékét, a hiányzók számát, illetve a csapatok aktuális formáját. A három új elem közül a kezdőcsapatok piaci értéke az összes modell legfontosabb magyarázó változói közé került, amely jól jelzi az előrejelzések komplexitásának bővítését. A piacok hatékonyabbá válásának, illetve a labdarúgásban jelen lévő kikerülhetetlen bizonytalansági faktor miatt azonban ez a plusz komplexitás magasabb realizált hozamokban nem tudott megmutatkozni.

Eredményeim összességében megerősítik a Premier League mérkőzéseire épülő sportfogadópiac hatékonyságát, a labdarúgásban jelen lévő véletlen faktor nagy szerepét, illetve a sikeres sportfogadó stratégiák hatékony mérkőzés-kiválasztásra vonatkozó képességét. Ez utóbbi azon legbővebb mérkőzésekből álló klaszter elkülönítésével érhető el, amelyben a legmagasabb az előrejelezhető kimenetek aránya. A paramétereknek megfelelő mérkőzések azonosításával megpróbálkoztam klasszifikációs és regressziós eljárással is, de csak utóbbival jártam sikerrel, megerősítve ezáltal a Knoll és Stübinger (2019 és 2020) által alkalmazott módszertan érvényességét.

V. Összegzés

Kutatásomban az angol első osztályú labdarúgó bajnokság 2018 augusztusa és 2023 decembere között lejátszott mérkőzéseit és a hozzájuk kapcsolódó sportfogadópiacokat vizsgáltam. A találkozók kimenetelét azok körülményeit (együttesek formája, kezdőcsapatának állapota) leíró változók és a csapatok közelmúltbeli teljesítménye alapján jeleztem előre. A kimenetek becslését – melyet osztályozási és regressziós feladatként külön-külön is elvégeztem - gépi tanuláson és neurális hálókön alapuló modellek segítségével végeztem. A becslések alapján különböző sportfogadási stratégiákat definiáltam, melyeket aztán az adatbázis 1372 mérkőzésből álló részhalmazán empirikus tesztelésnek vettem alá.

A különböző becsléseket megfelelően kombináló stratégiák találati aránya jóval meghaladta a 70%-ot, a legjobb pedig 76,52%-ot ért el, elérve ezáltal a hasonló kutatások (lásd 1. Táblázat) során tapasztalt legjobb eredmények szintjét. A feldolgozott szakirodalommal összhangban jelen kutatásban is számottevő különbség mutatkozott a klasszifikációs és regressziós modellekre épített stratégiák előrejelző pontossága között, természetesen az utóbbi javára. Mindazonáltal még a legpontosabban teljesítő fogadási lista is csak 2,13%-os mérkőzésenkénti hozamot volt képes generálni, mely elmaradt a hasonló témájú kutatások során tapasztalt eredményességtől. A stratégiák kockázatvállalási hajlandóságának elemzése alapján arra jutottam, hogy a stratégiák teljesítménye fordítottan arányos az általuk vállalt kockázattal. Másképp értelmezve a gazdaságilag sikeres fogadási módszerek legfőbb ismérve a nagy valószínűséggel bekövetkező kimenetek azonosítása volt. A jelenség hátterében a gólokat az esetek jelentős részében magyarázó véletlen hatások (Wunderlich és szerzőtársai, 2021), illetve a döntetlen kimenetel prediktálását övező magas bizonytalanság állhat.

Az optimális stratégia összeállításának logikáját a - kimenetel előrejelzésére legalkalmasabb - gólkülönbség-becslés érzékenységvizsgálata mutatta meg. Bár a hazai - vagy vendég győzelem bekövetkezési valószínűségét a kategorizáláskor választott magasabb abszolútértékű küszöbérték képes volt növelni, a nem várt kimenetel bekövetkezésének esélye minden küszöbérték alkalmazása esetén legalább 20-30% maradt. Az optimális fogadási stratégia tehát ezen szerkezetű legbővebb részhalmaz elhatárolásával állt elő, és végül a 1,5 gólos küszöbérték alkalmazásával volt elérhető.

