

TDK-dolgozat

2023

NEUMANN JÁNOS EGYETEM
GAZDASÁGTUDOMÁNYI KAR



**TELEPÜLÉSMARKETING TÁMOGATÁSA MESTERSÉGES
INTELLIGENCIA ALAPÚ PREDIKCIÓVAL**

**SUPPORTING LOCAL MARKETING WITH ARTIFICIAL
INTELLIGENCE-BASED PREDICTIONS**

Békési Gergő Bendegúz

Konzulens: Dr. Tózsá István

Kézirat lezárásának dátuma: 2023. 11. 14.

TARTALOMJEGYZÉK

ABSZTRAKT.....	V
ABSTRACT	VI
1. A TELEPÜLÉSMARKETING FONTOSSÁGA ÉS ALAPJAI	I
1.1. JELENTŐSÉG ÉS FŐBB ELEMELK	II
1.2. STRATÉGIÁI MEGKÖZELÍTÉS	V
1.3. A TELEPÜLÉS, MINT TERMÉK	VI
1.4. A VÁROSVÉRSÉNY EGYÉS RÉSZÉI ÉS AZ AKCIÓTERV	VII
1.5. GYÁKORLATI, OPERATÍV MEGFONTOLÁSOK	VIII
1.6. ADATVÉZÉRELK MEGKÖZELÍTÉS	IX
2. ADAT- ÉS MÓDSZERTANI LEÍRÁS.....	XI
2.1. ÜZLETI CÉL ÉS MEGVALÓSÍTÁSI MÓDSZER, BEMENETEK ÉS KIMENETEK	XI
2.1.1. Bemenetek.....	XII
2.1.2. Kimenetek	XIV
2.2. A MÓDSZERTAN ALAPJÁT KÉPEZŐ MATEMATIKAI-INFORMATIKAI, GÉPI TANULÁSI MODELLEK ISMERTETÉSE	XV
2.2.1. Lineáris regresszió	XV
2.2.2. Polinomiális regresszió	XVI
2.2.3. Döntési fa.....	XVIII
2.2.4. Véletlen erdő.....	XX
2.2.5. Skálázás	XXI
2.2.6. Különböző gépi tanulási módszertanok algoritmikus összevetése	XXII
3. LINEÁRIS- ÉS POLINOMIÁLIS REGRESSZIÓS, VALAMINT DÖNTÉSI FA ÉS VÉLETLEN ERDŐ EREDMÉNYEK.....	XXIX
3.1. EREDMÉNYEK ISMERTETÉSE	XXIX
3.2. EREDMÉNYEK ÉRTÉKELÉSE	XXXII
3.3. EREDMÉNYEK TOVÁBBI JAVÍTÁSA	XXXIV
3.3.1. Hiperparaméter optimalizáció megválasztása	XXXV
3.3.2. Bayes-féle optimalizálás Fa-struktúrájú Parzen Becslővel.....	XXXVI
3.4. OPTIMALIZÁLT EREDMÉNYEK BEMUTATÁSA	XXXVIII
4. KONKLÚZIÓ	XXXIX
IRODALOMJEGYZÉK	XL

TÁBLÁZAT- ÉS ÁBRAJEGYZÉK

Táblázatjegyzék

1. táblázat: A településmarketing érintett elemeinek összefoglalása	4
2. táblázat: A bemeneti adatok összefoglalása	13
3. táblázat: A kimeneti adatok összefoglalása	14
4. táblázat: Algoritmusok összefoglalása	27

Ábrajegyzék

1. ábra: A településmarketing sokrétősége	2
2. ábra: Audittevékenységek.....	6
3. ábra: Stratégia felépítése.....	9
4. ábra: Döntéshozás támogatása	10
5. ábra: Döntéshozás modellezés segítségével	12
6. ábra: Lineáris regresszió interpretálható vizualizációja	15
7. ábra: Polinomiális regresszió interpretálható vizualizációja	17
8. ábra: Döntési fa interpretálható vizualizációja	18
9. ábra: Véletlen erdő döntési fákból álló interpretálható vizualizációja	20
10. ábra: Dálnok relatív hibái.....	28
11. ábra: Esztelnek relatív hibái.....	29
12. ábra: Zabola relatív hibái	29
13. ábra: Csíkszentdomokos relatív hibái	30
14. ábra: Kovászna relatív hibái	30
15. ábra: Az öt település átlagos relatív hibái	32
16. ábra: Bayes-féle optimalizáció.....	35
17. ábra: Az optimalizált meta-modell relatív hibái.....	37

ABSZTRAKT

A 21. század egyik fő fókuszpontja az adatalapú technológiák témaköre. Ezek azok az innovációk, amelyek rendkívül sok tématerületen indukálnak fejlődést, vagy aknáznak ki olyan potenciált, amelyre korábban az emberi gondolkodás nem volt képes. Nincs ez másképp a marketing, azon belül pedig a településmarketing terén sem. Ahhoz, hogy egy település jövőjéért felelős döntéshozó meg tudja állapítani, hogy melyik döntésének milyen hatása volt az adott település lakosainak elégedettségére és hogy a jövőben milyen utakat érdemes választania, elengedhetetlen az adatalapú technológiák használata.

Munkám során statisztikai hivataloktól igényelhető és nyíltan elérhető adatok alapján végeztem becsléseket arra vonatkozóan, hogy milyen az elégedettség 1-1 településen sikeresség, tökevonás fejlesztés, értékek, élhetőség, turizmus és stratégia szempontjából. Lineáris regressziós, polinomiális regressziós, döntési fa és véletlen erdő gépi tanuló modelleket hasonlítottam össze standardizálás, normalizálás és nyers esetben is. Ezek tanítását és tesztelését keresztvalidációs technikával végeztem. Legjobb, ensemble alapú gépi tanuló modellem átlagosan kevesebb, mint 12 %-os hibával képes megjósolni az egyes települések elégedettségi mutatóit még rendkívül alacsony, 4 településes tanító adathalmaz esetén is.

Munkámban először röviden áttekintem a településmarketing dolgozathoz kapcsolódó főbb aspektusait, majd kitérek az egyes adatok, Dálnok, Esztelnek, Zabola, Csíkszentdomokos és Kovászna statisztikai, szabadon elérhető és felmért elégedettségi adataira. Ezt követően az előbbi két adattípust véve bemenetként, az elégedettséget kimenetként gépi tanuló modelleket tanítok és értékelek ki, valamint javítok ensemble learning és hiperparaméter optimalizáció segítségével. Munkám során újszerű megközelítésben eszközök tudok biztosítani településvezető döntéshozók számára a hatékony településmarketing érdekében. Az ezen a téren alkalmazott prediktív modellezés gépi tanulás segítségével teljesen újszerű megközelítés. A kicsiny mintaszám ellenére elért jó pontosság a számos különböző kipróbált gépi tanuló modellnek, valamint a Fa-struktúrájú Parzen Becslővel támogatott ensemble learningnek köszönhető. Ezen módszerek együttes alkalmazása egy új megközelítés.

ABSTRACT

One of the main focal points of the 21st century is the realm of data-driven technologies. These are the innovations that induce progress in a wide range of fields or harness potential that was previously beyond human cognition. This holds true in the field of marketing, particularly in the domain of local marketing. In order for decision-makers responsible for the future of a destination to determine the impact of their decisions on the satisfaction of the residents of that destination and to identify future paths worth pursuing, the use of data-driven technologies is essential.

During my work, I conducted estimations based on data that can be requested from statistical offices and is openly accessible. These estimations revolved around the satisfaction, success in attracting capital, development, values, livability, tourism, and strategy aspects of individual destinations. I compared linear regression, polynomial regression, decision tree, and random forest machine learning models in both standardized, normalized, and raw cases. I performed their training and testing using cross-validation techniques. My best ensemble-based machine learning model is capable of predicting the satisfaction indices of individual destinations with an average error of less than 12%, even in the case of an extremely small training dataset consisting of only four destinations.

In my work, I first briefly overview the main aspects related to local marketing, then I turn to the individual data, Dálnok, Esztelnek, Zabola, Csíkszentdomokos, and Kovászna's statistical, freely available, and surveyed satisfaction data. Taking the former two types of data as input, and satisfaction as output, I then teach and evaluate machine learning models, as well as improve them with ensemble learning and hyper-parameter optimization. During my work, I can provide a novel approach for decision-makers in the interest of effective local marketing. The predictive modeling applied in this field with the help of machine learning is a completely novel approach. The good accuracy achieved despite the small sample size is thanks to the many different tried machine learning models, as well as the Tree-structured Parzen Estimator supported ensemble learning. The combined application of these methods is a new approach.

1. A TELEPÜLÉSMARKETING FONTOSSÁGA ÉS ALAPJAI

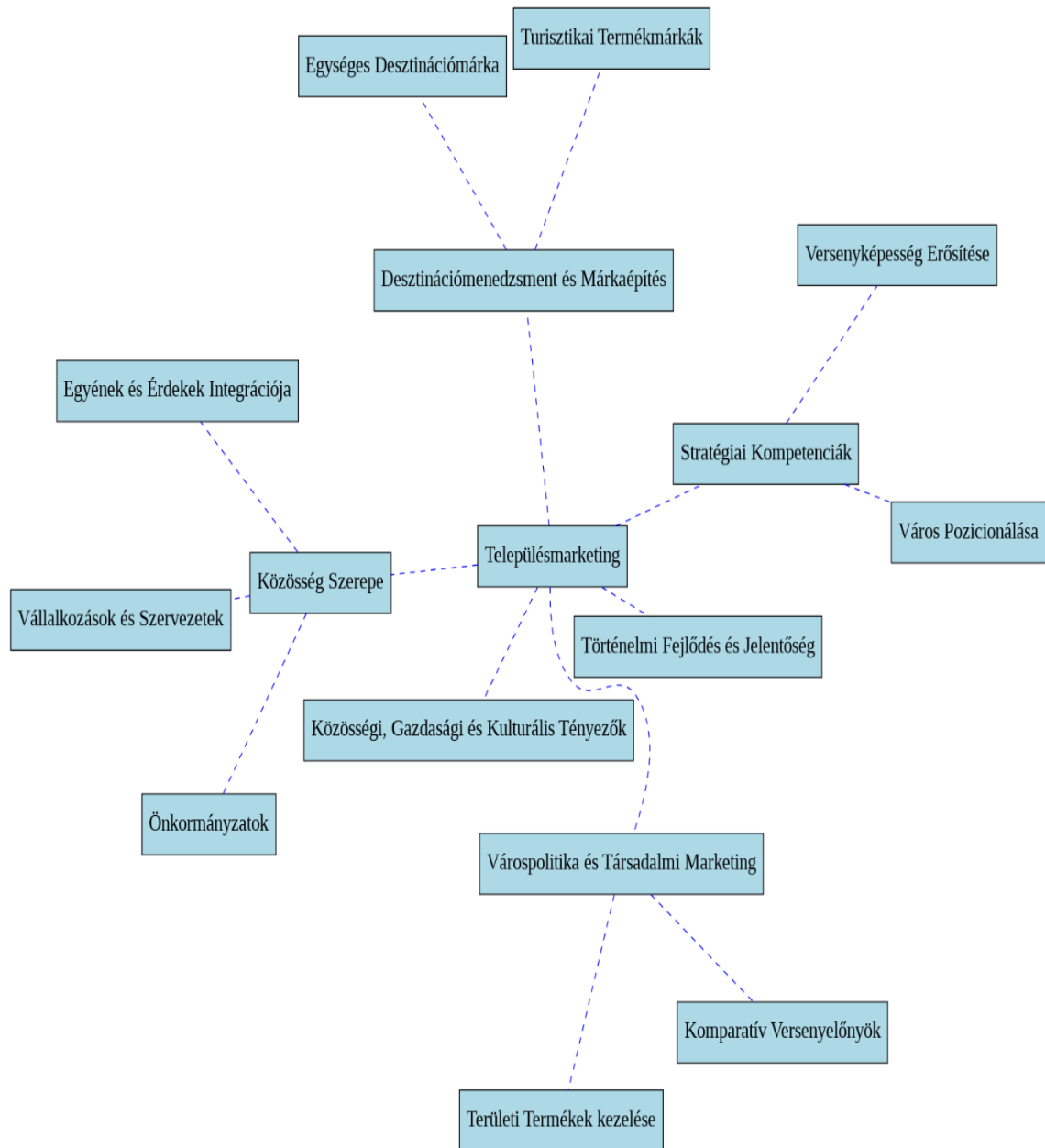
A településmarketing jelentőségének és történelmi fejlődésének megértése kulcsfontosságú a városfejlesztésben, különös tekintettel a várospolitikai szerepére. A településmarketing, mint a társadalmi marketing egy speciális ága, a komparatív versenyelőnyök kutatására, megvalósítására és kommunikációjára összpontosít, miközben a területi terméket egy komplex "bizalmi áruként" kezeli. Ez a megközelítés a marketing fejlődésének fontos állomása, amely a területfejlesztés hatékony eszközévé avanzsálódott (Puczkó, & Tózsá, 2017; Piskóti, 2022).

Ez a fajta marketing nem csupán az önkormányzatok feladata, hanem az egész közösség érdekében kell, hogy hasson, integrálva az egyes vállalkozások, szervezetek és egyének érdekeit. Ennek érdekében a településmarketing a társadalmi marketing egyedi kombinációjára támaszkodik, amely magában foglalja a kutatást, a célképzést, a stratégiaalkotást és az eszközök megvalósítását. Ezen belül a település imázsa kiemelkedő szerepet kap, mivel ez befolyásolja az információszerzési folyamatokat, az értékrendek kialakulását, valamint a terület valódi használatát és értékelését (Ma, de Jong, Hoppe, & de Bruijne, 2021; Nagy, Piskóti, Dankó, Molnár, & Marien, 2020; Jeney, Kulcsár, & Tózsá, 2013).

A marketing stratégiai kompetenciáinak jelentősége abban áll, hogy alkalmasak a várost pozicionáló, versenyképességét erősítő programok kialakítására. Ezek közé tartoznak az integrálóképes nemzetközi cégek, a kooperációkészséges közép- és kisvállalkozások, valamint a helyi egyetemek, kutatóintézetek és civilszervezetek. Ezek a szereplők hozzájárulnak a város vonzerőjének és identitásának erősítéséhez, miközben a regionális hatású programokat is elősegítik (Kuzior, Sobotka, Filipenko, & Kuzior, 2019; Garamhegyi, & Révész, 2000; Búzás, Kállay, & Lengyel, 2003; Gajzágó, 2019).

A desztinációmenedzsment, különösen a márkaépítés, szintén meghatározó szerepet játszik. Magyarországon ez a terület korábban elhanyagolt volt, ami a desztinációmarketing kulcsfontosságú része. A márkaépítés során figyelembe kell venni az egységes desztinációmárka és a turisztikai termékmárkák összehangolását. Egy jó példa erre a "tokaji bor", amely nem csak egy fontos agrártermék, hanem meghatározó országimázs-formáló elem is, rámutatva a helyi termékek jelentőségére a térség gazdasági és társadalmi fejlődésében. (Kavaratzis, & Hatch, 2021; Piskóti, Hidvéginé Molnár, Pataki, Schumpler, & Gulyás, 2007; Nagy, Piskóti, Molnár, Gulyásné Kerekes, & Karajz, 2015)

Tehát, a településmarketing és a várospolitikai szerepe a városfejlesztésben összetett és sokrétű. A közösségi, gazdasági és kulturális tényezők együttes fontossága hangsúlyozza, hogy ezek a területek hogyan járulnak hozzá a városok fejlődéséhez és prosperitásához. Ennek összefoglalása figyelhető meg az 1. ábrán.



1. ábra: A településmarketing sokrétűsége (Forrás: saját szerkesztés)

1.1. JELENTŐSÉG ÉS FŐBB ELEMEK

A településmarketing fogalmának és annak fő elemeinek megértése elengedhetetlen a városfejlesztés és a helyi gazdaság fejlődésének szempontjából. A településmarketing a helyi lakosság elégedettségét és jólétét helyezi előtérbe, és ezeket az érdekeket mindig szem

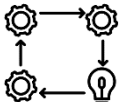
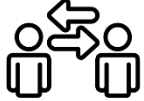
előtt tartja a marketing fejlesztési és kommunikációs tevékenységek során. Ez a megközelítés nem csupán a termékek és szolgáltatások eladására korlátozódik, hanem egy átfogó koncepciót képvisel, amely magában foglalja a régió és a település vezetésének, intézményeinek piacorientált stratégiai és operatív tervezését, koordinálását, valamint ennek realizálását és ellenőrzését. Ebben az összefüggésben a településmarketing nem egyenlő a településfejlesztéssel vagy a befektetőcsalogatással, hanem egy szélesebb kontextusban értelmezendő, ahol a település, mint "termék" az általa kínált adottságok, értékek, lehetőségek és élmények összessége (Stănciulescu, 2009; Puczkó, & Tózsá, 2017; Rechnitzer, 1995).

