

Szerencsejáték-vásárlási szokások vizsgálata „prediktív analitika” segítségével

Tessényi Judit,

a Szegedi Tudományegyetem
PhD-jelölt

E-mail: tessenyi.judit@szrt.hu

Kazár Klára,

a Szegedi Tudományegyetem
tudományos segédmunkatársa

E-mail: kazar.klara@eco.u-
szeged.hu

A prediktív analitika lényege, hogy minél több múltbeli adatból jelezzük előre (modellezzük) a jövőbeli várható viselkedést. Minél nagyobb mennyiségű ehhez szükséges adat áll rendelkezésünkre és minél pontosabban ismerjük az összefüggéseket az egyes vizsgálati elemek között, annál nagyobb megbízhatósággal határozhatjuk meg, hogy a vizsgált szereplőink (játékosaink, vevőink, klasztereink) miként reagálnak a jövőben, és közülük kik azok a kulcsszereplők, meghatározó elemek, akik megőrzése vagy befolyásolása különösen fontos.

Dolgozatunkban kísérletet teszünk a prediktív analitika rövid bemutatását követően a problémás szerencsejátékosok meghatározására, illetve a ma még „normálisan” játszó szereplők jövőbeli viselkedésváltozásának előrejelzését segítő tényezők rendszerezésére. Az SPSS számítógépes programcsomag, illetve annak analitikus módszerei nagy mennyiségű adat feldolgozását és elemzését teszik lehetővé, melyet egyetlen értékesítési pont (lottózó) vevőinek értékelésével mutatunk be.

TÁRGYSZÓ:

Prediktív analitika.

Játék.

Döntési fa.

Manapság számos szoftvergyártó és kutató cég hirdeti a prediktív analitikai vagy adatbányászati módszerek előnyeit. Ezen eszközök segítségével végrehajtható a vevők, ügyfelek elemzése, mely által jövőbeli viselkedésük, reakcióik előre jelezhető. Dolgozatunkban az üzleti életben kedvelt kifejezésként használt prediktív analitika fogalmát járjuk körül, és egy speciális ügyfélkör, a Szerencsejáték Zrt. vevői esetén való alkalmazhatóságát és jelentőségét írjuk le. Elemzésünk folyamán egy szege-di lottózó négyhavi tranzakciós adatsorát vettük alapul, mely a vizsgált időszakban 26038 elemezhető tranzakciót jelentett.

1. A prediktív analitika fogalma, módszerei

A prediktív analitika olyan statisztikai modelleket és egyéb empirikus módszereket jelent, amelyek empirikus előrejelzésekre szolgálnak, valamint a módszerek minőségi értékelésére, az előrejelzések gyakorlati alkalmazhatóságára, azaz előrejelző képességének javítására törekszenek (*Shmueli–Koppius* [2011]). Eltekintve a gyakorlati hasznosság alapkritériumától, a prediktív analitika fontos szerepet játszik a tudományos elméletben, a hipotézisek tesztelésében és relevanciavizsgálatában. Ezért elengedhetetlen része a tudományos kutatásnak (*Kaplan* [1964], *Dubin* [1969]).

Shmueli és *Koppius* [2010] meghatározása szerint két összetevőből áll:

- empirikus prediktív modellek (statisztikai modellek és egyéb módszerek, mint például adatbányászati algoritmusok), amelyek előrejelzést adnak a megfigyelések alapján,
- a módszerek értékelése, a modell prediktív ereje, azaz előrejelző képessége (prediktív pontossága). Az a modell képes generálni pontos előrejelzést, amelyben az „új” is értelmezhető időben, azaz a megfigyelések egy jövőbeni időszakában (*Shmueli–Koppius* [2010]).

A prediktív analitika kifejezés az üzleti életben alkalmazott, a tudomány világában kevésbé elterjedt. Az üzlet világában számos meghatározás található a módszerre. Egy ilyen: „A prediktív analitika segítségével a szervezet előre láthat változásokat, így stratégiákat tervezhet és hajthat végre az eredmények javítása érdekében. A prediktív analitikai megoldások már meglévő adatokon való alkalmazásával váratlan mintákat és összefüggéseket tárhat fel, illetve modelleket dolgozhat ki az ügyfélkap-

csolati párbeszédék irányításához.” (IBM Magyarország [2012]). A *Predictive Analytics World* [2012] szerint a prediktív analitika olyan üzleti világban alkalmazható intelligens technológia, amely az ügyfelek vagy egyéb szervezeti elemek értékelését végzi.

A tudományos közegben a meghatározással a prognosztika és a jövő kutatás mutat rokonságot, azonban a prediktív analitika esetén a nem kifejezetten előrejelzési célú technikák (csoportosító, szegmentáló eljárások stb.) is prediktív erővel bírnak. A példaként említett vevőszegmentálás során az ügyfelek csoportokba sorolása által, jövőbeli viselkedésükre fogalmazhatunk meg szabályokat, melynek köszönhetően kikövetkeztethető jövőbeli magatartásuk. Ebben az értelmezésben prediktív analitikának tekinthető bármely, jövőre vonatkozó következtetések levonását lehetővé tevő elemzési módszer.

Tanulmányunk vásárlók csoportosítására, viselkedésük vizsgálatára és lehetséges előrejelzésére irányul, amire a direkt, az adatbázis- vagy az ügyfélmarketing-alapú módszerek alkalmasak. Érdemes megjegyezni azonban, hogy az adatbázis-marketing nemcsak marketingakciók (fogyasztói magatartás, keresztértékesítési lehetőségek, promóció-reagálási készség) elemzésére, hanem bankok, befektetési társaságok és biztosítók által ügyfélértékelésre is használatos pénzügyi területen (Kamarkura et al. [2003]). Az SPSS-szoftver is rendelkezik direkt marketing modullal, amelyen belül az ügyfelek értékelésére (és ezáltal jövőbeli viselkedésük előrejelzésére) alkalmas RFM- (recency, frequency, monetary, azaz az ügyfél aktualitásához, a gyakorisághoz, illetve a pénzügyi értékhez kapcsolódó) analízist és a szegmentálási, osztályozási módszereket emelnénk ki, mert elemzésünk során az említett eszközöket használjuk.