Az egyes fogadási listák pénzügyi teljesítménye alapján a 2018 augusztusa és 2023 december közti időszakban nem sikerült szignifikáns statisztikai arbitrázslehetőséget azonosítanom a Premier League sportfogadópiacán, mely megerősíti a piac hatékonyságára

vonatkozó feltételezést. Mindazonáltal eredményeimre számos korlátozó tényező hatással lehetett, ezért a kutatást érdemes lenne továbbfejlesztett módszertannal nagyobb adatbázison megismételni. Korlátot jelentett a fogadásra használható mérkőzések relatíve alacsony száma, mely a modellezés során használt adatok csak 2018 óta nyilvánosan elérhető voltából eredt. A vizsgált időszak jelentős részét érintő Covid-19 járvány által okozott egészségügyi és gazdasági bizonytalanság a labdarúgásban is éreztette hatását, mely tovább csökkenthette a modellek előrejelző-képességét. Az alacsony realizált hozamokat ezek mellett magyarázhatják a fogadóirodák gyakori odds-korrekciói is (melynek célja a beérkező téték irányának kiegyenlítése), hiszen jelen kutatásban a Bet365 kezdőrúgás előtti utolsó hivatalos szorzóit vettem figyelembe. Végezetül a topbajnokságok mérkőzéseivel kapcsolódó fogadások magas likviditása is hozzájárulhat a nem szignifikáns pozitív hozamokhoz, így kevésbé népszerű, de kellő nyilvános információval rendelkező mérkőzések elemzésével valószínűleg erősebb eredmények érhetőek el.

A tapasztalt korlátok miatt úgy gondolom, hogy a kutatást érdemes lenne a jövőben továbbfejlesztett módszertannal megismételni. A csapatok teljesítménymutatóinak előrejelzésétől kezdve a szofisztikáltabb változószelekción és külső körülmények - például nyilvános oddsok segítségével történő - hatékonyabb modellezésén át a könnyen előrejelezhető kimeneteleket tartalmazó részhalmaz hatékonyabb kiválasztását célzó módszerek és stratégiák mind-mind olyan területek, amelyek pontosítása jobb eredménnyel kecsegtet. Végezetül pedig egyértelműen érdemes lenne bővíteni a vizsgált bajnokságok/mérkőzések körét is, hiszen a kevésbé ismert bajnokságok elemzésbe történő bevonása egyrészt a nagyobb mintán keresztül növelhetné a modellek becslési pontosságát, másrészt pedig kevésbé likvid/hatékony piacaik révén valószínűleg a Premier League-hez képest több statisztikai arbitrázs-lehetőséget kínálnak.

Jelen dolgozat a korábbi kutatásokhoz képes átfogóbb és komplexebb módszertannal próbálta előrejelezni az angol Premier League mérkőzéseinek kimenetelét. A korábbi kutatások során alkalmazott modellekhez képest a magyarázó változók halmazát kiegészítettem a kezdőcsapatok piaci értékét, a sérült és eltiltott játékosok számát, valamint a csapatok közelmúltbeli formáját leíró mutatókkal. A becslések alapján optimális stratégia alkalmazásával sikerült javítani a becslések pontosságát (egészen 76% fölé), a többnyire hatékonyan működő sportfogadó piac, illetve a mérkőzéseken jelen lévő általános bizonytalanság miatt azonban nem volt képes szignifikánsan pozitív hozamot generálni. Mindezek ellenére az alkalmazott modellek szerkezete és a felhasznált változók széles köre utat mutathat a későbbi kutatások

számára, melyek a futballmeccsek kimenetelének prediktálását vagy a sportfogadópiacok hatékonyságának elemzését célozzák. Az ugyanis valószínűsíthető, hogy a fogadóirodák és a fogadók között az arbitrázsprofit felkutatásáért folyó verseny nemhogy lassulni nem fog, de az adatelemzési módszerek és a piacok fejlődésével párhuzamosan még inkább kiéleződik.