A településmarketingnek döntő szerepe van a turizmusban is. A turisták, mint a településmarketing egyik fő célcsoportja, különös figyelmet igényelnek. A településeknek meg kell érteniük, hogyan kommunikálhatnak hatékonyan a potenciális látogatókkal, és hogyan prezentálhatják a városukat vagy régiójukat úgy, hogy az vonzó legyen a turisták számára. Ez magában foglalja a helyi attrakciók, kulturális és történelmi értékek, valamint a vendégszeretet és egyéb szolgáltatások kiemelését (Piskóti, Hidvéginé Molnár, Pataki, Schumpler, Gulyás, 2007; Piskóti, 2017; Asprogerakas, 2006).

A településmarketing stratégiája összetett és sokrétű. A hatóságoknak és intézményeknek olyan "terméket" kell kialakítaniuk, amely megfelel a polgárok, vállalkozók, turisták és befektetők igényeinek és elvárásainak. Ez azt jelenti, hogy a városi funkciók ellátását és a "várostermék" kialakítását a felhasználók igényei és várakozásai határozzák meg. A településmarketing kifejezés használata a városmarketing mellett arra utal, hogy a marketing szükségessége független a település méretétől. A nagyobb városok több erőforrással rendelkeznek a marketingmodell komplexebb alkalmazására, míg a kisebb települések kisebb ráfordításokkal, néhány területre koncentrálnak érhetnek el sikereket. A településmarketing célközönségei nem csupán a helyi lakosság vagy a turisták, hanem azok a személyek és szervezetek is, akik döntéseikkel, például a kormányzati támogatásokkal vagy infrastrukturális fejlesztésekkel, befolyásolják a terület sikeres működését (Beluszky, & Győri 1999; Puczkó, & Tózsá, 2017; Bonakdar, & Audirac, 2020).

Így elmondható, hogy a településmarketing egy átfogó és sokrétű tevékenység, amely nem korlátozódik pusztán a termékek vagy szolgáltatások értékesítésére. A turizmusban betöltött szerepe mellett a helyi közösség és a szélesebb társadalom igényeinek és várakozásainak kielégítésére is összpontosít, miközben figyelemmel kíséri és reagál a változó piaci és társadalmi környezetre. Ezt összegzi az 1. táblázat.

1. táblázat: A településmarketing érintett elemeinek összefoglalása (Forrás: saját szerkesztés)

Fő Komponensek és Célok	Lakosság elégedettsége és jóléte	
	Turizmus és turisták vonzása	
	Városi és régiós prezentáció	
	Vállalkozók, befektetők és egyéb érdekelt felek igényeinek kielégítése	
Stratégiai és Operatív Tevékenységek	Piacorientált tervezés és koordináció	
	Realizálás és ellenőrzés	
Külső és Belső Hatások	Piaci és társadalmi környezet változásai	
	Kormányzati támogatások és infrastrukturális fejlesztések	

1.2. STRATÉGIAI MEGKÖZELÍTÉS

A településmarketing stratégiai elemei – audit, szegmentáció, imázs, és kommunikáció – döntő fontosságúak a települések fejlődésében és vonzerejének növelésében. Ezek az elemek együttesen alkotják meg a település átfogó marketingstratégiáját, amelynek célja a helyi értékek és lehetőségek hatékony bemutatása és értékesítése (Braun, 2008; Tózsza, 2014).

Az érték audit kulcsfontosságú a településmarketingben, mivel alapvetően határozza meg a település aktuális állapotát és potenciálját. Az audit során felméri a település erősségeit, gyengeségeit, lehetőségeit és fenyegetéseit. Ez a folyamat magában foglalja a település gazdasági, társadalmi, kulturális, és környezeti értékeinek, valamint infrastrukturális adottságainak átfogó elemzését. Az érték audit eredményei segítik a döntéshozókat abban, hogy reális képet kapjanak a település jelenlegi helyzetéről, és ez alapján hatékonyan tervezhessék meg a további fejlesztéseket és marketingtevékenységeket (Tózsza, 2014; HBH Stratégia és Fejlesztés Kft. & Collective-Intelligence Kft., 2018; Yezhova, 2020).

A település auditálása során a fő hangsúly az értéken, imázson és tevékenységeken van. Az érték audit segít meghatározni a település legfőbb vonzerejét és jellegzetességeit, amelyek alapján a települést a potenciális látogatók és befektetők számára vonzóvá tehetik. Az imázs auditálása során a település jelenlegi megítélését vizsgálják, beleértve a külső percepciókat és a helyi lakosok nézőpontját is. A tevékenységek auditálása során pedig a már meglévő és potenciális turisztikai, kulturális, és gazdasági tevékenységeket értékelik (Pannyor, Györiné Kiss, & Gulyás, 2011; Yezhova, 2020).

A település, mint termék megközelítés azt jelenti, hogy a települést egyfajta "termékként" kezelik a marketingstratégiában. Ez magában foglalja a település egyedi tulajdonságainak, kulturális és természeti értékeinek, valamint gazdasági potenciáljának kiemelését és értékesítését. Ennek célja, hogy a település vonzóvá váljon a turisták, befektetők, és a helyi lakosok számára egyaránt (Tózsza, 2014; Kavartzis, 2007).

Megállapítható tehát, hogy a településmarketing stratégiájának sikere nagymértékben függ az érték audit, a szegmentáció, az imázs, és a kommunikáció hatékony összehangolásától. Ezen elemek együttes alkalmazásával a települések javíthatják vonzerejüket, növelhetik gazdasági és kulturális potenciáljukat, és erősíthetik pozíciójukat a versenyképes piacon. Ezt az audittevékenységet emeli ki a 2. ábra:



2. ábra: Audittevékenységek (Forrás: saját szerkesztés)

1.3. A TELEPÜLÉS, MINT TERMÉK

A települések értékesítése a marketing egy speciális ágazata, ahol a település maga a termék. Egyedi jellegzetességei, kulturális, történelmi és gazdasági attribútumai adják meg a település "termékének" különlegességét. Ennek eredményeként a települések marketingje sokrétű és komplex feladatot jelent, amely egyedi stratégiákat igényel (Puczkó, & Tózsza, 2017; Boisen, Terlouw, Groote, & Couwenberg, 2018).

Egy város vagy település értékesítésének alapvető célja, hogy vonzóvá tegye magát a potenciális lakosok, befektetők, turisták és üzleti vállalkozások számára. Ez a folyamat magában foglalja a település egyedi vonásainak kiemelését, mint például a történelmi örökség, a kulturális élet gazdagsága, az oktatási és munkalehetőségek minősége, valamint a helyi gazdasági környezet vonzereje (Cassel, 2008).

A települések értékesítése több dimenzióban is megközelíthető. Ezek között kiemelkedik a turizmus, az üzleti lehetőségek, az oktatás, a kulturális és sportesemények, valamint a lakhatási és életminőségi szempontok. Mindegyik terület sajátos stratégiát és megközelítést igényel, amely az adott célközönség igényeire és elvárásaira épül (Jeney, Kulcsár, Tózsza, 2013; Garamhegyi, & Révész, 2000; Audretsch, Link, & Walshok, 2015; Andersson, & Henrekson, 2015).

Például a turisztikai marketing esetében a cél a város különleges turisztikai vonzerejének kiemelése, mint például történelmi látnivalók, gasztronómiai élmények vagy természeti szépségek. Az üzleti marketing esetében pedig a helyi gazdaság erősségei, mint például

az adott iparági specializációk vagy a kedvező befektetési környezet kerülnek előtérbe (Štetić, Šimičević, Pavlović, & Stanić, 2014).

A városverseny egyre fontosabbá válik a globalizálódó világban, ahol a települések versenyeznek egymással a befektetések, turisták, tehetséges munkaerő és nemzetközi elismerés megszerzéséért. A sikeres városok olyanok, amelyek képesek kiemelni és hatékonyan kommunikálni egyedi értékeiket, erősségeiket, és ezáltal vonzó célponttá válnak a különböző érdekelt felek számára (Harmaakorpi, Kari, & Parjanen, 2008).

A városok sikere gyakran mérhető a lakosok életminőségének javulásán, a gazdasági növekedésen, a turizmus növekedésén, vagy akár az oktatási és kulturális lehetőségek bővülésén keresztül. Mindez a hatékony és jól megtervezett városmarketing stratégiák eredményeként adódik.

1.4. A VÁROSVERSENY EGYES RÉSZEI ÉS AZ AKCIÓTERV

A városverseny jelentőségének elemzése a településmarketing kontextusában különös figyelmet érdemel. A városok közötti verseny egyre inkább meghatározó szerepet tölt be a globális gazdaságban, mivel a települések versenyeznek a befektetésekért, tehetséges munkaerőért, turistákért és nemzetközi elismerésért. A sikeres települések jellemzőit vizsgálva feltűnik, hogy ezek a városok gyakran rendelkeznek erős gazdasági alappal, hatékony vezetéssel, jó infrastruktúrával és magas szintű életminőséggel. Ezek a tényezők hozzájárulnak a népességmegtartó képességhez, ami kritikus a városok hosszú távú fenntarthatóságához (Majláth, Reicher, Pecze, & Kolnhofer-Derecskei, 2022; Herrschel, 2013).

A városi sikerek mérhetősége több dimenzióban értelmezhető. A jövedelemtermelés, munkanélküliség szintje és a környezeti tisztaság mind meghatározó indikátorok. Ezek a mutatók segítenek megérteni egy város gazdasági és szociális helyzetét, és lehetőséget biztosítanak a várospolitikusoknak és tervezőknek, hogy értékeljék és javítsák a városi életminőséget (Komasi, Hashemkhani Zolfani, Prentkovskis, & Skačkauskas, 2022).

Egy hatékony településmarketing akcióterv létrehozása érdekében a városoknak összetett stratégiákat kell kidolgozniuk, amelyek figyelembe veszik a fenti tényezőket. Egy ilyen terv tartalmazza a célközönség meghatározását, az egyedi vonzerők és lehetőségek kiaknázását, valamint a helyi közösségek bevonását. A cél az, hogy egy vonzó, élhető és fenntartható városi környezetet alakítsanak ki, amely elősegíti a gazdasági növekedést és javítja az életminőséget (Hegyi, 2007; Lafuente, Araya, & Leiva, 2022).

Így, a városok közötti verseny és a településmarketing elengedhetetlen a modern városi környezetben. A sikeres városok jellemzői és a városi sikerek mérhetősége központi

elemek a településmarketingben, melyek segítenek a városoknak a versenyképesség javításában és a fenntartható növekedés biztosításában.

1.5. GYAKORLATI, OPERATÍV MEGFONTOLÁSOK

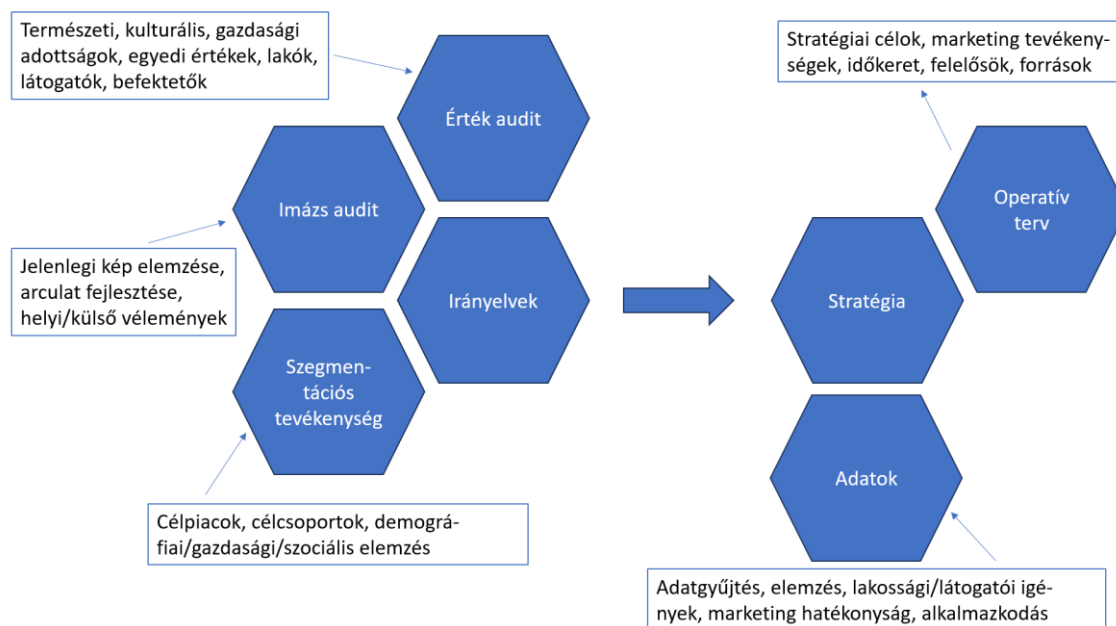
Operatív téren a már említett érték audit a településmarketing egyik kulcsfontosságú eleme, amely lehetővé teszi a városok számára, hogy alaposan felmérjék és kiaknázzák természeti, kulturális és gazdasági adottságaikat. Ez az audit segít meghatározni, hogy egy település milyen egyedi értékekkel rendelkezik, amelyek vonzóvá tehetik azt a lakók, látogatók és befektetők számára. A település szegmentációs tevékenysége ehhez kapcsolódik, mivel ennek során meghatározzák a potenciális célpiacokat és célcsoportokat. Ez a folyamat magában foglalja a demográfiai, gazdasági és szociális tényezők elemzését, hogy a város pontosan megérthesse, mely csoportokat kellene megcéloznia a marketing tevékenysége során (Piskóti, 2022; Puczkó, & Tózsza, 2017; Yezhova, 2021).

Továbbá a települések imázsának auditálása és arculatának kialakítása szintén létfontosságú. Ez magában foglalja a város jelenlegi képének és megítélésének elemzését, valamint azoknak a tényezőknek az azonosítását, amelyek javíthatnak ezen a képen. Az imázs audit során a városoknak figyelembe kell venniük a helyi közösségek nézeteit, valamint a külső érdekeltek és a látogatók perspektíváit. Ezen információk birtokában a városok képesek lesznek egy erős, vonzó és megkülönböztető arculatot kialakítani (Cevallos, 2022; Vanolo, 2017, Tózsza, 2014).

A településmarketing stratégiájának operatív tervének kidolgozása egy további fontos lépés, amely magában foglalja a fent említett elemzések alapján meghatározott stratégiai célok megvalósítását. Ez a terv részletesen leírja a különböző marketing tevékenységeket, a megvalósítás időkeretét, a felelős személyeket és a szükséges forrásokat. Egy jól megtervezett operatív terv segít a városnak koordinált és hatékony módon megvalósítani marketing céljait (Stănciulescu, 2009; Deffner, Karachalis, Psatha, Metaxas, & Sirakoulis, 2020).

Az említetteken túl, a településmarketing és adatok kapcsolata elengedhetetlen. Az adatgyűjtés és elemzés lehetővé teszi a városok számára, hogy megértsék a lakosság és a látogatók igényeit, valamint nyomon kövessék a marketing tevékenységek hatékonyságát. Az adatok segítségével a városok jobban alkalmazkodhatnak a változó körülményekhez és finomíthatják marketing stratégiáikat a jövőben (Sgambati, & Carvalho, 2023).

Ezek az elemek összekapcsolva képezik a településmarketing stratégia alapját, amely lehetővé teszi a városok számára, hogy versenyképesek és vonzóak maradjanak a globális piacon. Ezt foglalja össze a 3. ábra:



3. ábra: Stratégia felépítése (Forrás: saját szerkesztés)

1.6. ADATVEZÉRELT MEGKÖZELÍTÉS

Az érintett adatgyűjtés és -elemzés szerepe tehát a településmarketing területén rendkívül fontos. A mai adatvezérelt világban a településeknek olyan stratégiákat kell kidolgozniuk, amelyek megfelelően kihasználják az elérhető adatokat a döntéshozatali folyamatokban. A pontos és releváns adatok gyűjtése lehetővé teszi a városok számára, hogy mélyreható betekintést nyerjenek a lakosság igényeibe, preferenciáiba és viselkedési mintáiba. Ez nemcsak az aktuális helyzet megértését segíti elő, hanem prognózisokat is lehetővé tesz a jövőbeli trendekre vonatkozóan (Ye, Fan, Miao, He, 2022; Jeney, Kulcsár, Tózsza, 2013).