Az RFM-analízis ügyfélérték-számítást végez tranzakció- vagy ügyfélalapú adatbázis alapján. Az eljárás segítségével következtetni lehet például arra, hogy kik lehetnek az adott vállalat kulcsügyfelei, kik képezhetik a lemorzsolódó ügyfeleket, vagy többletinformáció szerezhető a vásárlók vásárlási vagy promóciós kampányra való reagálásáról (Chang–Tsai [2011]). A módszer minden ügyfélhez egy háromjegyű RFM-pontszámot rendel. Ennek első számjegye az ügyfél aktualitásához kapcsolódik (recency), a második a gyakorisághoz (frequency), a harmadik pedig a vevő pénzügyi értékéhez (monetary). Az aktualitás az utolsó tranzakció óta eltelt időt, a gyakoriság a vizsgált időszak alatt bekövetkező tranzakciók számát jelenti, míg a pénzügyi érték az adott időszak során megvásárolt termékek mennyiségéből származik (Bose–Chen [2009]). A program ezekre építve, kvantilis eloszlás alapján hozza létre az osztályközöket, avagy a pontszámokat. 1–9-ig terjedő értékeket lehet meghatározni, a pontszámok generálását a szoftver egymástól függetlenül vagy egymásba ágyazottan is végezheti. Az utóbbi lehetőség arra utal, hogy először az aktualitás szerint történik az osztályközökbe sorolás, majd ezen belül a gyakoriság osztályközének kialakítása, és ebben a pénzügyiérték-dimenzió pontjainak kiosztása. A módszer

gyakorlati haszna az osztályközökre épülő pontszámképzésen túl, hogy segítségével tranzakció-alapúból ügyfélalapú adatbázishoz juthatunk, továbbá az „egyszerű osztályközképzéshez” képest lehetőség van az osztályközök (pontszámok) egymásba ágyazott kezelésére is.

A szegmentálási, azaz a vevők osztályozására, csoportosítására irányuló módszerek között a marketing- és piackutatással foglalkozó szakirodalmak (*Malhotra* [2008], *Veres–Hoffmann–Kozák* [2006]) elsősorban a klaszter- és a diszkriminanciaelemzést szokták ismertetni. Az általunk alkalmazott szoftver esetén azonban az osztályozó technikák között szerepel a döntési fa is, mellyel a vizsgálati egyedek (jelen esetben a játékosok) irányított csoportosítása végezhető el. Az ezzel való következtetés *Abrahams et al.* [2009] szerint a prediktív algoritmusok osztályába tartozik. A fa-struktúra felépítése: a levelek (node-ok) jelentik a csoportokat, az ágak pedig a jellemzőket és a besorolást, ami alapján létrejött az adott csoportosítás. A fa szerkezetének kialakítása folyamán, a bemeneti változók értékeinek vizsgálatakor, az adatbázis különböző részekre osztása történik meg. A szerzők szerint a módszer könnyen alkalmazható, és az eredmény nehézség nélkül értelmezhető, hiszen csak a vágási szabályokat kell leolvasni a fáról az egyes csoportok jellemzéséhez (*Abrahams et al.* [2009]).

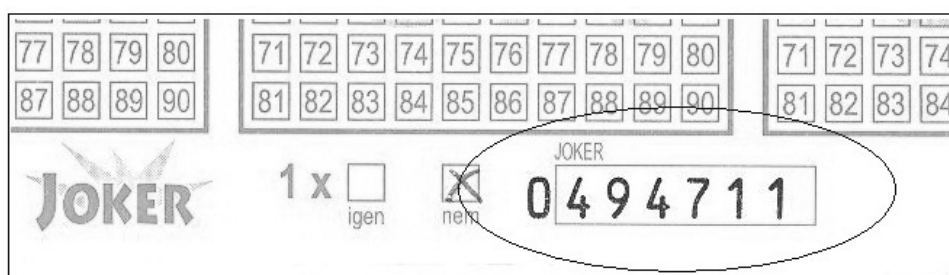
2. A prediktív analitika konkrét alkalmazása szerencsejáték-értékesítés és „problémás” játékosok esetén

A felnőtt magyar lakosság valamivel több mint fele (54 százaléka, körülbelül 4,3 millió fő) rendszeresen (legalább 2-3 havonta egyszer) játszik a Szerencsejáték Zrt. valamelyik játékával (*Kutatópont* [2011]), viszont az internethasználók csupán 3 százaléka teszi ezt interneten keresztül (*Medián* [2010]). A jelen dolgozat tárgyát képező, „hagyományos” módon játészó, azaz lottózóba betérő fogadó/játékos anonim, tehát személyes azonosítás a játékok feladása során nem történik. A szerencsejáték-szervezői szakszargonban a fogadó a sportfogadásokon játszó játékost, míg a játékos kifejezetten a számsorsjátékot és kaparós sorsjegyen játszókat takarja. Ezen megkülönböztetés elemzésünk szempontjából azért is hasztalan lenne, mert számos lottó játészó (játékos) egyúttal totózik vagy tippmixezik is (fogad), az egyes játéktípusok között átfedések észlelhetők (*Ipsos* [2010]).

Hogyan lehet mégis elemezni a játékszokásokat a megfigyelések és az eladói interjúk vagy a piackutatás módszerein túl? A játékok feladása online rendszeren keresztül történik, azaz minden egyes feladás azonnal egy központi szerveren kerül rögzítésre, mely tartalmazza, hogy mikor, milyen játékokkal, mely tétekekkel és termé-

szetesen milyen számokkal játszott az adott játékos. Ráadásul a joker számként használt hatjegyű szám abban az esetben is szerepel az adatbázisban, ha az illető nem játszotta meg a joker játékot, ugyanis ennek hét karakterű változata a segédszelvény nyomdai sorszáma is egyben. (Lásd az 1. ábrát.) Így a névtelen játékosokat egy-egy önálló számsorral tudjuk azonosítani, mivel lehetőség van az adott azonosítóval több alkalommal való játszásra is. Továbbá, a vevő szerencsejáték-feladásainak végeztével a fizetendő összeg összesítéséhez a gép lezárja a tranzakciót, tehát az adatsorból azt is láthatjuk, hogy mely tranzakciók tartoztak a játékoshoz.

1. ábra. A nyomdai sorszám, mint ügyfél-azonosító



Forrás: Saját szerkesztés.

A jelenleg rendelkezésre álló informatikai rendszerünk, illetve adatbázisunk alapján is élhetünk a prediktív analitika által kínált lehetőséggel, miszerint a játékosainkat a segédszelvényük alapján beazonosítjuk és figyelemmel kísérhetjük játékszokásaikat, kapcsolódó játékaikat. Az elemzésünk során visszatérő tranzakcióknak az olyan vásárlásokat tekintjük, melyek esetén a szelvényt egyenél több alkalommal fordul elő az adatbázisban a vizsgált négy hónap alatt. A vevők tanulmányozásánál kiemelt figyelmet érdemel a problémás játékosok azonosítása, és játszási szokásaik vizsgálata a játékszenvedély kialakulásának megelőzése érdekében. A problémás kifejezés olyan mértéktelen vagy túlzott mértékű szerencsejáték-vásárlásra, -használatra utal, amely a személy életvitelét lényegesen befolyásolja (Balázs–Kun–Demetrovics [2009]).