VI. Hivatkozások jegyzéke

- Alexander, C., & Dimitriu, A. (2005). Indexing and Statistical Arbitrage. *The Journal of Portfolio Management*, 31(2), 50–63. <https://doi.org/10.3905/jpm.2005.470578>
- Angelini, G., & De Angelis, L. (2019). Efficiency of online football betting markets. *International Journal of Forecasting*, 35(2), 712–721. <https://doi.org/10.1016/j.ijforecast.2018.07.008>
- Archontakis, F., & Osborne, E. (2007). Playing It Safe? A Fibonacci Strategy for Soccer Betting. *Journal of Sports Economics*, 8(3), 295–308. <https://doi.org/10.1177/1527002506286775>
- Are Betting Markets Inefficient? Evidence From Simulations and Real Data—David Winkelmann, Marius Ötting, Christian Deutscher, Tomasz Makarewicz, 2024. (é. n.). Elérés 2024. április 14., forrás <https://journals.sagepub.com/doi/full/10.1177/15270025231204997>
- Baboota, R., & Kaur, H. (2019). Predictive analysis and modelling football results using machine learning approach for English Premier League. *International Journal of Forecasting*, 35(2), 741–755. <https://doi.org/10.1016/j.ijforecast.2018.01.003>
- Beal, R., Middleton, S. E., Norman, T. J., & Ramchurn, S. D. (2021). Combining Machine Learning and Human Experts to Predict Match Outcomes in Football: A Baseline Model. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(17), 15447–15451. <https://doi.org/10.1609/aaai.v35i17.17815>
- Bendeguz K. H. (2022). A hazai pálya előnye a labdarúgásban. Tudományos Diákköri Konferencia Dolgozat. Budapesti Corvinus Egyetem. Témavezető: Madari Zoltán.
- Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, 15(1), 27–33. <https://doi.org/10.1016/j.aci.2017.09.005>
- Carloni, L., De Angelis, A., Sansonetti, G., & Micarelli, A. (2021). A Machine Learning Approach to Football Match Result Prediction. In C. Stephanidis, M. Antona, & S. Ntoa (Szerk.), *HCI International 2021—Posters* (Köt. 1420, o. 473–480). Springer International Publishing. https://doi.org/10.1007/978-3-030-78642-7_63

- Chinwe Peace, I. (2014). An Improved Prediction System for Football a Match Result. IOSR Journal of Engineering, 04(12), 12–020. <https://doi.org/10.9790/3021-04124012020>
- completesports.com, 2023: Top European Betting Pages: Exploring The Best Platforms For Wagering. Letöltés helye: <https://www.completesports.com/top-european-betting-pages-exploring-the-best-platforms-for-wagering/>, Letöltés ideje: 2024. 04. 09.
- Csikai, M. (2016): Sportfogadási és pénzügyi piacok kapcsolata. (2016). MSc Diplomamunka. Budapesti Corvinus Egyetem, Biztosítási és pénzügyi matematika MSc. Témavezető: Badics Milán Csaba.
- DBMR, 2022: Europe Sports Betting Market – Industry Trends and Forecast to 2030. Letöltés helye: <https://www.databridgemarketresearch.com/reports/europe-sports-betting-market>, Letöltés ideje: 2024. 04. 11.
- Elaad, G., Reade, J. J., & Singleton, C. (2020). Information, prices and efficiency in an online betting market. Finance Research Letters, 35, 101291. <https://doi.org/10.1016/j.frl.2019.09.006>
- Feddersen, A., Humphreys, B. R., & Soebbing, B. P. (2023). Contest incentives, team effort, and betting market outcomes in European football. European Sport Management Quarterly, 23(3), 605–621. <https://doi.org/10.1080/16184742.2021.1898432>
- FIFA.com, 2023: One Month On: 5 billion engaged with the FIFA World Cup Qatar 2022™. Letöltés helye: <https://inside.fifa.com/tournaments/mens/worldcup/qatar2022/news/one-month-on-5-billion-engaged-with-the-fifa-world-cup-qatar-2022-tm>, Letöltés ideje: 2024. 04. 10.
- GII, 2022: Europe Sports Betting Market - Industry Trends and Forecast to 2030. Letöltés helye: <https://www.giiresearch.com/report/dbmr1189357-europe-sports-betting-market-industry-trends.html>, Letöltés ideje: 2024. 04. 09.
- Goddard, J., & Asimakopoulou, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. Journal of Forecasting, 23(1), 51–66. <https://doi.org/10.1002/for.877>
- Groll, A., Ley, C., Schauburger, G., & Van Eetvelde, H. (2018). Prediction of the FIFA World Cup 2018—A random forest approach with an emphasis on estimated team ability parameters (arXiv:1806.03208). arXiv. <https://doi.org/10.48550/arXiv.1806.03208>