Az adatalapú döntéshozatal a településmarketing stratégiákban a hatékonyság és a célzott megközelítések felé vezet. A városok, amelyek adatokat használnak a stratégiai döntéseik meghozatalához, képesek jobban személyre szabni a marketingkampányaikat, optimalizálni a költségvetést és elérni a kívánt célcsoportokat. Az adatelemzés segítségével a városok megérthetik, mely marketingcsatornák és üzenetek a leginkább hatékonyak, és így jobban fókuszálhatnak az erőforrásaikra (Bibri, 2021; Jeney, Kulcsár, Tózsza, 2013).

Számos előnye van az adatalapú megközelítésnek a településmarketingben. Az egyik legnagyobb előny a megnövekedett pontosság és relevancia, amelyet az adatok biztosítanak. Az adatok lehetővé teszik a városok számára, hogy jobban megértsék és kiszolgálják a lakosságukat, miközben növelik a marketingkampányok hatékonyságát. Emellett az adatalapú

megközelítések segítenek a költséghatékony döntések meghozatalában, mivel lehetővé teszik a városok számára, hogy pontosabban célzott és hatékony marketingtevékenységeket végezzenek (Goovaerts, Van Biesbroeck, & Van Tilt, 2014; Gupta, & Chandra, 2020).

Mindazonáltal vannak kihívások is ebben a megközelítésben. Az adatok gyűjtésének és elemzésének folyamata bonyolult és időigényes lehet, és megköveteli a megfelelő technológiai infrastruktúra és szakértői tudás meglétét. Ezenkívül a magánélet és az adatvédelem kérdései is kiemelt figyelmet igényelnek, mivel a városoknak biztosítaniuk kell az összegyűjtött adatok biztonságos és etikus kezelését (Wang, Miwa, Morikawa, 2020).

A településmarketing gyakorlati alkalmazásában az adatalapú megközelítés integrálása lehetővé teszi a városok számára, hogy dinamikusabb és interaktívabb marketingstratégiákat hozzanak létre. Ez magában foglalhatja a közösségi média, mobilalkalmazások és egyéb digitális platformok használatát, amelyek segítenek a városoknak a közönséggel való kapcsolattartásban, az üzenetek személyre szabásában és a városi életminőség javításában (Malecki, 2017).

Tehát az adatgyűjtés és -elemzés, valamint az adatalapú döntéshozatal kulcsfontosságú a modern településmarketing stratégiákban. Ezek a módszerek lehetővé teszik a városok számára, hogy hatékonyabban kommunikáljanak a lakosokkal és látogatókkal, miközben növelik marketingtevékenységeik hatékonyságát és költséghatékonyságát. Az adatalapú megközelítések előnyei jelentősek, de megfelelő figyelmet kell fordítani a kihívások kezelésére is.

Ez a munka is ezt helyezi a középpontba. Ha egy városvezető vagy felelős döntéshozó meg tudja becsülni bizonyos statisztikai adatok alapján egy település lakosságának elégedettségét, akkor ez jelentős eszköz a kezében. Ennek oka, hogy így olyan döntéseket tud hozni a jövőben, amelyek növelik az elégedettséget az adott településen, így lehetséges lesz a hatékony forrásallokáció megvalósítására. Ehhez azonban szükséges egy prediktív modell felépítése, amely képes az elérhető adatok segítségével, a nem elérhető elégedettségi adatok predikciójára. Erre fókuszál a következő fejezet, a 4. ábra pedig ennek folyamatát szemlélteti. A döntéshozó ezen folyamat segítségével szimulálhatja jövőbeli potenciális döntéseinek eredményeit, a legjobb választás érdekében.



4. ábra: Döntéshozás támogatása (Forrás: saját szerkesztés)

2. ADAT- ÉS MÓDSZERTANI LEÍRÁS

Munkám tehát az 1. fejezetben leírtaknak megfelelően a helymarketingre fókuszál, egy adott terület vonzerejének és egyúttal versenyképességének növelése érdekében. Célja, hogy meghatározza és mérhetővé tegye az attraktivitást a településmarketing kontextusában, ezzel segítve a döntéshozókat az optimális döntések meghozatalában.

Ehhez Dálnok, Esztelnek, Zabola, Csíkszentdomokos és Kovászna települések adatai kerültek felhasználásra. Kiválasztásukat indokolja, hogy egymáshoz földrajzilag közel, a Keleti-Kárpátok és a Kárpátkanyar Hargita felőli részén találhatók, illetve, hogy népességszám szerint három különböző sávban helyezkednek el, azaz a kisközségek (2000 lakos alatt), nagyközségek (2000 lakos fölött), valamint a városok dimenziójában. Ezzel a rajtuk végzett prediktív modellezés eredményességének esetén a kapott modellek hasonló településeken is alkalmazhatók, segítve az ottani döntéshozók munkáját. Ez azért is lehet fontos, mert Erdély ezen részén a településmarketing még nem kiforrott, így nagy segítség lehet itt egy ilyen fejlesztett eszköz.

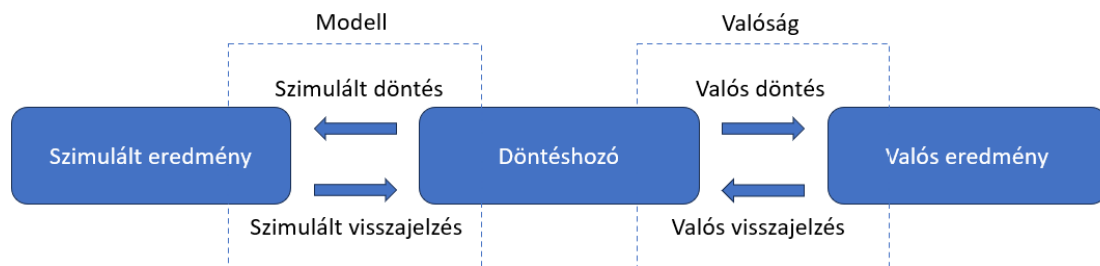
2.1. ÜZLETI CÉL ÉS MEGVALÓSÍTÁSI MÓDSZER, BEMENETEK ÉS KIMENETEK

Az olyan adatokat, amelyek statisztikai hivatalból vagy adatbányászati módszertanok segítségével elérhetők, egy döntéshozó emberi munkaórában mérve relatív kis költséggel tud beszerezni. Ezek az adatok is alkalmasak lehetnek arra, hogy bizonyos következtetéseket tudjon az adott vezető levonni a jövőbeli jó irányok meghatározására vonatkozólag, ahogy azt az 1.6 fejezet is érintette. Azonban az emberi, lakossági elégedettség ezekből az információkból nem tükröződik triviális módon.

Márpedig, ahogy az az 1. fejezetből is világos a régió- és településmarketing célja a térség versenyképességének, előnyeinek és vonzerejének feltárása és növelése. Ez magában foglal egy alapos elemzést, érdekegyeztetést, és stratégiaalkotást, amely a helyi lakosság gazdasági, közösségi, kulturális és ökológiai jólétének javítására irányul. A helyi lakosság elégedettsége és jóléte pedig kiemelt jelentőséggel bír a marketing stratégiák és kommunikációs tevékenységek kialakításában. Így egy döntéshozónak kiemelt fontosságú, hogy erre a statisztikai és bányászható információkból nem triviálisan következő adatra fény derüljön. Ezen könnyebben és nehezebben megtudható tények között természetesen van összefüggés, amelynek modellezése jó megoldás a felmerülő problémára.

Ebben segíthet egy gépi tanuló modell, amely képes összekapcsolni a könnyen és nehezen megszerezhető adatokat, és így egy átfogóbb képet ad a helyzet elemzéséről és később

a lehetséges stratégiák kialakításáról. Amennyiben a statisztikai és bányászati úton megszerzhető adatok mellé, mint összetartozó értékpárhalmazokat, összegyűjtik az elégedettségi adatokat is, akkor illeszthető egy gépi tanuló modell a problémára. Ez lényegében egy vektor-vektor típusú leképezés. Ez a gépi tanuló modell valójában az összegyűjtött adatok esetéből tanul és így később alkalmazható lesz olyan esetekre, amikor csak a könnyen beszerezhető információk állnak rendelkezésre. Persze kiemelten fontos figyelni a tanítás során olyan aspektusokra, mint az általánosítás, a túlillesztés elkerülése. A bemutatott problémakört a 4. ábra részletesebb változata, az 5. ábra szemlélteti.



5. ábra: Döntéshozás modellezés segítségével (Forrás: saját szerkesztés)

Az 5. ábra alapján látható, hogy a cél a valóság modellezése mesterséges intelligenciával. Minél pontosabb ez a modell, annál jobbak lesznek az egyes vezetők valós döntései is, mivel a szimulációkból annál inkább képesek lesznek a valóságra is igaz következtetéseket levonni és a szimulációban jó döntéseik így a valóságban is jól fognak működni. Ez a jó és általános modell pedig úgy érhető el, ha nem csupán tanítjuk a meglévő adatokkal az algoritmust, hanem teszteljük is. Így az adataink egy részét erre használva szimulálhatjuk azt az esetet is, amikor a tanított modell egy még nem látott bemenetet kap. Ha itt is jól teljesít a gépi tanuló algoritmus, akkor megalapozottan gondolható, hogy az ismeretlen adatokon is jól fog működni. Ezzel a kérdéskörrel részletesebben a 2.2.6.1 alfejezet foglalkozik.

2.1.1. Bemenetek

Az egyes elégedettségi becslésekhez Dálnok, Esztelnek, Zabola, Csíkszentdomokos és Kovászna településekről állt rendelkezésre több Román Statisztikai Hivatal által szolgáltatott információ és nyíltan elérhető bemeneti adat is. Így felhasználható volt:

- a Lakosság száma,
- a 2019. évi saját bevétel,
- a 2020. évi saját bevétel,
- a 2019. évi összes bevétel,

- a 2020. évi összes bevétel,
- a Facebook elérések száma,
- a honlap látogatók száma,
- a havonta átlagosan a közösségi médiában megjelenő posztok száma,
- a Facebook kedvelések száma,
- a 2019. évi rendezvények száma,
- a 2020. évi rendezvények száma,
- a 2019. évi sajtómegjelenések száma,
- valamint a 2020. évi sajtómegjelenések száma.

Az ennek megfelelő 2. táblázat alább figyelhető meg:

2. táblázat: A bemeneti adatok összefoglalása (Forrás: saját szerkesztés)

Adattípus	Dálnok	Esztel- nek	Zabola	Csík- szentdo- moks	Kovászna
Lakosság [-]	956	1.182	4.597	6.110	10.114
Saját bevétel 2019 [RON]	853.392	1.204.119	5.664.135	3.656.612	15.629.954
Saját bevétel 2020 [RON]	1.059.376	1.287.837	3.329.468	3.404.590	16.466.476
Összes bevétel 2019 [RON]	2.793.300	2.891.012	5.664.135	8.964.416	19.117.290
Összes bevétel 2020 [RON]	2.374.104	3.373.354	7.425.185	11.430.885	24.226.781
Facebook elérés [-]	710	1.100	1.841	4.400	1.453
Honlap látogató [-]	35	85	80	50	120
Posztok száma (havonta) [-]	9	5	2	18	6
Lájkok száma [-]	45	135	35	140	115
Rendezvények száma 2019 [-]	13	32	25	35	33
Rendezvények száma 2020 [-]	7	10	10	15	10
Sajtómegjelenések száma 2019 [-]	36	32	58	65	250
Sajtómegjelenések száma 2020 [-]	28	39	27	37	150

2.1.2. Kimenetek

A bemenetek alapján prediktálandó kimenetek kérdőívek segítségével kerültek beszerzésre. Jelen munka fókusz a megfelelő matematikai modellállítási módszertan kidolgozása, nem pedig a tökéletes statisztikai felmérés. Ennek tükrében az egyes településeken kitöltött 60-60 kérdőív bár nem méri tökéletesen az egyes települések lakosságának elégedettségét, de valamilyen plusz információval szolgál róla a bemeneti adatokhoz képest arról. Ezen plusz információnak kiderítése pedig általánosságban érték egy települést vezető döntéshozó számára. Így a kimenetek becslésére vonatkozó helyes módszertan további „ismert bemenet, de már ismeretlen kimenet” típusú szituációk esetén is hasznos lesz a későbbiekben.

A kérdőív kérdéseire 1 és 10 közötti zárt intervallumban kellett pontot adni. A kérdések az alábbiak voltak:

- Ön szerint a település fejlődés szempontjából sikeres település?
- Érdeemes lenne a településnek nagyobb hangsúlyt fektetni a tökevonzás fejlesztésére?
- Ön szerint a lakosság mennyire ismeri a település értékeit?
- Ön szerint a lakosság szeret a településen élni?
- Milyen mértékben támogatja a hivatal a turizmust a településen?
- Mennyire tartja sikeresnek a stratégiát?

Az 1-es értékelés jelentette az „egyáltalán nem”, a 10-es a „teljes mértékben igen” választ. Az eredményeket a 3. táblázat foglalja össze:

3. táblázat: A kimeneti adatok összefoglalása (Forrás: saját szerkesztés)

Adattípus	Dálnok	Esztelnek	Zabola	Csíkszent-domokos	Kovászna
Sikeresség	10	9	7	9	7
Tökevonzás fejlesztés	10	10	10	9	8
Értékek	10	6	7	7	7
Élhetőség	4	5	4	4	4
Turizmus	10	10	8	7	9
Stratégia	10	9	10	8	8

2.2. A MÓDSZERTAN ALAPJÁT KÉPEZŐ MATEMATIKAI-INFORMATIKAI, GÉPI TANULÁSI MODELLEK ISMERTETÉSE

Munkámban a legmegfelelőbb matematikai modellezés érdekében több tanuló modell típust és skálázási módszert is kipróbáltam. A következő alfejezetekben a matematikai modell készítéséhez szükséges, összehasonlítás alapját képező lineáris regresszió, polinomiális regresszió, döntési fa és véletlen erdő módszertanokat ismertetem. Ezeket a módszereket kipróbáltam skálázás mellett és anélkül is, így a használható skálázási módszertanokat is bemutatom, valamint az implementációt is ismertetem.