2.1. Korlátok

Vizsgálatunkat egy szegedi lottózó négyhavi forgalmi adatai alapján végezzük el. A játékosok viselkedését a környezet is befolyásolhatja (például a városrész, ahol a lottózó elhelyezkedik, az üzlet típusa, atmoszférája), amit azonban jelen kutatásunkba nincs lehetőségünk beépíteni.

Az elemzésbe bevont játékosokra vonatkozóan szintén korlátozásokkal kell élnünk. A forgalmi adatok a lottózóban megvásárolt játékokra vonatkoznak, ily módon az interneten vagy telefonon keresztül játszó e körön kívül maradnak. A már említett azonosító minden játéktípus szelvényén szerepel a számsorsjátékot kivéve, így a sorsjeggyel játszókat nem tudjuk beazonosítani, a vizsgálatban ezek az ügyfelek sem szerepelnek. A nyomdai sorszám azonosítóként kezelésével további problémaként merül fel, hogy egyes játékosok eldobják a megvásárolt szelvényt és újjal (másik sorszámmal) térnek vissza, így a visszatérő játékosok, tulajdonképpen csak azokat a visszatéréseket jelentik, amely esetekben az eredeti szelvényt ténylegesen többször is felhasználják.

Praktikus cél, hogy az ügyfél teljes egészében beazonosítható legyen, személyes információkat szerezzünk róla. Előfordulhat, hogy a játszási szokásokra építve homogén csoportokat kapunk, azonban tagjaik demográfiai tényezők szerinti jellemzői eltérők, amit viszont a rendelkezésre álló tranzakciós adatok alapján nem tudunk megállapítani. Hasonlóképp, a sikeresebb ügymenetet és a piaci igények ismeretét szolgálná a vélemény-formálók mélyebb szintű beazonosítása, de e tekintetben problémaként merül fel (főként a nyertesek esetén) az anonimitás és a személyi jogok figyelembe vétele.

3. Primer kutatás

A játszási szokások vizsgálatához a Szerencsejáték Zrt. egy lottózójának 2010. január első munkanapjától 2010. április utolsó munkanapjáig terjedő forgalmazási adatait vettük figyelembe, jelenleg a 2010. év első harmada áll rendelkezésre forgalmi adatok szempontjából. A lottózó kiválasztásánál fontos szempont volt, hogy forgalmi nagyságát és játékos összetételét tekintve – beleértve a termékstruktúrát is – ne mutasson specialitásokat vagy lényegi eltérést az országos átlagadatokhoz mérten. A vizsgálat azokra a sikeres értékesítési tranzakciókra terjedt ki, melyek esetén létezik a már bemutatott, azonosításra szolgáló szelvényszám, így a kaparós sorsjegyes játéktípus nem szerepel benne. Az ismertetett szempontok figyelembe vételével az adattáblába 26 038 tranzakció került be, ami 8948 féle szelvényszámból, azaz ügyfélből tevődik össze.

3.1. A vizsgálat célja, hipotézisei és módszerei

Elemzésünk során múltbeli adatokat vizsgáltunk, majd szabályszerűségeket próbáltunk megfogalmazni a jövőbeli cselekvések meghatározásához. A tranzakciós

adatbázis alapján azt tanulmányoztuk, hogy a visszatérő tranzakciók értékesebbek-e, az egyes játéktípusok esetén milyen ezek aránya, illetve az akciókra kimutatható-e reagálás.

Feltételezésünk szerint az egy tranzakcióban elköltött összeg magasabb visszatérés esetén, kimutatható kapcsolat a játék típusa és a visszatérés között, illetve az akciók napokon magasabb az egy tranzakcióban elköltött összegek átlaga. A három kérdés megválaszolása érdekében független t -próbákat és keresztábla-elemzést alkalmaztunk SPSS 19.0 program segítségével.

A hipotézisek vizsgálatán túl kíváncsiak voltunk a vevők jellemzőire is. Szociodemográfiai változók híján mélyebb kutatásuk az RFM-analízis segítségével, csoportosításuk pedig a program döntési fa moduljával történt.

3.2. Az egy tranzakcióban elköltött összeg és a visszatérés vizsgálata

A mintában 9 543 tranzakciót tekintünk visszatérőnek, mivel ezen vásárlások esetén a szelvéyszám egynél több alkalommal fordul elő az adatbázisban a vizsgált négy hónap alatt. A maradék 16 495 tranzakciót nem tartjuk annak. A visszatérő tranzakciók esetén az átlagos szelvény ára (az egy tranzakcióban elköltött összegek átlaga) 1 010,14 Ft, a nem visszatérő tranzakcióknál pedig 595,02 Ft. A kétféle tranzakcióban kifizetett szelvényárak (egy tranzakcióban elköltött összeg) várható értékének összehasonlításához Wald-tesztet alkalmaztunk, mivel a két csoport esetén a szórásnégyzetek nem tekinthetők azonosnak ($p = 4,14 \cdot 10^{-24}$). A teszt eredménye szerint ($p = 2,66 \cdot 10^{-47}$) az egy tranzakcióban elköltött várható értékek nem tekinthetők azonosnak ötszázalékos szignifikanciaszinten. A sokasági átlagok különbségét tartalmazó konfidencia-intervallum alapján megállapítható, hogy a visszatérő tranzakciók esetén az egy tranzakcióban elköltött összeg átlaga szignifikánsan magasabb a nem visszatérőkhöz képest, azaz a visszatérő tranzakció értékesebbnek tekinthető a Szerencsejáték Zrt. szempontjából.

3.3. A játék típusa és a visszatérés vizsgálata

A vizsgálatunk során kíváncsiak voltunk arra, hogy az egyes játéktípusok esetén hogyan alakulnak a visszatérési arányok, azaz azt vizsgáltuk, hogy a játék típusa és a visszatérés függetlennek tekinthető-e egymástól. Khi-négyzet-tesztet alkalmazva megállapítható, hogy ötszázalékos szignifikanciaszinten a két változó nem tekinthető függetlennek egymástól ($p < 0,05$, $\chi^2 = 4613,431$). A játék típusa és a visszatérés közötti kapcsolatot vizsgálva, a Cramer-féle együttható alapján ($C = 0,421$), közepes erősségű kapcsolat mutatható ki a két változó között. Az 1. táblázatból látható, hogy

a visszatérő tranzakciók aránya Kenónál a legmagasabb, Góltotó esetén pedig 0 volt a vizsgált időszakban.