- GVR, 2024: Europe Sports Betting Market Size, Share & Trends Analysis Report By Type, By Platform, By Sports Type (Football, Basketball, Baseball, Horse Racing, Cricket, Hockey, Others), By Country, And Segment Forecasts, 2024 – 2030. Letöltés helye: <https://www.grandviewresearch.com/industry-analysis/europe-sports-betting-market-report>, Letöltés ideje: 2024. 04. 09.
- Improved Least Squares Football, Basketball, and Soccer Predictions. (1980). IEEE Transactions on Systems, Man, and Cybernetics, 10(2), 116–123. <https://doi.org/10.1109/TSMC.1980.4308442>
- Khan, S., & V. B, K. (2019). Comparing machine learning and ensemble learning in the field of football. International Journal of Electrical and Computer Engineering (IJECE), 9(5), 4321. <https://doi.org/10.11591/ijece.v9i5.pp4321-4325>
- Knoll, J., & Stübinger, J. (2020). Machine-Learning-Based Statistical Arbitrage Football Betting. KI - Künstliche Intelligenz, 34(1), 69–80. <https://doi.org/10.1007/s13218-019-00610-4>
- Levitt, S. D. (2004). Why are Gambling Markets Organised so Differently from Financial Markets? The Economic Journal, 114(495), 223–246. <https://doi.org/10.1111/j.1468-0297.2004.00207.x>
- Matej, U., Gustav, Š., Ondřej, H., & Filip, Ž. (2021). Optimal sports betting strategies in practice: An experimental review. IMA Journal of Management Mathematics, 32(4), 465–489. <https://doi.org/10.1093/imaman/dpaa029>
- Mattera, R. (2023). Forecasting binary outcomes in soccer. Annals of Operations Research, 325(1), 115–134. <https://doi.org/10.1007/s10479-021-04224-8>
- Parke, A., & Parke, J. (2019). Transformation of Sports Betting into a Rapid and Continuous Gambling Activity: A Grounded Theoretical Investigation of Problem Sports Betting in Online Settings. International Journal of Mental Health and Addiction, 17(6), 1340–1359. <https://doi.org/10.1007/s11469-018-0049-8>
- Pipatchatchawal, C., & Phimoltares, S. (2021). Predicting Football Match Result Using Fusion-based Classification Models. 2021 18th International Joint Conference on Computer Science and Software Engineering (JCSSE), 1–6. <https://doi.org/10.1109/JCSSE53117.2021.9493837>