2.2.1. Lineáris regresszió

A lineáris regresszió (Bishop, 1995) egy statisztikai módszer, amelynek célja egy vagy több független változó (magyarázó változó) és egy függő változó közötti kapcsolat modellezése és elemzése. A módszer azon alapul, hogy egy lineáris hipersík írja le az adatok közötti összefüggést. A hipersík egyenlete a következőképpen áll elő:

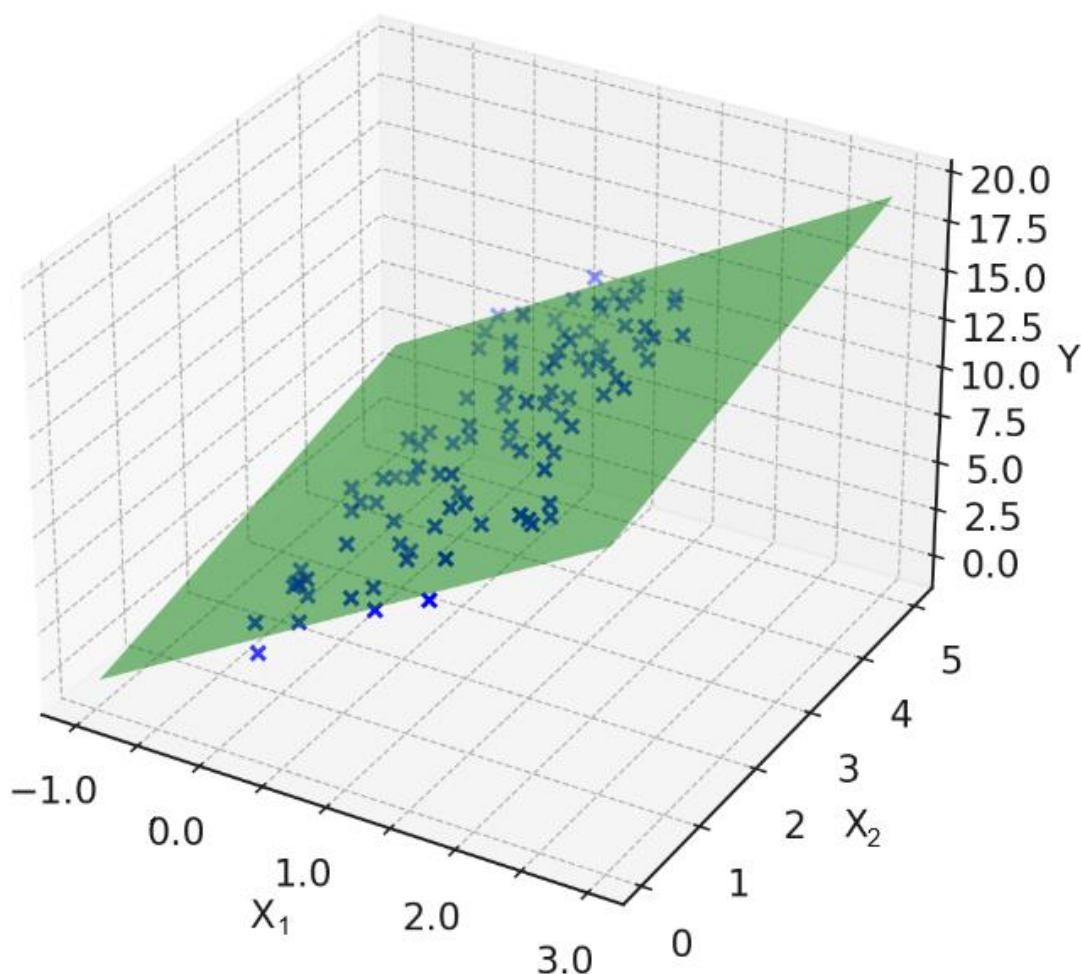
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon, \quad (1)$$

ahol:

- Y a függő változó,
- $X_1, X_2, \dots, X_n$ a magyarázó változók,
- $\beta_0, \beta_1, \beta_2, \dots, \beta_n$ a regressziós együtthatók,
- ϵ pedig a modell hibája.

Itt β_0 a metszéspont, amely azt mutatja meg, hogy hol metszi a hipersík az Y tengelyt, míg a $\beta_1, \beta_2, \dots, \beta_n$ együtthatók azt mutatják, hogy az $X_1, X_2, \dots, X_n$ magyarázó változók egy egységnyi változása mennyivel változtatja meg az Y változót, feltéve, hogy a többi magyarázó változó állandó marad. Az ϵ változó reprezentálja a modell és az adatok közötti különbséget.

A lineáris regresszió célja a regressziós együtthatók (β) becslése, amelyek minimalizálják a modell hibáinak négyzetösszegét (a valós értékek és a modell által prediktált értékek közötti különbségek négyzeteinek összege). A legkisebb négyzetek módszere (OLS - Ordinary Least Squares) a leggyakrabban használt módszer az együtthatók becslésére. Ennek vizualizációja szemléltetés céljából a könnyen interpretálható háromdimenziós esetben a 6. ábrán figyelhető meg:



6. ábra: Lineáris regresszió interpretálható vizualizációja (Forrás: saját szerkesztés)

2.2.2. Polinomiális regresszió

Az előzőekhez hasonlóan a polinomiális regresszió (Kleijnen, 2017) is egy statisztikai módszer, amely egy függő és egy vagy több független változó közötti kapcsolat modellezésére szolgál. Az eljárás lényege, hogy a független változók polinomiális függvényeit használja a függő változó megbecsléséhez. A lineáris regresszióval szemben a modell nemcsak az eredeti függő változókat veszi figyelembe, hanem azok hatványait és egymással való kölcsönhatásait is, így a kapcsolatot komplexebben képes leírni. A modell alapvető formája a következő:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \beta_{n+1} X_1^2 + \dots + \beta_{2n} X_n^2 + \beta_{2n+1} X_1 X_2 + \dots + \epsilon, \quad (2)$$

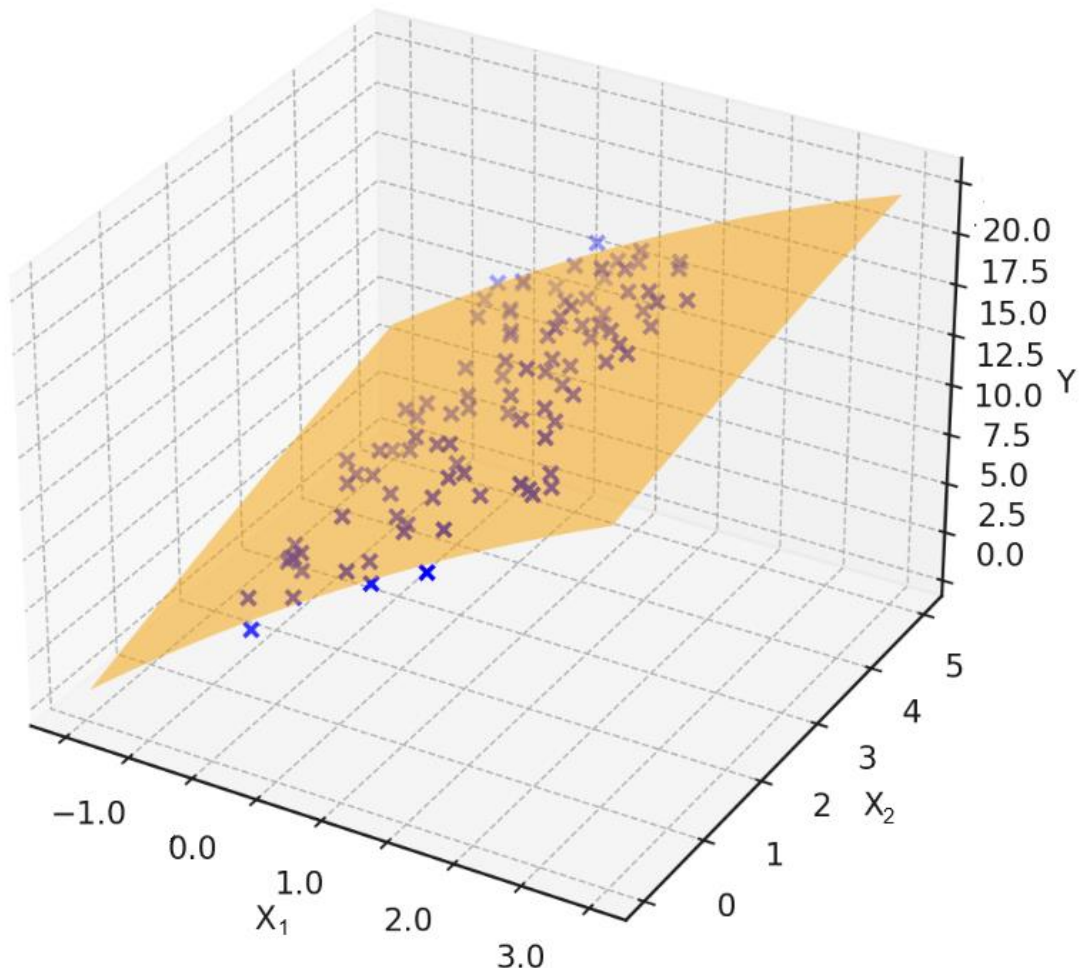
ahol:

- Y a függő változó,
- $X_1, X_2, \dots, X_n$ a független változók,

- $\beta_0, \beta_1, \beta_2, \dots$ a modell paraméterei,
- ϵ pedig a hibatag.

A modell paramétereit itt is a legkisebb négyzetek módszerével szokták becsülni, azaz olyan paramétereket keresünk, amelyek minimalizálják a hibák (a megfigyelt és a modell által prediktált értékek közötti különbségek) négyzetösszegét.

A többváltozós polinomiális regresszió előnyei közé tartozik az, hogy képes megragadni a változók közötti nemlineáris kapcsolatokat és interakciókat. Azonban óvatossá kell lennie annak, aki ezt a módszert alkalmazza, mivel könnyen túlzottan illeszkedő (overfitting) modelleket eredményezhet, amelyek a tanulódathalmazon jól teljesítenek, de új adatokra rosszul általánosítanak. Továbbá, ahogy a polinomok foka növekszik, a modellezés bonyolultabbá válik, és az egyes változók közötti kapcsolatok értelmezése nehezebb lesz.



7. ábra: Polinomiális regresszió interpretálható vizualizációja (Forrás: saját szerkesztés)

A többváltozós polinomiális regresszió modellek validációjához és értékeléséhez gyakran használnak különféle mértékeket, mint például az R^2 négyzetes hibát. Az említett 2 módszerhez hasonló módszerek esetében nagyon fontos éppen az overfitting elkerülésének érdekében a későbbiekben bemutatott skálázás. A lineáris regresszió esetéhez hasonlóan mutatja a 7. ábra a polinomiális regresszió interpretálható, háromdimenziós vizualizációját. Itt ahogy az megfigyelhető, a felület nem tökéletes sík, így adott esetben jobban illeszkedhet az ismert pontokra.

2.2.3. Döntési fa

A döntési fák (Breiman, Friedman, Stone, & Olshen, 2017) használata egy népszerű gépi tanulási technika, amelyet a prediktív modellezés széles spektrumában használnak. A módszer alkalmas mind kategorikus, mind numerikus adatok feldolgozására és könnyen értelmezhető modellek létrehozására. A regressziós döntési fák kifejezetten a folytonos változók vagy numerikus adatok előrejelzésére használhatók.

A regressziós döntési fa modellezésének alapelve, hogy először, a modell egy döntési fát épít fel, amely a tanító adatokon alapul. A fa minden csomópontja egy feltételt tartalmaz, ami az egyik bemeneti változóval kapcsolatos. Az adatokat aszerint osztják szét, hogy azok megfelelnek-e vagy nem felelnek meg a feltételnek. A levelek (azaz a fa végpontjai) az előrejelzett kimeneti értékeket tartalmazzák, amelyek azon adatpontok átlagos célfüggvény-értékén alapulnak, amelyek az adott levélre kerültek.

A predikció készítéséhez egy új adatpontot a fa gyökerétől kezdve vezetünk végig, mindig az adott csomópontban szereplő feltétel alapján eldöntve, hogy melyik irányba haladjunk tovább, mígnem elérünk egy levél csomópontot, amely tartalmazza az előrejelzési értéket.

Matematikai szempontból a regressziós döntési fa az alábbi módon közelíti meg a kimeneti változó Y értékét:

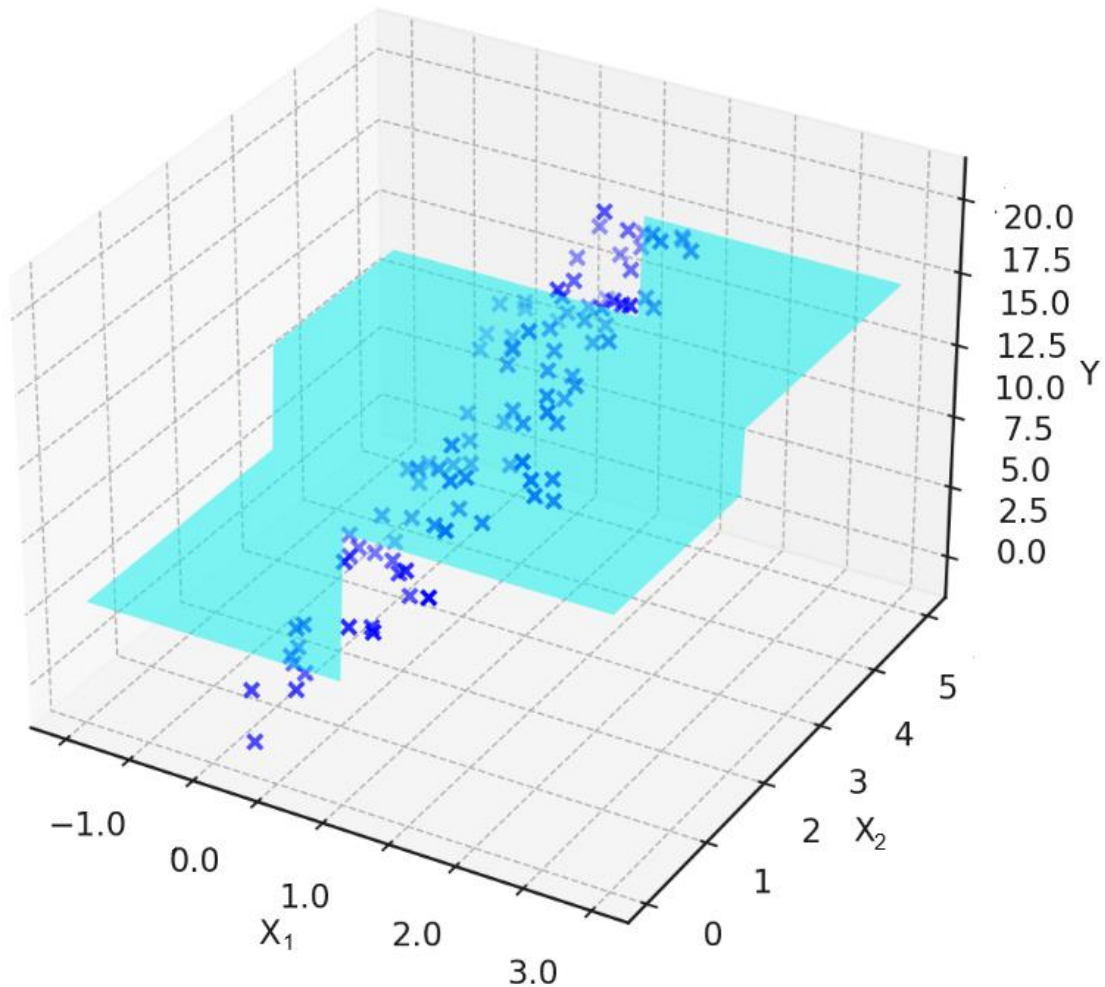
$$Y_{pred} = \frac{1}{|S|} \sum_i Y_i \quad (3)$$

Itt:

- Y_{pred} a prognosztizált kimeneti érték,
- S az adatpontok halmaza, amelyek azonos levélbe kerülnek a fában,
- Y_i pedig az i -edik adatpont kimeneti értéke.

A fa építése során az algoritmus célja, hogy minél jobban csökkentse a különböző levélcsoportokon belüli varianciát vagy a négyzetes hibát, vagyis az egyes adatpontok és a

hozzájuk tartozó levelek átlagértéke közötti négyzetes eltérések összegét. A cél egy olyan fa struktúra létrehozása, amely minimalizálja a részletező kérdések számát (tehát a fa mélységét), ugyanakkor jól megjósolja a független változó értékét, amikor új adatok érkeznek.



8. ábra: Döntési fa interpretálható vizualizációja (Forrás: saját szerkesztés)

A döntési fa ilyen módon kapott háromdimenzióban interpretálható vizualizációja, a 8. ábrán látható. Megfigyelhető, hogy a változók terének különböző térrészeiben más-más konstans szintet jósol a modell.

Az ide kötődő előnyök közé tartozik a modell egyszerűsége és interpretálhatósága, valamint a képesség, hogy kezelje a nemlineáris kapcsolatokat és a hiányzó adatokat. Azonban a kihívások közé tartozik a túl komplex modellek kialakításának veszélye és az, hogy a döntési fák hajlamosak lehetnek a túltanulásra, különösen, ha a fa túl mélyre nő. Az ensemble módszerek (mint például a következőkben bemutatott véletlen erdő) gyakran

használhatók a döntési fák hátrányainak csökkentésére és a teljesítményük javítása érdekében.

2.2.4. Véletlen erdő

A véletlen erdő (Segal, 2004) vagy más néven "Random Forest", egy erőteljes gépi tanulási algoritmus, amely az ensemble learning technikák egyike. Ez az algoritmus különösen jól alkalmazható mind a klasszifikációs, mind a regressziós problémák megoldására. A véletlen erdő alapvetően számos döntési fát (az „erdő”) kombinál, hogy egy stabilabb és pontosabb előrejelzést hozzon létre.