1. táblázat

*A visszatérés játéktípusonkénti megoszlása
(százalék)*

Játék típusa	Visszatérés		Összesen
	Nem visszatérő tranzakciók	Visszatérő tranzakciók	
Góltotó	100,0	0,0	100,0
Kenó	33,7	66,3	100,0
Lottó 5/90	48,5	51,5	100,0
Lottó 6/45	49,6	50,4	100,0
Lottó 7/35	60,3	39,7	100,0
Luxor	54,6	45,4	100,0
Puttó	70,6	29,4	100,0
Tippmix	93,7	6,3	100,0
Totó	99,8	0,2	100,0
Összesen	63,3	36,7	100,0

Forrás: Saját adatok.

50 százalékos feletti visszatérési arányszám a Kenó, az Ötös és a Hatos lottó tekintetében fordul elő. Tipikusan nem visszatérő játékok a Góltotó, a Tippmix és a Totó.

3.4. Az akciók hatása

A vizsgált időszak alatt 5 038 olyan tranzakció fordult elő, amelyek napján volt meghirdetett akció, 21 ezer tranzakció esetén pedig nem. Az akciós napokon az adott játékból egy meghatározott értékösszeg feletti vásárláskor a szelvények extra sorsoláson is részt vehettek, melynek nyereségének tárgyát pénzösszeg vagy személygépkocsi képezte. Az akció alatt létrejövő tranzakcióknál az egy tranzakcióban elköltött összeg átlagos értéke 754,16 Ft, míg az akciós időszakon kívüliek körében 745,44 Ft volt. Az akciós és akció nélküli napokon létrejövő tranzakciók esetén a szórásnégyzetek azonosnak tekinthetők ötszázalékos szignifikanciaszinten ($p = 0,112$), így a kifizetett szelvényárak várható értékének összehasonlításához független kétmintás t -próbát alkalmaztunk. A teszt eredménye szerint a két csoport egy tranzakcióban elköltött várható értékei is azonosnak tekinthetők ötszázalékos

szignifikanciaszinten ($p = 0,789$). Így az akció és akció nélküli időszakokban megvalósuló tranzakcióknál a szelvényárak várható értékei nem különböznek, azaz az akciónak nincs kimutatható szignifikáns hatása az egy tranzakcióban elköltött átlagos összeg tekintetében (ötszázalékos szignifikanciaszinten).

Kíváncsiak voltunk, hogy a teszt ugyanezt az eredményt hozza-e, ha azt a visszatérő és a nem visszatérő tranzakciókra külön-külön is elvégezzük. A visszatérő tranzakcióknál 1 776 tranzakció akció ideje alatt, 7 767 tranzakció pedig akció nélküli időszakban jött létre. Az akciós és akció nélküli napokon létrejövő tranzakciók esetén a szórásnégyzetek azonosnak tekinthetők ötszázalékos szignifikanciaszinten ($p = 0,423$), ezért a kifizetett szelvényárak várható értékének összehasonlításához itt is független kétmintás t -próba alkalmazható. Ennek eredménye szerint a két csoport egy tranzakcióban elköltött várható értékei azonosnak tekinthetők ($p = 0,816$) ötszázalékos szignifikanciaszinten ($p = 0,816$). A visszatérő tranzakcióknál nincs szignifikáns hatása az akcióknak. Ez esetünkben azt jelenti, hogy adott szelvényhez nem rendelnek több játékmezőt a játékosok az akció hatására.

A nem visszatérő tranzakciók esetén 3 262 tranzakció akció ideje alatt, 13 233 tranzakció pedig akció nélküli időszakban jött létre. Az akciós és akció nélküli napokon létrejövő tranzakciók esetén a szórásnégyzetek azonosnak tekinthetők ötszázalékos szignifikanciaszinten ($p = 0,465$), így a kifizetett szelvényárak várható értékének összehasonlításához szintén független kétmintás t -próba alkalmazható. A teszt eredménye szerint a két csoport egy tranzakcióban elköltött várható értékei is azonosnak tekinthetők ötszázalékos szignifikanciaszinten ($p = 0,645$), azaz a nem visszatérő tranzakcióknál sincs szignifikáns hatása az akcióknak az egy tranzakcióban elköltött átlagos összeg tekintetében.

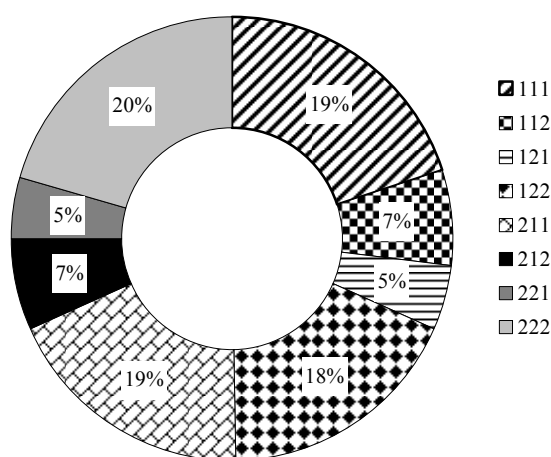
A vizsgált tranzakciók esetén tehát megfogalmazható, hogy a visszatérő tranzakciók értékesebbek, játéktípusonként eltér a visszatérő tranzakciók aránya, azonban az akcióknak nincs szignifikáns hatása az egy tranzakcióban elköltött átlagos összeg tekintetében. E kérdéseken túl a vevők jellemzése is célunk volt. Az ügyfélalapú vizsgálatokhoz és az ügyfélalapú adatbázis létrehozásához RFM-analízist alkalmaztunk, majd a további vizsgálatokat döntési fa alkalmazásával folytattuk le.

3.5. RFM-analízis

Az SPSS-szoftver RFM-analízis moduljának bemeneteként az ügyfél-azonosítót, a tranzakciók dátumát és a vásárolt mennyiséget (mely jelen esetben a megjátszott alapjáték számával azonosítható) kell átadni. Az aktualitás pontszámát a program az utolsó tranzakció óta eltelt idő, a gyakoriságét az adott időintervallum alatt létrejövő tranzakciók száma, a pénzügyi értékét pedig az egy tranzakcióban vásárolt mennyiségek alapján számítja ki. Elemzésünk során e három változóra építve, az osztálykö-

zök meghatározásánál (pontszámok kiosztásánál) az egymástól független lehetőséget választottuk (a tanulmány 1. fejezetében ismertetett lehetőségek közül), és a felosztás a medián alapján történt. Ennek eredményeként nyolcféle pontszám jött létre, amelyek egyben az ügyfelek csoportjait jelképezik. A 2. ábrán az általunk kapott nyolcféle pontszám előfordulásának relatív gyakorisága látható.