- Popescu, M.-C., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer Perceptron and Neural Networks. 8(7).
- Raju, M. A., Mia, Md. S., Sayed, Md. A., & Riaz Uddin, Md. (2020). Predicting the Outcome of English Premier League Matches using Machine Learning. 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), 1–6. <https://doi.org/10.1109/STI50764.2020.9350327>
- Richard, M., & Vecer, J. (2021). Efficiency Testing of Prediction Markets: Martingale Approach, Likelihood Ratio and Bayes Factor Analysis. *Risks*, 9(2), 31. <https://doi.org/10.3390/risks9020031>
- Rigatti, S. J. (2017). Random Forest. *Journal of Insurance Medicine*, 47(1), 31–39. <https://doi.org/10.17849/insm-47-01-31-39.1>
- Rodrigues, F., & Pinto, Â. (2022). Prediction of football match results with Machine Learning. *Procedia Computer Science*, 204, 463–470. <https://doi.org/10.1016/j.procs.2022.08.057>
- Rudrapal, D., Boro, S., Srivastava, J., & Singh, S. (2020). A Deep Learning Approach to Predict Football Match Result. In H. S. Behera, J. Nayak, B. Naik, & D. Pelusi (Szerk.), *Computational Intelligence in Data Mining* (Köt. 990, o. 93–99). Springer Singapore. https://doi.org/10.1007/978-981-13-8676-3_9
- Rue, H., & Salvesen, O. (2000). Prediction and Retrospective Analysis of Soccer Matches in a League. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3), 399–418. <https://doi.org/10.1111/1467-9884.00243>
- Similarweb.com, 2024: Top Websites Ranking. Letöltés helye: <https://www.similarweb.com/top-websites/gambling/sports-betting/>, Letöltés ideje: 2024. 04. 09.
- Sofascore.com, 2024: <https://www.sofascore.com/tournament/football/england/premier-league/17#id:52186>
- Statista.com, 2023: Total amount wagered on European soccer worldwide in the 2020/2021 season, by league. Letöltés helye: <https://www.statista.com/statistics/1263462/value-betting-on-european-soccer/>, Letöltés ideje: 2024. 04. 09.

- Statista.com, 2024a: Online Sports Betting - Worldwide. Letöltés helye: <https://www.statista.com/outlook/dmo/eservices/online-gambling/online-sports-betting/worldwide>, Letöltés ideje: 2024. 04. 09.
- Statista.com, 2024b: Online Sports Betting – Europe. Letöltés helye: <https://www.statista.com/outlook/dmo/eservices/online-gambling/online-sports-betting/europe>, Letöltés ideje: 2024. 04. 09.
- Stübinger, J., Mangold, B., & Knoll, J. (2019). Machine Learning in Football Betting: Prediction of Match Results Based on Player Characteristics. *Applied Sciences*, 10(1), 46. <https://doi.org/10.3390/app10010046>
- Tax, N. & Yme Joustra. (2015). Predicting The Dutch Football Competition Using Public Data: A Machine Learning Approach. <https://doi.org/10.13140/RG.2.1.1383.4729>
- transfermarkt.com, 2024: Market value of European leagues and cups. Letöltés helye: <https://www.transfermarkt.com/wettbewerbe/europa>, Letöltés ideje: 2024. 04. 09.
- Vaidya, S., Sanghavi, H., & Gevaria, K. (2016). Football Match Winner Prediction. *International Journal of Computer Applications*, 154(3), 31–33. <https://doi.org/10.5120/ijca2016912066>
- Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. *Journal of Forecasting*, 28(5), 426–444. <https://doi.org/10.1002/for.1085>
- Wunderlich, F., Seck, A., & Memmert, D. (2021). The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. *Journal of Sports Sciences*, 39(20), 2322–2337. <https://doi.org/10.1080/02640414.2021.1930685>
- Zeileis, A., Leitner, C., & Hornik, K. (2016). Predictive bookmaker consensus model for the UEFA Euro 2016.
- Zeileis, A., Leitner, C., & Hornik, K. (2018). Probabilistic Forecasts for the 2018 FIFA World Cup Based on the Bookmaker Consensus Model.
- Рейкін, В. (2021). Financial and Economic Aspects of Professional Football Industry: Trends Analysis and Development Prospects. *Financial and Credit Activity Problems of Theory and Practice*, 1(36), Article 36. <https://doi.org/10.18371/fcaptop.v1i36.228089>