Alapvetően a működés több döntési fa létrehozásával indul. A véletlen erdő algoritmus sok döntési fát generál, amelyeket aztán összekapcsol. Minden fa építése során a tanító adatok egy véletlen alhalmazát választják ki (bootstrap módszerrel, ami azt jelenti, hogy az elemek ismétlődhetnek), hogy biztosítsák a modellek diverzitását.

Ezt követően a regressziós esetben a modell az egyes fák előrejelzéseinek átlagát veszi, hogy végleges előrejelzést kapjon. Klasszifikációs feladatok esetén a leggyakoribb osztály (a "szavazatok" többsége) lenne az előrejelzés.

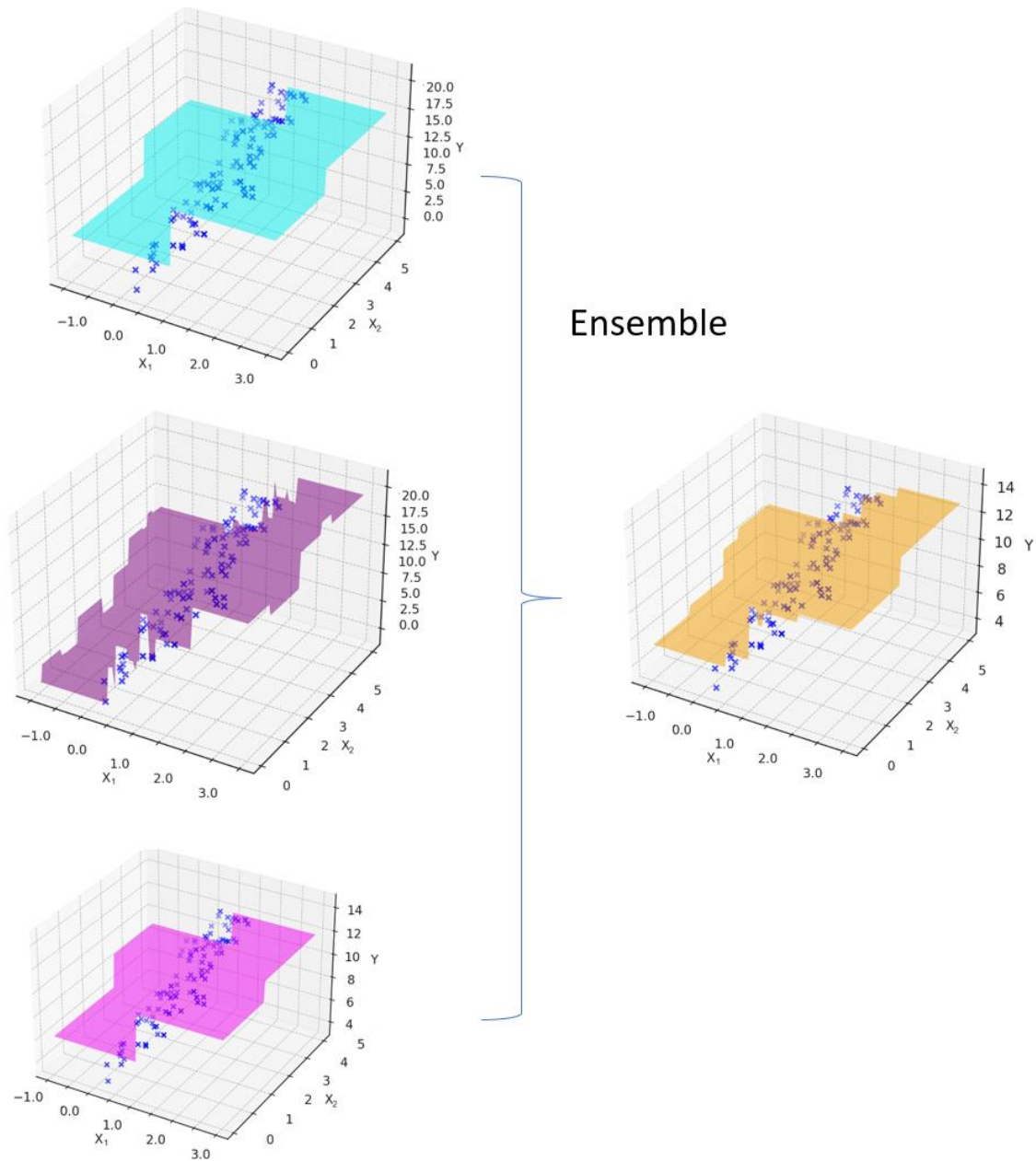
Matematikailag, véletlen erdő regresszió esetén, K döntési fa használata mellett, az előrejelzés a következőképpen számítható:

$$Y_{pred} = \frac{1}{K} \sum_k Y_k \quad (4)$$

ahol:

- Y_{pred} a prediktált kimenet,
- Y_k a k -adik fa előrejelzése,
- K pedig a fák száma az "erdőben".

A véletlen erdő algoritmus kiváló választás lehet sokféle alkalmazáshoz, különösen, amikor egyensúlyt kell találni az előrejelzési pontosság és a modell interpretálhatósága között. Ennek vizualizációja látható a 9. ábrán, amely 3 fa kimeneti értékeit együttesen veszi figyelembe egy jobb döntés meghozatalának érdekében.



9. ábra: Véletlen erdő döntési fákból álló interpretálható vizualizációja (Forrás: saját szerkesztés)

2.2.5. Skálázás

Ezen módszerek kapcsán nagyon fontos beszélni a már említett skálázásról (Ahsan, Mahmud, Saha, Gupta & Siddique, 2021), ez ugyanis kritikus szerepet játszik a gépi tanulási modellek teljesítményében, különösen azon algoritmusok esetében, amelyek a változók közötti távolságot használják az előrejelzések generálásához vagy a változók súlyozásához. A változók nagy mértékű skálakülönbségei torzíthatják a modellek tanulását, mivel a nagyobb

értékek dominánssá válhatnak a kisebbekkel szemben. Esetünkben ilyen a lineáris regresszió és a polinomiális regresszió is. A két legelterjedtebb skálázási módszer a normalizálás és a standardizálás.

2.2.5.1. Normalizálás

A normalizálás során az adatokat úgy skálázzuk és/vagy eltoljuk, hogy azok értékei egy $[0,1]$ intervallumba kerüljenek. Azaz, az adatpontokat az alábbi módon transzformáljuk:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5)$$

ahol:

- X_{norm} a normalizált érték,
- X az eredeti érték,
- X_{min} az adott adattípus minimális értéke,
- X_{max} pedig az adott adattípus maximális értéke.

2.2.5.2. Standardizálás

A standardizálás során az adatokat úgy alakítjuk, hogy azoknak átlaga 0 és szórása 1 legyen. Ez az eljárás gyakran ismert "z-transzformáció" néven is:

$$X_{std} = \frac{X - \mu}{\sigma} \quad (6)$$

ahol:

- X_{std} a standardizált érték
- X az eredeti érték,
- μ az adott adattípus átlaga,
- σ pedig az adott adattípus szórása.

A változók skálájának normalizálása vagy standardizálása segíthet javítani a gépi tanulási modellek teljesítményét, és stabilabb, összehasonlíthatóbb eredményeket biztosíthat, továbbá a gradiens alapú és más optimalizálási algoritmusok gyorsabban konvergálnak, ha az összes változó azonos skálán van, mivel a keresési tér egyenletesen van méretezve minden dimenzióban.

2.2.6. Különböző gépi tanulási módszertanok algoritmikus összevetése

A kutatásnak fókuszát képezte a fent bemutatott módszertanok automatizált összehasonlítása is. A Python 3 programozási nyelv (Van Rossum, & Drake, 2009), valamint a

Sklearn (Pedregosa, Varoquaux, Gramfort, Michel, Thirion, Grisel, & Duchesnay, 2011), Pandas (McKinney, 2010), Numpy (Harris, Millman, Van Der Walt, Gommers, Virtanen, Cournapeau, & Oliphant, 2020) és Matplotlib (Hunter, 2007) könyvtárak segítségével történt az említett gépi tanuló modellek feltanítása a korábban említett input és output adatok segítségével.

2.2.6.1. Keresztvalidáció

Mivel a korábbiakban említetteknek megfelelően kifejezetten kevés adatból lehetett dolgozni, így a keresztvalidáció eszközéhez kellett folyamodni. A keresztvalidáció (cross-validation) egy technika a gépi tanulás modelljeinek validációjához és értékeléséhez, amely segít megelőzni az overfittinget, és biztosítja, hogy a modell jól generalizáljon ismeretlen adatokra. Az alapötlet, hogy a rendelkezésre álló adatkészletet szándékosan felosztjuk különböző részekre, hogy bizonyos részeket a modell tanítására, míg másokat annak tesztelésére vagy validálására használjunk.

A keresztvalidáció, a gépi tanulási modellek értékelésének kulcsfontosságú lépése, számos változatban jelenik meg a tudományos és ipari alkalmazásokban. A K-Fold keresztvalidáció (Anguita, Ghelardoni, Ghio, Oneto, & Ridella, 2012) a legelterjedtebb típusa ennek a módszernek, amelyben az adatkészletet "K" részre osztjuk, majd a modellt "K" alkalommal futtatjuk, minden egyes iterációban egy különböző részt, vagy "foldot" használva tesztadatként, míg a maradék részt tanulási célra használjuk. Az így kapott eredményeket átlagoljuk, így egyetlen értékelési metrikát kapunk.

A Leave-One-Out (Wong, 2015) keresztvalidáció egy speciális eset, ahol a "K" értéke egyenlő az összes rendelkezésre álló adatpont számával, ami azt jelenti, hogy minden egyes iterációban egy adatpontot hagyunk ki tesztelésre, a maradékot pedig tanításra használjuk. Ez a módszer, bár igen pontos, számítási szempontból költséges lehet. Egy másik változat, a Leave-P-Out (Arlot, & Celisse, 2010) keresztvalidáció, szintén minden lehetséges "P" elemű alhalmazt kipróbál, tehát szintén számításigényes.

A keresztvalidáció minden formája esszenciális abban, hogy segít a fejlesztőknek elkerülni a túltanulást, valamint lehetővé teszi, hogy objektíven mérjék és összehasonlítsák különböző modellek teljesítményét, biztosítva, hogy a kiválasztott modell a lehető legjobban fog teljesíteni új, ismeretlen adatokon is. Jelen esetben, mivel összesen 5 település adataival lehetett dolgozni így a legeffektívebb, Leave-One-Out technikát volt célszerű alkalmazni. Így lehetett megbecsülni az input adatokból az output adatokat. A hiba mértékének számítása a négyzetes hiba elvén történt.

2.2.6.2. Futtatások első lépései

Mivel az említett 4 modell és 5 település képezte a munka fókuszát a skálázatlan, a normalizált és a standardizált esetben is, így összesen 60 db futtatást kellett végezni. Ennek első részlete alább látható:

```
from google.colab import drive

# Mount Google Drive
drive.mount('/content/drive')

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, r2_score
import numpy as np
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.preprocessing import PolynomialFeatures
import matplotlib.pyplot as plt

input_data = pd.read_excel("drive/MyDrive/adatok.xlsx", sheet_name="Input")
output_data = pd.read_excel("drive/MyDrive/adatok.xlsx", sheet_name="Output")

X = input_data.drop('Település/Input adatok', axis=1)
y = output_data.drop('Település/Output adatok', axis=1)
```

A kódrészlet a korábban említetteknek megfelelően Python nyelven íródott, és az adatait a Google Drive-ról olvassa be a Google Colab felhasználása mellett. A Google Colab egy online, a böngészőből elérhető kódfejlesztési környezet, ami lehetővé teszi a felhasználóknak, hogy Python kódokat futtassanak az általuk használt böngésző segítségével. Elsőként a Google Drive-ot csatolja a Colab környezethez, ami lehetővé teszi számára a Google Drive-on tárolt fájlok elérését. Az ezt követő megfelelő importálások után egy "adatok.xlsx" nevű Excel fájlt olvas be, két különböző lappal, azaz "Input" és "Output" lapokkal, melyekből az input és output adatokat szerzi be.

A Pandas könyvtárat használja a kód az adatok beolvasásához és előfeldolgozásához, ami egy népszerű könyvtár adatfeldolgozásra és adatelemzésre. A beolvasott adatokból eldob néhány oszlopot, egész pontosan a nem numerikus jellegűeket, ami itt a 'Település/Input adatok' és a 'Település/Output adatok' oszlopokat jelenti.

A Sklearn könyvtárból számos modult importál, amik a későbbiekben kerülnek felhasználásra a modellezéshez és az eredmények kiértékeléséhez. Ezek között találunk lineáris regressziót, döntési fát és véletlen erdőt, mint regressziós modelleket, valamint az adatok

szétosztásához és a modell teljesítményének kiértékeléséhez szükséges funkciókat és metrikákat is. Továbbá néhány előfeldolgozási eszközt és polinomiális jellemzők előállítására szolgáló eszközt is importál a kód.

Az Numpy könyvtár egy széles körben alkalmazott könyvtár numerikus számításokra, amely itt az adatok vektoros feldolgozásához használható, míg a Matplotlib az adatok és az eredmények vizualizációjára alkalmas. Az ezt követő kódsorok az alábbiak szerint alakulnak:

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
```

Látható, ahogy a megfelelő *random_state* és *test_size* beállítással megtörténik az adatok validációs és tanító adathalmazokra osztása a keresztvalidáció követelményeinek megfelelően. A kódrészlet a *train_test_split* függvényt alkalmazza, amely a scikit-learn könyvtárban található és segít abban, hogy az adatokat tanuló- és teszhalmazokra oszthassuk. Ebben a kontextusban *X* a jellemzőket tartalmazó változó, amely az előzőleg tárgyalt *input_data DataFrame*-ből származik, a nem kívánt oszlopok eltávolításával, konkrétan a 'Település/Imput adatok' oszlop elhagyásával. Ugyanakkor *y* a célváltozót tárolja, és ez származik az *output_data DataFrame*-ből, a 'Település/Output adatok' oszlop eltávolításával. A *test_size* paraméter 0.2 értéke azt specifikálja, hogy az össz adat 20%-át szeretnénk használni teszhalmazként, míg a maradék 80%-át tanító adathalmazként. A *random_state* paraméter értéke 4, ami biztosítja, hogy minden futtatásnál ugyanazokat az elemeket válogatja ki a teszhalmazba, így a reprodukálhatóság garantált. Végül, az *X_train*, *X_test*, *y_train*, *y_test* változók tárolják a felosztott adathalmazokat. Az *X_train* és *y_train* a modell tanításához használt jellemzők és célváltozók, míg az *X_test* és *y_test* a tesztelési fázisban felhasznált adatokat tárolják. Az egész sor így segít előkészíteni a modellünk tanítási és tesztelési fázisait, biztosítva, hogy a két adathalmaz különböző, és lehetőségünk van validálni a modell teljesítményét az ismeretlen, teszt adatokon. Ez a felosztás Dálnok település becsléséhez használatos. Más eseteken ezzel megegyezően kell elvégezni az adatszétválasztást, más *random_state* mellett.

2.2.6.3. Lineáris és polinomiális regressziós modellek futtatása

Innentől különböző modellek és skálázási megfontolások esetén különböző nyelvi elemek használata szükséges. Amennyiben normalizálás vagy standardizálás történik, akkor szükséges a megfelelő skálázó objektum létrehozása:

```
X_scaler = StandardScaler()
```

Vagy normalizálás esetén:

```
X_scaler = MinMaxScaler()
```

Mindkét skálázót először az *fit* metódussal illesztik az adathoz, majd a *transform* metódussal skálázzák az adatokat. A *fit_transform* metódus egy lépésben végzi mindkét műveletet. Így a tanító adatok esetén a *fit_transform* a teszt adatok esetén pedig a *transform* művelet használatos. Ezáltal a tanítóadatok alapján számított értékeket alkalmazzuk a tesztadatokra, garantálva a konzisztenciát a skálázás terén:

```
scaled_X_train = X_scaler.fit_transform(X_train)
scaled_X_test = X_scaler.transform(X_test)
```

Ugyanezek a műveletek természetesen a kimeneti adatok esetén is szükségesek:

```
y_scaler = StandardScaler()
scaled_y_train = y_scaler.fit_transform(y_train)
scaled_y_test = y_scaler.transform(y_test)
```

Ezt követően polinomiális regresszió alkalmazása esetén szükséges kiszámítani a polinomiális jellemzőket. Erre a más korábban érintett *sklearn.preprocessing PolynomialFeatures* osztályát lehet használni:

```
# Choose the degree of the polynomial features
poly_features = PolynomialFeatures(degree=2)

# Transform your training and test data to include polynomial features
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)
```

Értelemszerűen skálázás alkalmazása esetén a skálázott adatokon végzendő el az említett művelet.

Így már létrehozható és futtatható a regressziós modell. Ennek is a megfelelően előfeldolgozott adatokat kell odaadni. Az itt bemutatott példában skálázott adatokat kap:

```
model = LinearRegression()
model.fit(scaled_X_train, scaled_y_train)
scaled_y_pred = model.predict(scaled_X_test)
```

Ezt követően skálázás alkalmazása esetén még szükséges elvégezni az inverz transzformációt is a végeredmény eléréséhez:

```
y_result = y_scaler.inverse_transform(scaled_y_pred)
```

Ezzel pedig sikerült elérni a megfelelő becslést. Összefoglalva ebben az alfejezetben lineáris és polinomiális regressziós modellek futtatására került sor. Az adatok vagy skálázatlanul

maradtak, vagy normalizálásra, vagy standardizálásra kerültek, a *MinMaxScaler*, illetve *StandardScaler* objektumok használatával, majd az rájuk polinomiális esetben polinomiális jellemzőket számoltunk ki. A regressziós modell a megfelelően előfeldolgozott adatokon került futtatásra, majd a kapott eredményeket szükség esetén inverz transzformációval kellett visszaalakítani az eredeti skálára.

2.2.6.4. *Döntési fa és véletlen erdő modellek futtatása*

A skálázás szükség esetén a véletlen erdők és a döntési fák kapcsán is a korábbiakkal megegyező módon történik, azonban itt nem alkalmazandók polinomiális jellemzők. Az egyetlen különbség a modellek létrehozásában van. Döntési fák esetén:

```
model = DecisionTreeRegressor(min_samples_split=3, min_samples_leaf=2)
```

Véletlen erdők esetén pedig:

```
model = RandomForestRegressor(min_samples_split=3, min_samples_leaf=2)
```

A feljebb látható *min_samples_split* és a *min_samples_leaf* két fontos paraméter a döntési fa és az erdő alapú modellekben. Ezek a paraméterek segítenek a túlillesztés (overfitting) elkerülésében. A *min_samples_split* meghatározza, hogy hány minta szükséges legalább egy csomópont elágazásához egy döntési fában. A nagyobb értékek növelhetik a modell általánosító képességét, mivel a fa mélyebb elágazásainak létrehozása több mintát igényel. Azonban, ha ez az érték túl magas, a modell lehet, hogy nem képes eleget tanulni az adatokból, mely alulteljesítést (underfitting) eredményezhet. A *min_samples_leaf* paraméter pedig azt határozza meg, hogy hány minta szükséges legalább egy levél csomópontban. A nagyobb értékek itt is növelhetik a modell általánosító képességét, mivel a levélben több mintát igényelnek, ami megakadályozza a túl mély fák kialakítását. Azonban itt is igaz, hogy ha ez az érték túl magas, a modell lehet, hogy nem lesz képes eleget tanulni, ami alulteljesítést eredményez.

Mivel jelen esetben összesen 4 település bemeneti adataival lehet dolgozni 1 település adatainak megbecslése érdekében, így érdemes ezeket a paramétereket is kisebb értékűre választani. Jelen esetben mindig olyan fa tud létrejönni, amelynek egyetlen elágazása, így 1-1 levele van 2-2 adattal. Az erdők is ilyen fákból épülnek fel.

Ezzel a módszertanok és megvalósításuk ismertetésre került. A bemutatottak alapján képesek a modellek a statisztikai és bányászott adatokból ez egyes településmarketing szempontból kulcsfontosságú elégedettségi mutatókra következtetni, így segítve a döntéshozók munkáját. A következő fejezetben az itt ismertetett módszertan eredményeit mutatom be. A fejezet algoritmusainak táblázatos összefoglalója pedig alább látható.

4. táblázat: Algoritmusok összefoglalása (Forrás: saját szerkesztés)

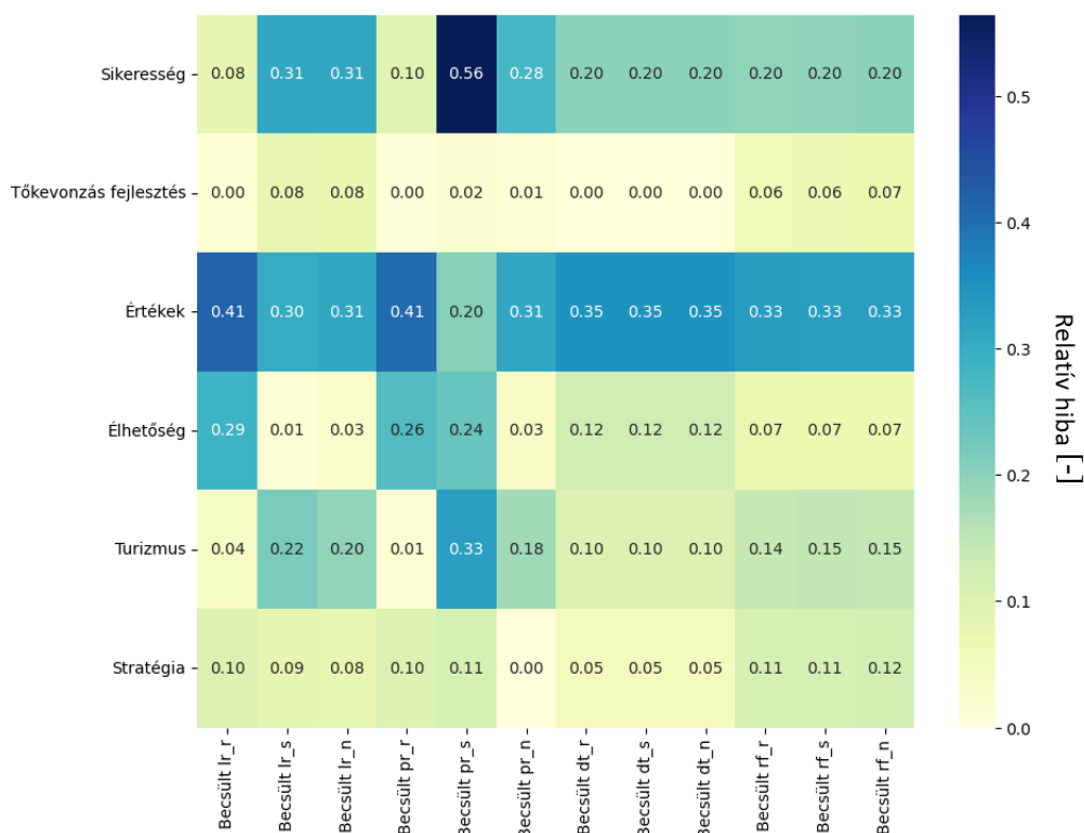
Algoritmus	Alapelve	Fő jellemzők	Alkalmazás	Kihívások
Lineáris regresszió	Függő változó és egy vagy több független változó lineáris kapcsolatának modellezése	-Egyszerű, lineáris kapcsolatokra alkalmas -Gyors -Könnyen interpretálható eredmények	Alapvető statisztikai elemzések	- Nem alkalmas komplex, nemlineáris kapcsolatok modellezésére
Polinomiális regresszió	Függő változó és független változók polinomiális kapcsolatának modellezése	-Nemlineáris kapcsolatok elemzésére alkalmas -Magasabb rendű interakciók és hatványok figyelembevétele -Komplexebb modellek	Összetett kapcsolatok elemzése	-Overfitting veszélye -Számításigényes -Nehézkesebb interpretáció
Döntési fa	Feltételes elágazásokon alapuló modellezés	-Egyszerű, fa alapú struktúra -Könnyen interpretálható -Képes kezelni a hiányzó adatokat és nemlineáris kapcsolatokat	Klasszifikációs és regressziós feladatok	-Overfitting hajlam -Komplexebb fák nehézkes kezelése -Nem mindig a legpontosabb módszer
Véletlen erdő	Sok döntési fa kombinálása	-Robusztus, stabil előrejelzések -Képes kezelni a túlтанulást (overfitting)	Klasszifikációs és regressziós feladatok	- Magas számítási igény -Kevesebb interpretálhatóság

3. LINEÁRIS- ÉS POLINOMIÁLIS REGRESSZIÓS, VALAMINT DÖNTÉSI FA ÉS VÉLETLEN ERDŐ EREDMÉNYEK

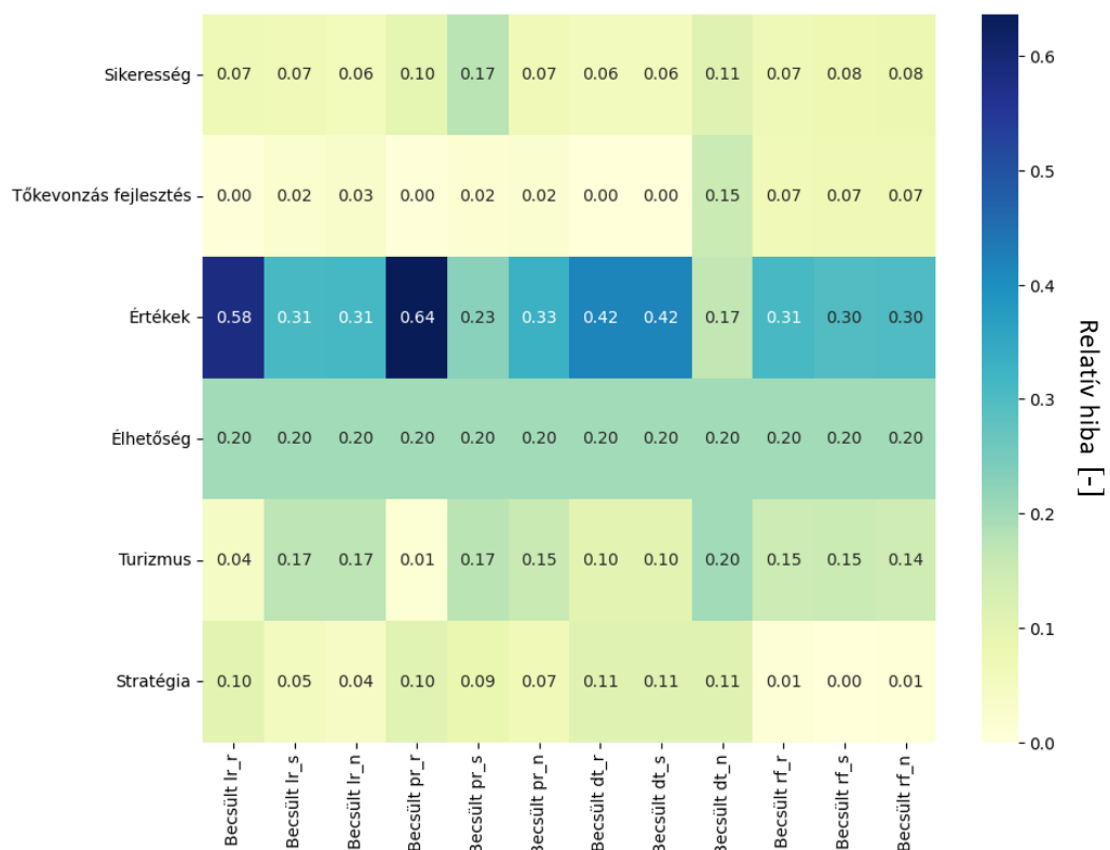
Ebben a fejezetben eredményeimet ismertetem. Megvizsgálom, hogy mely kimeneti adatokat voltak képesek az ismertetett algoritmusok pontosabban és mely adatokat kevésbé pontosan megbecsülni. Az egyes települések esetén ezekre külön-külön kitérek és az egyes felhasználási esetekhez modellt javaslok.

3.1. EREDMÉNYEK ISMERTETÉSE

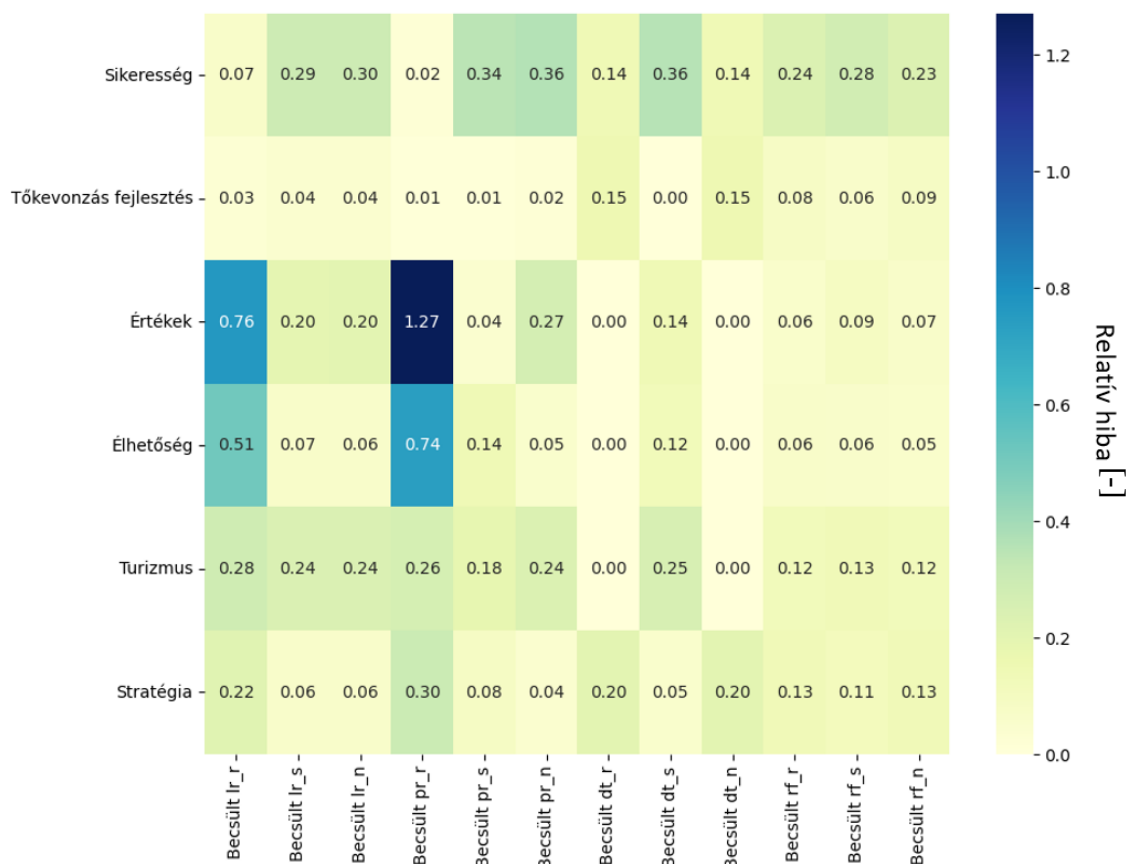
A következő ábrákon rendre Dálnok, Esztelnek, Zabola, Csíkszentdomokos és Kovászna eredményei láthatók. Minden esetben vizsgálom, hogy melyik paraméter volt a legpontosabban becsülhető, illetve, hogy melyik modell illet a legjobban az adott településhez. Az ábra feliratozásnál használt nevezéktanomban az „lr” lineáris regressziót, a „pr” polinomiális regressziót, a „dt” döntési fát (decision tree), az „rf” pedig véletlen erdőt (random forrest) jelöli. Hasonló módon az „r” a nyers adatot (raw) az „s” a standard skálázott adatot, az „n” pedig a normalizált adatot jelöli. Ennek megfelelően a következő eredményeket kaptam.



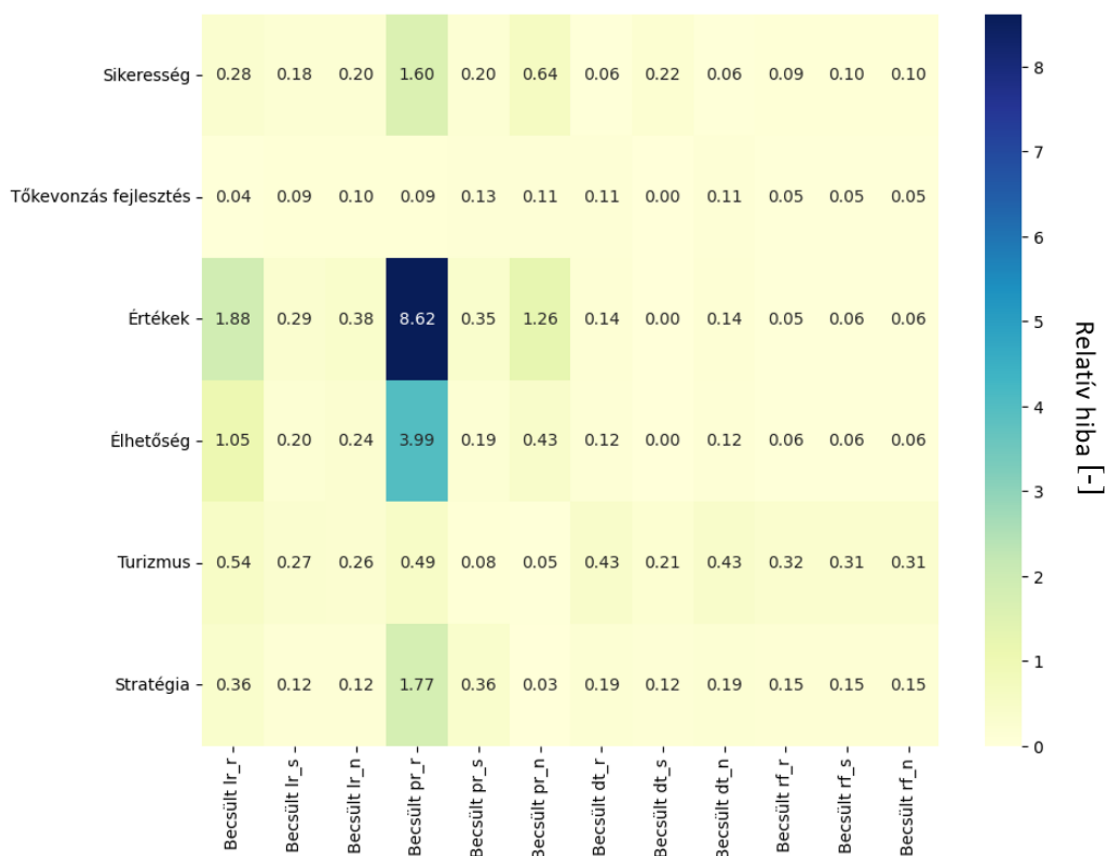
10. ábra: Dálnok relatív hibái (Forrás: saját szerkesztés)



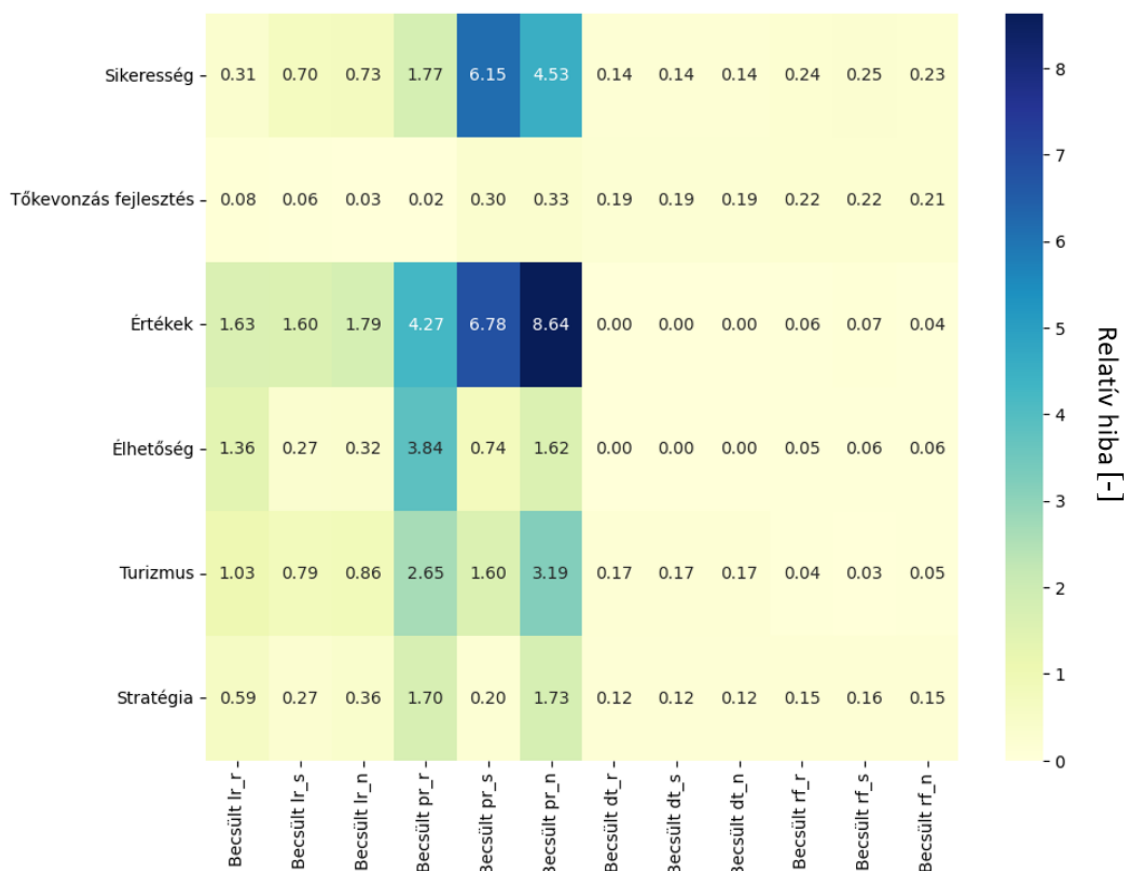
11. ábra: Esztelnek relatív hibái (Forrás: saját szerkesztés)



12. ábra: Zabola relatív hibái (Forrás: saját szerkesztés)



13. ábra: Csíkszentdomokos relatív hibái (Forrás: saját szerkesztés)



14. ábra: Kovászna relatív hibái (Forrás: saját szerkesztés)

3.2. EREDMÉNYEK ÉRTÉKELÉSE

Az eredmények értékeléséhez relatív hibát használtam. Ez egy statisztikai mérőszám, amelyet gyakran használnak a valós és becsült értékek közötti eltérés mérésére. Ez a metrika különösen hasznos, amikor a mért adatok nagysága eltérő lehet, és szükség van egy standardizált hibaérték kiszámítására, amely független a mért adatok nagyságától. A következőképpen számolandó:

$$r = \left| \frac{\hat{z} - z}{z} \right|, \quad (7)$$

ahol:

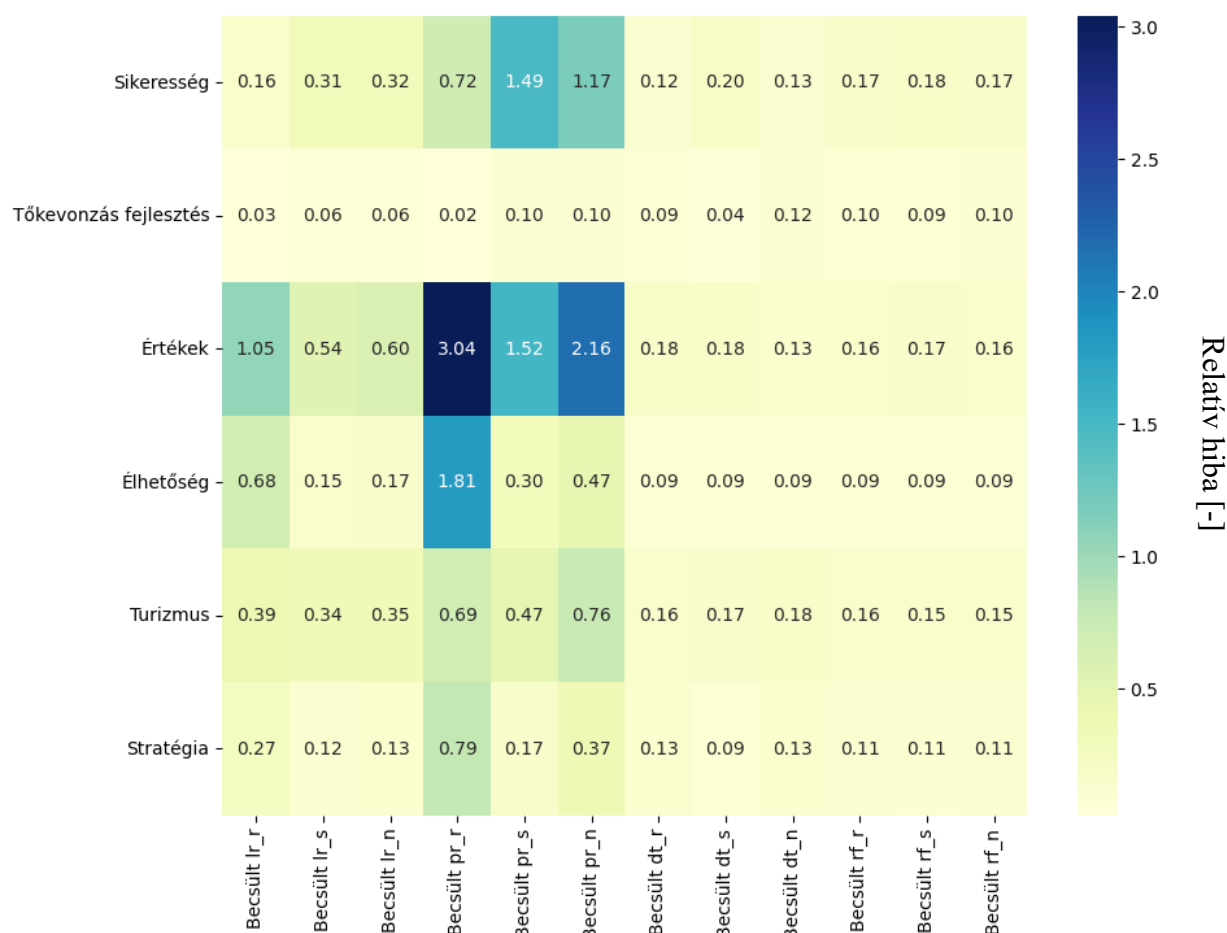
- r a relatív hiba,
- $\hat{z}$ a becsült adat,
- z pedig a valós adat.

Az eredmények értékelése kapcsán rögtön szembetűnő, hogy a polinomiális regresszió csak a megfelelő skálázással adhat jó eredményt a legtöbb esetben. Ez megegyezik a 2.2.2 alfejezet szerint elvártakkal, hiszen a polinomiális regresszió magasabb rendű hatványokat és interakciókat foglal magába, ami növelheti a numerikus instabilitást, ha az adatok nagyon különböző skálájúak. A skálázás csökkenti ezt a problémát, mivel az összes jellemző hasonló tartományba kerül. Ez lényegében lehetővé teszi, hogy az összes jellemző azonos mértékegységben legyen, ami megkönnyíti az egyes jellemzők relatív fontosságának értékelését. De egyébként még standardizálással vagy normalizálással is némely helyen felléphet a már szintén tárgyalt overfitting, különösen az értékek ismeretének megbecslése terén, így érdemes odafigyelni rá, hogy hol alkalmazzuk ezt a modellt.

Minden település kapcsán a tökevonzás fejlesztésre való igény a legjobban megbecsülhető. Átlagosan a modellek nagyon jól tudják ezt prediktálni, a legtöbb település esetén még olyan modell is van, amelyik pontosan is el tudja találni az értékét. Ez kifejezetten hasznos lehet egy polgármesternek például, hiszen ez alapján olyan településmarketing stratégiát tud kidolgozni, amiben a lakosság igényeinek megfelelő hangsúlyt helyez a tökevonzás fejlesztésére. Ehhez a ponthoz tartozik egyébként az egyetlen eset, amikor a nyers adatokon végzett polinomiális regresszió is igazán eredményes. Így kijelenthető, hogy a tökevonzás fejlesztésre való igény nagyon nagy biztonsággal becsülhető. Ez Dálnok, Esztelnek, Zabola és Csíkszentdomokos esetében a standardizált döntési fa kapcsán hibátlan becslést jelent, míg Kovászna kapcsán 2%-os hibát nyers adatokat használó polinomiális regresszió

óval. De ezen utóbbi modell egyébként a többi településen is átlagosan 2%-os hibával teljesít. Így elmondható, hogy a tökevonzás fejlesztésre való igény jól becsülhető nyers adatokat használó polinomiális regresszióval, viszont más célra nem nagyon érdemes használni ezt a megközelítést.

Szintén jól becsülhető az élhetőségről alkotott vélemény és a település stratégiával való elégedettség. Erre főként a döntési fák és a véletlen erdők alkalmasak. Az átlagos hibák a standardizált döntési fa esetén 9-9% értékűek, de külön bontva az egyes településeket természetesen kereshetők ennél jobb értékek is. Összességében szintén a döntési fák és véletlen erdők teljesítenek a legjobban a nehezebben becsülhető sikeresség megítélés, turizmus támogatottság megítélés és értékismeret jellemzők terén. Esetükben rendre a nyers döntési fa, a standardizált vagy normalizált véletlen erdő és a normalizált döntési fa javasolt átlagosan 12%, 15% és 13% hibával. Ezt jól összefoglalja a 15. ábra:



15. ábra: Az öt település átlagos relatív hibái (Forrás: saját szerkesztés)

Az is megállapítható, hogy amennyiben egyetlen modell típussal szeretnénk megoldani minden becslést, akkor a települések és elégedettségi jellemzők szerinti átlagos hiba

minimalizálása esetén a standardizált döntési fa a legjobb választás, amelynek hibája átlagosan 12,8% értékű. Így egy döntéshozó az itt bemutatott módszertan segítségével 87,2%-os pontossággal tudja megbecsülni döntéseinek elégedettségre gyakorolt hosszútávú hatását. Ez egy ilyen kicsiny mintamennyiség esetén már önmagában is kiemelkedő eredmény (Kokol, Kokol, & Zagoranski, 2022), azonban ez még tovább javítható ensemble learning segítségével (Dong, Yu, Cao, Shi, & Ma, 2020). Erről szól a következő alfejezet.

3.3. EREDMÉNYEK TOVÁBBI JAVÍTÁSA

Az ensemble learning egy olyan gépi tanulási technika, amely több modellt kombinál egy erősebb, pontosabb prediktív modell létrehozása érdekében. Az ensemble technikák különböző módon használják a modell kombinációkat, mint például bagging, boosting és stacking (Ganaie, Hu, Malik, Tanveer, & Suganthan, 2022).

A bagging, vagy bootstrap aggregating, egy olyan módszer, amely véletlenszerűen generált alhalmazokat használ az adatokból, és minden alhalmazon külön-külön tanít egy modellt. Ezeket a modelleket párhuzamosan futtatják, és a végső döntést általában a modellek átlagos vagy többségi szavazásával hozzák meg. Egy klasszikus példa a baggingre a már használt véletlen erdő, amely sok döntési fa kombinációját használja (Kadiyala, & Kumar, 2018).

A boosting, ellentétben a bagginggel, szekvenciálisan építi fel a modelleket, ahol minden új modell a korábbiak hibáit próbálja javítani. Az egyes modellek súlyozása az alapján történik, hogy mennyire jól teljesítenek az adathalmazon. A gyengén teljesítő modellek kevesebb súlyt kapnak, míg a jobban teljesítők nagyobb súlyt. Például egy ismert boosting algoritmus az AdaBoost (Freund, & Schapire, 1996), amely adaptívan állítja a súlyokat az iterációk során (Mayr, Binder, Gefeller, & Schmid, 2014).

A stacking egy másik ensemble technika, amely különböző modelleket kombinál azáltal, hogy egy meta-modellt tanít az egyes modellek kimeneteire. Ez a meta-modell arra tanul, hogy hogyan kombinálja a különböző modellek predikcióit a legjobb eredmény elérése érdekében (Naimi, & Balzer, 2018).

A különböző modellek súlyválasztása kulcsfontosságú az ensemble learning hatékonyságában. A súlyok beállítása lehetővé teszi, hogy a rendszer kiemelve a jobban teljesítő modelleket, miközben csökkenti a kevésbé hatékonyak befolyását. Ez növeli az ensemble modell általános pontosságát és megbízhatóságát.

Ezen módszertanok általános alkalmazhatóságához azonban nagy mennyiségű validációs adat szükséges az overfitting elkerülésének érdekében (Ghojogh, & Crowley, 2019).

Így ebben az esetben a kevésbé direkt kapcsolatot használó hiperparaméter optimalizáció kérdéskörével igyekeztem javítani becsléseimen.

3.3.1. Hiperparaméter optimalizáció megválasztása