2. ábra. Az ügyfelek RFM-pontszám szerinti megoszlása
($N = 8\,848$)



Forrás: Saját adatok.

A létrejövő kódkombinációk egy-egy ügyfélcsoportként is felfoghatók. Az 111-es RFM-pontszámú ügyfelek olyan vásárlókat jelentenek, akik a kevésbé aktuális vevők közé tartoznak (az utolsó tranzakció óta eltelt idő hosszabb a 2-es kódú csoporthoz képest), nem túl gyakran hajtanak végre vásárlásokat, és alacsony pénzügyi érték rendelhető hozzájuk, azaz az általuk vásárolt mennyiség az alacsony osztályközbe sorolható. Másik példaként, a 212-es RFM-pontszámú vásárlók az aktuális, kevésbé gyakran vásárló, de magas pénzügyi értékkel rendelkező ügyfeleket jelentik.

A játékok szempontjából az lehet kulcsügyfél, aki aktuális ügyfele a vállalatnak, és gyakran játszik. Ilyen megközelítésben a fontos ügyfelek szegmensét a 222 és 221-es pontszámmal rendelkezők alkothatják. A 8048 játékosnak 20,5 százalékát teszik ki az aktuális, magas gyakorisággal játszó, magas pénzügyi értékű vevők, 4,5 százalékát pedig az aktuális, magas gyakorisággal játszó, alacsony pénzügyi értékűek. Azonban ezeket a csoportokat fontos lenne tovább vizsgálni, több információt szerezni róluk, hiszen a gyakran játszóknak között nemcsak stabil, rendszeresen visszatérő törzsvásárlók, hanem az esetlegesen problémás játékosok vagy játékszenvedélyvel rendelkezők is előfordulhatnak. A további vizsgálathoz a program döntési fa modulját alkalmaztuk.

3.6. A játékosok vizsgálata döntési fával

Elemzésünk során kíváncsiak voltunk arra, hogy a fontos és a nem fontos felosztás hogyan állítható elő az ügyfelek heti átlagos elköltött összegei, heti átlagosan megjárt szelvénytárhelyszámai és az egy szelvényben elköltött átlagos összegek alapján. Ez a csoportosítás (mint célváltozó) az RFM-analízisből származik, a 222 és 221-es pontszámú ügyfelek jelentik a fontos ügyfelek kategóriáját, míg a többi előforduló pontszám (111, 112, 121, 122, 211, 212) a nem fontosokét.

Az analízis az ügyfelek „csoportosítását” az „azonosító”, a „tranzakciók dátuma” és a „vásárolt mennyiség” változók alapján végezte el, a pénzügyi értéket is a vásárolt mennyiség alapján számította ki. A döntési fával való modellezés esetén így a magyarázó változók köre az RFM-analízishez képest az ügyfelek heti átlagos elköltött összegeivel és az egy szelvényben elköltött átlagos összeggel bővült. A két módszer esetén „közös” magyarázó változó a heti átlagosan megjárt szelvénytárhelyszám, ami a döntési fa két új változójával nem korrelál erős mértékben. A heti átlagosan megjárt szelvénytárhelyszám és a heti átlagos elköltött összeg között szignifikáns ($p = 3,3 \cdot 10^{-257}$), gyenge ($r_{xy} = 0,35$), a heti átlagosan megjárt szelvénytárhelyszám és az egy szelvényben elköltött összeg között nincs szignifikáns kapcsolat ($p = 0,159$) ötszázalékos szignifikanciaszinten ($p = 0,159$).

A döntési fa által létrejött csoportosítás jónak tekinthető, a kevésbé fontos csoportnak 77,7 százalékát a szoftver is a kevésbé fontos csoportba sorolta. A fontos csoport 79,9 százaléka került a döntési fa szerint létrehozott fontos csoportba, és összességében a döntési fa végeredménye szerint az ügyfelek 78,2 százaléka került a helyes csoportba. Ezt az eredményt a programon belül a CRT (classification and regression tree – osztályozási és regressziós fák elnevezésű) növekedési módszer segítségével tudtuk elérni, mely a szakirodalomban CART néven is ismert. A módszer lényege Lee et al. [2006] alapján az, hogy először az előrejelző változók (jelen esetben a heti átlagos elköltött összeg, a heti átlagosan megjárt szelvénytárhelyszám és az egy szelvényben elköltött átlagos összeg) azon értékeit (optimális vágási pontjait) kell megtalálni, melyek mentén az eljárás el tudja kezdeni a megfigyelési egységek csoportokba sorolását. Ezt követően az előrejelző változókra meghatározott vágási pontok közül a legalkalmasabbnak vélt érték szerint történik a felosztás. A létrejövő végpontokon (faágakon) belül addig kell ismétetni az ismertetett lépéssorozatot, amíg a stabil faszerkezet ki nem alakul.¹

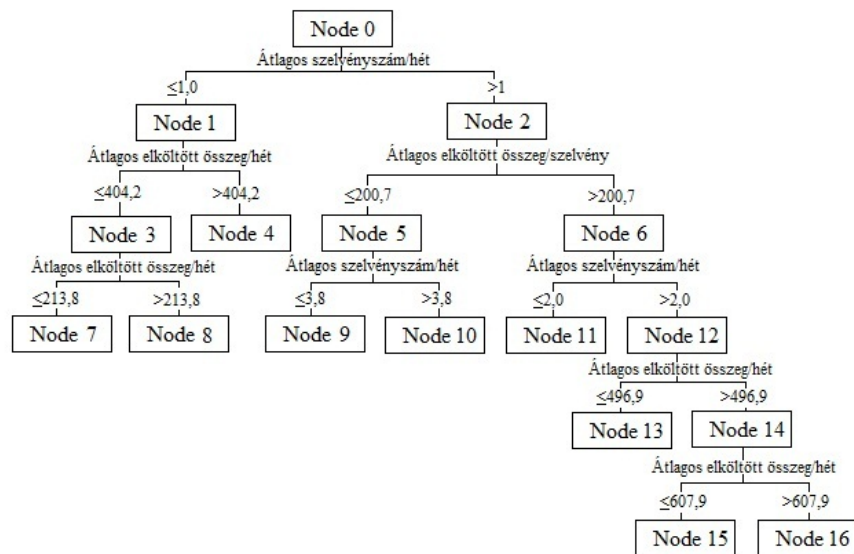
A 3. ábrán látható a döntési fa kimeneti ábrája, melyből leolvasható az osztályozási folyamat. Az első lépésben a heti átlagosan megjárt szelvénytárhelyszám alapján történt a vágás, az egyik ágon a heti átlagban egy alkalommal vagy annál kevesebb-

¹ A stabil faszerkezet olyan állapotot jelent, amely újabb, az előzőekben ismertetett lépéssorozatok ismételt elvégzésével sem vezet az első felosztáshoz képest eltérő eredményre (Dwyer-Holte [2007]).

szer, a másik ágon a heti átlagban egy alkalomnál többször játszóknak vannak. A heti átlagban legfeljebb egy alkalommal játszókon belül a következő vágás a heti átlagos kiadás összege alapján jött létre, az egyik ág a 404,2 Ft-ot vagy annál kevesebbet, a másik a 404,2 Ft-nál többet költőket tartalmazza, ami egyben egy végpontot is jelent (Node 4). A heti szinten átlagosan 404,2 Ft-ot vagy annál kevesebbet költőkön belül a heti átlagosan elköltött összeg alapján történt még egy vágás, így kaptuk a heti átlagban legfeljebb 213,8 Ft-ot (Node 7), illetve a 213,8 és legfeljebb 404,2 Ft között költők csoportját (Node 8).

A fa másik fő ágán az átlagban heti szinten egy alkalomnál többször játszóknak találhatók, ezen a részen a következő vágásra az egy szelvényben elköltött átlagos összeg alapján került sor. Létrejött az egy szelvényben átlagosan legfeljebb 200,7 Ft-ot és az annál többet költő ügyfelek ága. Az egy szelvényben átlagosan legfeljebb 200,7 Ft-ot költőkön belül még egy vágás történt az egy héten megjátszott szelvények átlagos száma szerint, így kialakult a heti szinten átlagban legfeljebb 3,8 (Node 9) és az annál több szelvényt játszóknak csoportja (Node 10).

3. ábra. A döntési fa kimeneti ábrája



Forrás: Saját adatok.

A fa heti szinten átlagban egy alkalomnál többször játszó és egy szelvényben átlagban 200,7 Ft-nál többet költők ágán (Node 6) belül a következő vágásra az egy héten megjátszott átlagos szelvénytárs alapján került sor. Megkülönböztethető a heti szinten átlagban 1-nél több alkalommal, de legfeljebb 2 szelvénytárs (Node 11) és a

2-nél több szelvényrel játszó ága. Az utóbbiak két részre oszthatók a heti átlagos elköltött összeg alapján: a legfeljebb 496,9 Ft-ot (Node 13) és az annál nagyobb összeget kiadók ágára, ami még szintén tovább bontódik az egy héten átlagosan elköltött összeg alapján. Így létrejön még két csoport: a heti szinten átlagban 496,9 Ft-nál többet, de legfeljebb 607,9 Ft-ot, illetve a 607,9 Ft-nál többet költőké. Összességében a fának kilenc végpontja van, ami kilenc vásárlói csoportot jelent.

Érdeemes megvizsgálni ezek cél-, azaz a fontosságváltozóhoz való viszonyulását is. (Lásd a 2. táblázatot.) Mindegyik esetén látható a csoportok elemszáma, annak mintán belüli megoszlása, a csoportokban a fontos ügyfelek aránya, elemszáma, illetve az utóbbi mintán belüli megoszlása. Az utolsó oszlopban található index azt fejezi ki, hogy az adott csoportban a fontos ügyfelek aránya hány százalék a mintabeli arányhoz képest.

2. táblázat

A döntési fa alapján létrejövő csoportok néhány jellemzője

Csoport (Node)	Csoport		Fontos ügyfelek		Fontos ügyfelek aránya (százalék)	Index (százalék)
	N	Megoszlás (százalék)	N	Megoszlás (százalék)		
11	166	1,9	136	6,1	81,9	327,6
13	270	3,0	178	8,0	65,9	263,6
16	2851	31,9	1474	65,9	51,7	206,7
10	128	1,4	51	2,3	39,8	159,3
15	155	1,7	43	1,9	27,7	110,9
9	378	4,2	98	4,4	25,9	103,7
4	2160	24,1	203	9,1	9,4	37,6
8	1180	13,2	49	2,2	4,2	16,6
7	1660	18,6	6	0,3	0,4	1,4
<i>Összesen</i>	<i>8948</i>	<i>100,0</i>	<i>2238</i>	<i>100,0</i>	<i>25,0</i>	<i>100,0</i>

Megjegyzés. A megoszlásértékek kerekítés miatt nem adják ki a 100 százalékot.

Forrás: Saját adatok.

Ilyen értelemben azok a csoportok lehetnek érdekesek, amelyeknél az index 100 százalék felett van (9., 15., 10., 16., 13. és 11. csoportok). Azonban érdemes azt is figyelembe venni, hogy a 10., 15. és 9. csoportokban a fontos ügyfelek aránya csak 50 százalék alatti, így a fontos ügyfeleket a 11., 13. és 16. csoportok jelentik.

A 11. csoportban a fontos ügyfelek aránya 81,9 százalék, ahol olyan játékosok találhatóak, akik hetente átlagosan egynél több, de legfeljebb kettő szelvényrel játszanak, és az egy szelvényben elköltött összeg átlaga 200,7 Ft-nál nagyobb. Ez a csoport

a heti szinten megvásárolt szelvények átlagos száma miatt tekinthető stabilnak, így „stabilan vásárlóknak” nevezhető.

A 13. csoportban, melynél a játékosok hetente átlagosan kettőnél több szelvényvel játszanak, egy-egy szelvényre átlagosan 200,7 Ft-nál többet költenek és a heti átlagos költség legfeljebb 496,9 Ft, a fontos ügyfelek aránya 65,9 százalék. Ez a heti szinten elköltött átlagos összegek maximuma miatt tekinthető stabilnak, ezért tagjait „stabilan költőknek” lehet hívni.

A 16. csoportban a fontos ügyfelek aránya 51,7 százalék. Itt a vevők hetente átlagosan kettőnél több szelvényvel játszanak, az egy szelvényre költött összegek átlaga 200,7 Ft-nál magasabb, és a heti átlagos költség 607,9 Ft-nál több. Közülük kerülhetnek ki a problémás játékosok, így a csoport elnevezése „lehetséges problémások” lehet.

3.7. A döntési fa eredménye és a játékok szerinti tipizálás

Kíváncsiak voltunk arra is, hogy a döntési fa alapján létrejött célcsoportokban (stabilan vásárlók, stabilan költők, lehetséges problémások) felfedezhető-e tipikus játékok. Minden vevő esetén ismert az egyes játékok relatív játszási gyakorisága, így a következőkben a célcsoportok relatív játszási gyakorisága várható értékeinek összehasonlítását végezzük el az egyes játékokra vonatkozóan, varianciaanalízis segítségével. A célközönség részét nem képező csoportokat összevontuk egy csoportba, ők alkotják a kevésbé fontos ügyfeleket. Ennek köszönhetően mindegyik csoport (stabilan vásárlók, stabilan költők, lehetséges problémások, kevésbé fontos ügyfelek) elemszáma 100-nál magasabb. A Levene-teszt eredménye alapján ötszázalékos szignifikanciaszinten a Góltotó változó esetén teljesül a varianciahomogenitás ($p = 0,162$), míg a többi változónál (játéknál) nem ($p < 0,05$). (Lásd a Függelék F1. táblázatát.) A Góltotóra végzett F -próba eredménye alapján ötszázalékos szignifikanciaszinten a relatív játszási gyakoriság várható értékei azonosnak tekinthetők az egyes csoportokban. (Lásd az F2. táblázatot.)

A többi játék esetén (Kenó, Ötöslottó, Hatoslottó, Skandináv lottó, Luxor, Puttó, Tippmix, Totó) a relatív játszási gyakoriságok várható értékeinek összehasonlítására Welch-tesztet végeztünk el. Eszerint ötszázalékos szignifikanciaszinten a relatív játszási gyakoriságok várható értékei egyik változónál sem tekinthetők azonosnak a döntési fa csoportjai tekintetében (mind a nyolc változó esetén a $p < 0,05$). (Lásd az F3. táblázatot.)

A csoportok játékok szerinti definiálását Post Hoc tesztek segítségével határoztuk meg. (Lásd az F4. táblázatot.) Akkor mondjuk, hogy egy játék egy csoportra tipikusan jellemző, ha az adott csoport esetén a relatív játszási gyakoriság várható értéke szignifikánsan magasabb a többinél (ötszázalékos szignifikanciaszinten). A Kenó és a Skandináv lottó változók tekintetében nem figyelhető meg ez a szabályszerűség. A

Puttó esetén a kevésbé fontos ügyfelek csoportja rendelkezik szignifikánsan nagyobb relatív játszási gyakoriság várható értékkel ($p < 0,05$) más csoportokhoz képest. Az Ötöslottónál a stabilan vásárlók és a stabilan költők várható relatív játszási gyakoriságértékei is szignifikánsan magasabbak a többiekénél (mindkét esetben a $p < 0,05$). A Hatoslottóra stabilan költők, míg a Luxor esetén a stabilan vásárlók relatív játszási gyakoriságának várható értéke a nagyobb ($p < 0,05$) más csoportokhoz viszonyítva. A Tippmix és a Totó esetén is a lehetséges problémások csoportjánál magasabb szignifikánsan a relatív játszási gyakoriság várható értéke (mindkét esetben a $p < 0,05$).

A tipizálási szabályunknak megfelelően a stabilan vásárlókra jellemzőbb a Luxor és az Ötöslottó, a stabilan költőkre a Hatoslottó, a lehetséges problémásokra pedig a Tippmix és a Totó játzsása.

4. Összegzés

Összegzésként az elemzésünkben megállapítható, hogy a vizsgált esetekben a visszatérő tranzakciók értékesebbek a nem visszatérő tranzakciókhoz képest, játéktípusonként különbözők a visszatérési arányok, azonban az akcióknak nincs kimutatható hatása. Az RFM-analízis segítségével sikerült meghatározni a fontos és kevésbé fontos ügyfelek körét, ezt a csoportosítást használtuk a döntési fa célváltozójának. Az utóbbi módszer segítségével az ügyfelek heti átlagos elköltött összegei, heti átlagosan megjárt szelvény számai és egy szelvényben elköltött átlagos összegei alapján elvégeztük a játékosok osztályozását. Három csoportban a fontos ügyfelek aránya ötven százalék felett van, mely arányszám több mint kétszeres a fontos ügyfelek egész mintabeli arányához képest. Ezeket a csoportokat játszási és költési jellemzőik alapján stabilan vásárlóknak, stabilan költőknek és lehetséges problémásoknak neveztük el. Felfedezhető náluk a többihez képest tipikusabb játékok: az első csoportnál a Luxor és az Ötöslottó, a másodikonál a Hatoslottó, a harmadikonál pedig a Tippmix és a Totó.

A prediktív analitika keretében alkalmazott módszerekkel (statisztikai tesztek, RFM-analízis, döntési fa) sikerült a játékosok jellemzését, tipizálását és ezáltal várható jövőbeli viselkedését leírni. Ennek köszönhetően a játékosok azonosításával megvalósítható az értékesítés biztonságos és felelős növelése, továbbá a problémás játékosok tekintetében prevenciós lépések tehetők meg, amelyek az érintettek felvilágosításán, megfelelő kezelési lehetőségek felé történő terelésén túl a későbbiekben akár a játékmennyiségük korlátozásáig vagy drasztikus esetben a játékból történő kizárásáig terjedhet. A prediktív analitika a játékszervezéstől eltérő üzletágakra szintén kiterjeszhető, mely általánosságban alkalmas az ügyfelek, a vevők elemzésére, értékelésére és várható viselkedésük leírására.

Függelék

F1. táblázat

Varianciahomogenitás-teszt a relatív játszási gyakoriságra a döntési fa csoportjaiban

Játék típusa	Levene-teszt mintán felvett érték	1. szabadságfok	2. szabadságfok	<i>p</i>
Góltotó	1,712	3	8944	0,162
Kenó	26,702	3	8944	0,000
Ötöslottó	299,105	3	8944	0,000
Hatoslottó	321,177	3	8944	0,000
Skandináv lottó	205,834	3	8944	0,000
Luxor	46,163	2	8675	0,000
Puttó	2134,634	3	8944	0,000
Tippmix	215,201	3	8944	0,000
Totó	34,795	3	8944	0,000

Forrás: Ez és a további táblázatok saját adatok alapján készültek.

F2. táblázat

A Góltotó játék ANOVA-táblája (a relatív játszási gyakoriságok döntésifa-csoportok szerinti összehasonlítása)

Szóródás oka	<i>SS</i>	Szabadságfok	Becsült szórásnégyzet	<i>F</i>	<i>p</i>
Belső	0,002	3	0,001	0,426	0,734
Külső	16,391	8944	0,002		
Összesen	16,393	8947			

F3. táblázat

Welch-teszt eredmények (a relatív játszási gyakoriságok döntésifa-csoportok szerinti összehasonlítása)

Játék típusa	A Welch-teszt mintán felvett érték	1. szabadságfok	2. szabadságfok	<i>p</i>
Kenó	24,794	3	594,026	0,000
Ötöslottó	88,510	3	543,689	0,000
Hatoslottó	555,141	3	519,252	0,000
Skandináv lottó	35,049	3	528,270	0,000
Puttó	426,212	3	582,559	0,000
Tippmix	430,696	3	1150,217	0,000
Totó	38,905	3	698,875	0,000
Luxor	12,983	2	436,927	0,000

F4. táblázat

A relatív játszási gyakoriság várható értékeinek páronkénti összehasonlítása Tamhane-tesztel – a csoportátlagok különbségei (I-J)

Játék	Fakategória (I)	Fakategória (J)			
		Nem fontos csoportok	11. csoport	13. csoport	16. csoport
Kenó	Nem fontos csoportok	–	–0,017	0,034	0,010
	11. csoport	0,017	–	0,051	0,027
	13. csoport	–0,034	–0,051	–	–0,024
	16. csoport	–0,010	–0,027	0,024	–
Ötöslottó	Nem fontos csoportok	–	–0,160	–0,216	–0,052
	11. csoport	0,160	–	–0,056	0,108
	13. csoport	0,216	0,056	–	0,163
	16. csoport	0,052	–0,108	–0,163	–
Hatoslottó	Nem fontos csoportok	–	–0,181	–0,337	–0,167
	11. csoport	0,181	–	–0,156	0,014
	13. csoport	0,337	0,156	–	0,170
	16. csoport	0,167	–0,014	–0,170	–
Skandináv lottó	Nem fontos csoportok	–	–0,002	0,055	0,064
	11. csoport	0,002	–	0,057	0,066
	13. csoport	–0,055	–0,057	–	0,009
	16. csoport	–0,064	–0,066	–0,009	–
Puttó	Nem fontos csoportok	–	0,231	0,250	0,233
	11. csoport	–0,231	–	0,019	0,002
	13. csoport	–0,250	–0,019	–	–0,017
	16. csoport	–0,233	–0,002	0,017	–
Tippmix	Nem fontos csoportok	–	0,156	0,150	–0,055
	11. csoport	–0,156	–	–0,006	–0,211
	13. csoport	–0,150	0,006	–	–0,205
	16. csoport	0,055	0,211	0,205	–
Totó	Nem fontos csoportok	–	0,028	0,031	–0,018
	11. csoport	–0,028	–	0,003	–0,046
	13. csoport	–0,031	–0,003	–	–0,049
	16. csoport	0,018	0,046	0,049	–
Luxor	Nem fontos csoportok	–	–0,058	–	–0,015
	11. csoport	0,058	–	–	0,043
	13. csoport	–	–	–	–
	16. csoport	0,015	–0,043	–	–

Irodalom

- ABRAHAMS, A. S. – BECKER, A. B. – SABIDO, D. – D’SOUZA, R. – MAKRIYIANNIS, G. – KRASNODEBSKI, M. [2009]: Inducing a Marketing Strategy for a New Pet Insurance Company Using Decision Trees. *Expert Systems with Applications*. Vol. 36. No. 2. pp. 1914–1921.
- BALÁZS H. – KUN B. – DEMETROVICS ZS. [2009]: A kóros játékszenvedély típusai. *Psychiatria Hungarica*. 24. évf. 4. sz. 238–247. old.
- BOSE, I. – CHEN, X. [2009]: Quantitative Models for Direct Marketing: A Review from Systems Perspective. *European Journal of Operational Research*. Vol. 195. No. 1. pp. 1–16.
- CHANG, H. C. – TSAI, H. P. [2011]: Group RFM Analysis as a Novel Framework to Discover Better Customer Consumption Behavior. *Expert Systems with Applications*. Vol. 38. No. 12. pp. 14499–14513.
- DUBIN, R. [1969]: *Theory Building*. The Free Press. New York.
- DWYER, K. – HOLTE, R. [2007]: Decision Tree Instability and Active Learning. In: *Kok, J. N. – Koronacki, J. – Lopez de Mantaras, R. – Matwin, S. – Mladenic, D. – Skowron, A.* (eds.): *ECML 2007*. Springer–Verlag. Berlin. pp. 128–139.
- IBM MAGYARORSZÁG [2012]: *Szoftverleírás*. <http://www-01.ibm.com/software/hu/analytics/spss/>
- IPSOS [2010]: *Játékszenvedély és felelős játékszervezés – Kutatási jelentés a Szerencsejáték Zrt. részére*. Budapest.
- KAMARKURA, W. A. – WEDEL, M. – DE ROSA, F. – MAZZON, J. A. [2003]: Cross-Selling Through Database Marketing: A Mixed Data Factor Analyzer for Data Augmentation and Prediction. *International Journal of Research in Marketing*. Vol. 20. No. 1. pp. 45–65.
- KAPLAN, A. [1964]: *The Conduct of Inquiry: Methodology for Behavioural Science*. Chandler Publishing. New York.
- KUTATÓPONT [2011]: *Szerencsejáték – személyes megkérdezésem alapuló tracking kutatás*. Budapest.
- LEE, T. S. – CHIU, C. C. – CHOU, Y. C. – LU, C. J. [2006]: Mining the Customer Credit Using Classification and Regression Tree and Multivariate Adaptive Regression Splines. *Computational Statistics & Data Analysis*. Vol. 50. No. 4. pp. 1113–1130.
- MALHOTRA, N. K. [2008]: *Marketingkutatás*. Akadémiai Kiadó. Budapest.
- MEDIÁN [2010]: *Az internethasználati szokások, az internetezők csoportjai – Szerencsejáték Zrt. által megrendelt kutatás*. Budapest.
- PREDICTIVE ANALYTICS WORLD [2012]: http://www.predictiveanalyticsworld.com/predictive_analytics.php
- SHMUELI, G. – KOPPIUS, O. [2010]: Predictive Analytics in Information Systems Research. *MIS Quarterly*. Vol. 35. No. 3. pp. 553–572.
- VERES Z. – HOFFMANN M. – KOZÁK Á. [2006]: *Bevezetés a piackutatásba*. Akadémiai Kiadó. Budapest.

Summary

The essence of predictive analytics is to model expected behavior for the future based on data from the past. The more needed data are available and the more accurately one knows connections,

between elements, the higher reliability can be achieved, consequently, the future reactions of the examined actors (players, customers, segments) can be determined. By the help of the specification and characterization of the key actors or players, actions can be developed for the prevention, preservation or influence of these customers. In this paper the authors try to describe predictive analytics, determine the gamblers with problems, and organize the factors which promote the anticipation of the possible changes in the behavior of the normally playing gamblers. By the means of the analytical methods of the SPSS software, the analysis, classification and evaluation of lottery customers are introduced.