A hiperparaméter optimalizálás a tanuló algoritmusok fejlesztésének egyik legsarlatosabb pontja (Bergstra, Bardenet, Bengio, & Kégl, 2011). Az adatalapú probléma megoldásának lényeges eleme az alkalmazott modell hiperparamétereinek megfelelő finomhangolása, jelen esetben a meta-modell súlyainak beállítása.

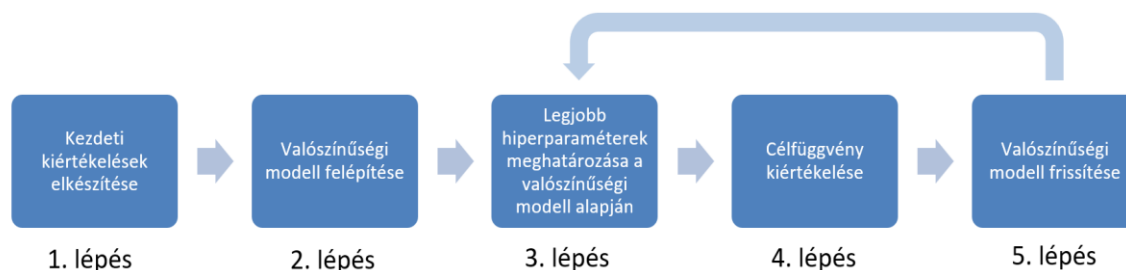
A hiperparaméterek optimalizálására többféle lehetőség is létezik (Feurer, & Hutter, 2019). A szoftverfejlesztési szempontból legegyszerűbb a manuális módszer. Ekkor kézzel történik az egyes értékek módosítása és az új futtatások indítása is, így figyelve, hogy milyen beállítások esetén kapható jobb megoldás a tanuló algoritmustól. Egy kicsivel nagyobb odafigyelést igénylő megoldás, amikor több egymásba ágyazott ciklus segítségével egy adott tartományon belül az összes lehetséges modell konfiguráció megvizsgálásra kerül. Ez abból a szempontból előnyösebb, hogy nem igényel folyamatos emberi jelenlétet, azonban továbbra is rendkívül időigényes. Így, ha az optimalizálás gyorsasága is szempont, akkor a hiperparaméterternek csak egy lokális optimumát lehet vele megtalálni. A véletlenszerű keresés is egy lehetőség, azonban itt fennál a veszélye a legjobb hiperparaméter konfigurációk kihagyásának. Ugyanakkor gyorsabb lehet, mint a rács optimalizálás.

A leghatékonyabb és legkorszerűbb módszer pedig az algoritmikus optimalizálás, mivel ez egy adott futtatáson belül intelligensen képes kihasználni a korábbi iterációk eredményeit. Alapvetően a legtöbb korszerű hiperparaméter optimalizáló algoritmus alapja a Bayes-optimalizáció (Frazier, 2018), noha vannak olyan esetek, amikor olyan speciális az optimalizálandó rendszer, hogy lehetőség van gradiens-alapú optimalizációra (Bengio, 2000), vagy kellően gyorsan kiértékelhető a rendszer ahhoz, hogy alkalmazhatók legyenek evolúciós algoritmusok (Simon, 2013). A Bayes-féle hiperparaméter optimalizáció lényege, hogy az optimalizációs iteráció minden lépésében felépít egy valószínűségi modellt az optimalizálandó célfüggvényről és ezt a modellt használja arra, hogy meghatározza, hogy a következő lépésben milyen hiperparaméter konstrukcióval kell kiértékelni a tényleges célfüggvényt. A Bayes-féle optimalizációt megvalósító algoritmusok, másként a szekvenciális modellalapú optimalizálási (Sequential Model-Based Optimization, SMBO) algoritmusok között a legjelentősebb különbség az, hogy milyen módon készítik el a célfüggvény valószínűségi modelljét. A két legnépszerűbb algoritmus, a Gauss-folyamat (Gaussian

Process, GP) és a Fa-struktúrájú Parzen Becslő (Tree-structured Parzen Estimator, TPE) közül az utóbbit választottam, mivel ez több teszt alapján is eredményesebb volt (Bergstra, Bardenet, Bengio, & Kégl, 2011).

3.3.2. Bayes-féle optimalizálás Fa-struktúrájú Parzen Becslővel

A Bayes-féle hiperparaméter optimalizálást a következő folyamatára szemlélteti:



16. ábra: Bayes-féle optimalizáció (Forrás: saját szerkesztés)

Látható, hogy az 1. és a 2. lépést csak egyszer kell elvégezni, míg a 3., 4. és 5. lépést ciklikusan ismételni kell. A megállási kritériumot jellemzően időhöz, iterációs számhoz, vagy célfüggvény értékhez szokás kötni. Az 1. lépés a szimulációs tér kezdeti véletlenszerű feltérképezését jelenti, hogy legyen alapja a valószínűségi modell felépítésének. Esetemben az Optuna könyvtár (Akiba, Sano, Yanase, Ohta, & Koyama, 2019) kísérletezési alapon meghatározott legjobb alapbeállítást, a 10 random iterációs beállítást használtam. A 2. és 5. lépés a TPE specifikus modellezés, a 3. lépés a TPE specifikus kritérium alapján való döntés, a 4. lépés pedig a becslés futtatás.

Érdeemes elsőként a kritériummal foglalkozni. Mitől lesz jó egy hiperparaméter konstrukció? Ennek meghatározására a TPE a Várható Fejlődés (Expected Improvement, EI) függvényét használja:

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y) \cdot p(y|x) dy, \quad (8)$$

ahol az eddig nem említett

- x az aktuális hiperparaméter konstrukció,
- y^* a célfüggvény küszöbértéke. Az Optuna kísérletezési alapon meghatározott legjobb alapbeállítása szerint az alsó decilis, a 25. iteráció után pedig az alsó decilis és a 25. legkisebb érték közül a kisebbik,
- y a célfüggvény aktuális értéke,
- $p(x|y)$ pedig a célfüggvény feltételes valószínűségi modellje.

Már ebből a formulából is világosan látszik, hogy olyan célfüggvény értékeket keresek ezzel az algoritmussal, amik kisebbek, mint a küszöbérték, hiszen ekkor lehet pozitív az

integrál értéke. Tehát minimalizálom a hibát. Ismerve a várható fejlődés fogalmát célszerű rátérni a TPE legfőbb elemére, a valószínűségi modellre:

$$EI_{y^*}(x) = \int_{-\infty}^{y^*} (y^* - y) \cdot \frac{p(x|y) \cdot p(y)}{p(x)} dy, \quad (9)$$

ahol az eddig nem említett:

- $p(x|y)$ a hiperparaméter konstrukció feltételes valószínűségi modellje,
- $p(y)$ a célfüggvény valószínűségi modellje,
- $p(x)$ pedig a hiperparaméter konstrukció valószínűségi modellje.

Továbbá a TPE bevezeti az $l(x)$ és $g(x)$ függvényeket úgy, hogy:

$$p(x|y) = \begin{cases} l(x) & | y < y^* \\ g(x) & | y \geq y^* \end{cases} \quad (10)$$

Itt az $l(x)$ és $g(x)$ modellek Gaussi Keverékmódként (Gaussian Mixture Model, GMM) (Resch, 2010) állnak elő úgy, hogy minden egyes adatra egy eloszlást kell illeszteni a hozzá jobbról és balról legközelebbi adatok közül a távolabb lévő távolságával megegyező szórással. Így, ha (9)-ből kiemelem $\frac{1}{p(x)}$ -t:

$$EI_{y^*}(x) = \frac{1}{p(x)} \int_{-\infty}^{y^*} (y^* - y) \cdot p(x|y) \cdot p(y) dy, \quad (11)$$

akkor a Várható Fejlődés (10) alapján felírható úgy is mint:

$$EI_{y^*}(x) = \frac{1}{p(y < y^*) \cdot l(x) + (1 - p(y < y^*)) \cdot g(x)} \int_{-\infty}^{y^*} (y^* - y) \cdot l(x) \cdot p(y) dy, \quad (12)$$

amiből kiszámítható:

$$EI_{y^*}(x) = \frac{p(y < y^*) \cdot y^* - \int_{-\infty}^{y^*} p(y) dy}{p(y < y^*) + (1 - p(y < y^*)) \cdot \frac{g(x)}{l(x)}} \quad (13)$$

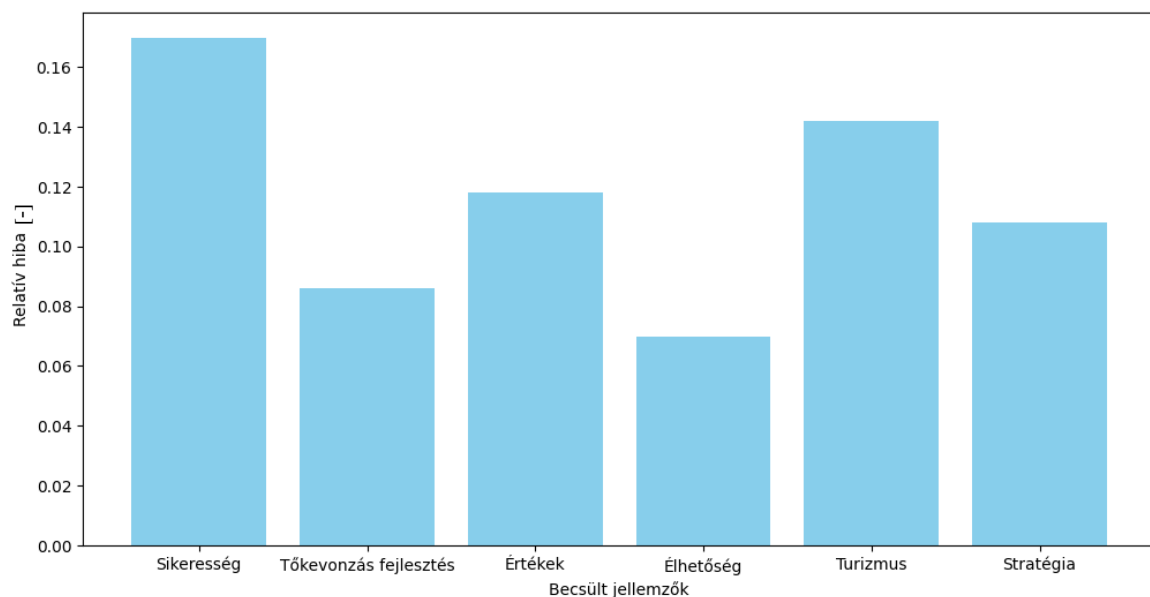
Ebből pedig látszik, hogy ha a $\frac{g(x)}{l(x)}$ kifejezés csökken, akkor a (13) tört nevezője is csökken, ezáltal a tört, azaz az $EI_{y^*}(x)$ nő, vagyis ekkor van a legnagyobb esélyem javítani a becslésem. Ez gyakorlatilag azt jelenti, hogy a küszöbérték feletti célfüggvény értékek

küszöbérték alatti célfüggvény értékekhez viszonyított aránya csökken. Tehát a TPE algoritmus mindig a kisebb hibájú becsléseket eredményező hiperparaméter konstrukciókhoz közeli konstrukciót igyekszik választani.

3.4. OPTIMALIZÁLT EREDMÉNYEK BEMUTATÁSA

Ezzel az optimalizációval sikerült súlyokat választani az egyes becslést végző modelleknek. Ennek alapján létrejött egy olyan ensemble modell, amely minden más modellnél jobban teljesít, de nem illeszkedik túl az adatokra. Ez egyszerűségéből is látható, hiszen a legjobb modellnek az adódott, amely 81%-ban veszi figyelembe a standardizált döntési fa és 19%-ban a normalizált döntési fa becslését.

Ezzel a módszertannal az egy modelles esethez képest az átlagos hiba tehát csökkenthető. A 12,8%-os érték helyett lecsökken a teljes értékelési tartomány egynolcadára, azaz 12% alá, egész pontosan 11,6%-ra. Ez azt jelenti, hogy egy döntéshozó a fejlesztett eszköz segítségével a településen lakók véleményéről el tudja dönteni, hogy egy 8 fokú skálán hol helyezkedik el átlagosan. Ezen legjobb meta-modell eredményei a 17. ábrán láthatók.



17. ábra: Az optimalizált meta-modell relatív hibái (Forrás: saját szerkesztés)

Megfigyelhető, hogy a hiba típusok egymáshoz viszonyított aránya jelentősen nem változik. Ami könnyebben becsülhető volt az könnyebben becsülhető maradt és vice-versa. Azonban az átlagos hiba nagyságát $\frac{12,8\% - 11,6\%}{12,8\%} = 9,2\%$ -kal sikerült csökkenteni, ami jelentős lehet egy vezető számára.

4. KONKLÚZIÓ

Jelen munka részletes elemzést nyújt a településmarketing és a hozzá kapcsolódó adatalapú technológiák szerepéről a 21. században. A kutatás fókuszja a települések vonzerének és versenyképességének növelése volt, különös tekintettel a lakosság elégedettségének mérésére és előrejelzésére. Kiemelésre kerül, hogy a gépi tanulás és adatalapú technikák alkalmazása kulcsfontosságú a hatékony településmarketing stratégiák kidolgozásában.

Kitértem rá, hogy a településvezetők számára kulcsfontosságú az adatokon alapuló döntéshozatal és a prediktív modellezés. Az elemzés során különböző gépi tanulási modelleket használtam, és azokat ensemble learning módszerrel kombináltam, Fa-struktúrájú Parzen Becslővel optimalizáltam. Alkalmazott kutatásom kiterjedt a lineáris- és polinomiális regresszió, a döntési fa, a véletlen erdő, a normalizálás és a standardizálás használatára is. Ez lehetővé tette, hogy még kis mintaszám esetén is pontos előrejelzéseket lehessen kapni a lakosság elégedettségére vonatkozóan.

Munkám kiemeli, hogy a településmarketing sikere szorosan összefügg a helyi lakosság igényeinek megértésével és kielégítésével. Megállapítható, hogy a modern informatikai-matematikai technikák és a gépi tanulás alkalmazása jó jövőbeli irány lehet településmarketing számára. Az eredmények azt mutatják, hogy az adatvezérelt megközelítések jelentősen hozzájárulhatnak a településvezetők döntéshozatali képességének javításához, lehetővé téve számukra, hogy pontosabban előrejelezzék és értékeljék a különböző stratégiák hatását a lakosság elégedettségére. Ennek kapcsán 11,6%-os átlagos becslési hibát, azaz 88,4%-os pontosságot sikerült elérnem a sikeresség megítélés, a tökevonzás fejlesztési igény, az értékismeret, az élıhetőség megítélés, a turizmus támogatottsággal való elégedettség és a stratégiával való megelégedés becslésének terén.

Összefoglalva, a dolgozat új megvilágításba helyezi az adatok és a gépi tanulás jelentőségét a településmarketing területén. A vizsgálatok eredményei alátámasztják, hogy az adatvezérelt döntéshozatal és a prediktív modellezés kulcsfontosságú eszközök a települések jövőbeli fejlődésének előmozdításához, hozzájárulva ezzel a helyi közösségek jólétéhez és a települések versenyképességének növeléséhez.

IRODALOMJEGYZÉK

- Ahsan, M. M., Mahmud, M. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, 9(3), 52.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019, July). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2623-2631).
- Andersson, M., & Henrekson, M. (2015). Local competitiveness fostered through local institutions for entrepreneurship. *The Oxford handbook of local competitiveness*, 145-190.
- Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012, April). The K-fold Cross Validation. In *ESANN* (pp. 441-446).
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection.
- Asprogerakas, E. (2006). City competition and urban marketing: The case of tourism industry in Athens.
- Audretsch, D. B., Link, A. N., & Walshok, M. L. (Eds.). (2015). *The Oxford handbook of local competitiveness*. Oxford University Press.
- Beluszky, P., & Györi, R. (1999). A magyarországi városhálózat és az EU-csatlakozás. *Tér és Társadalom*, 13(1-2), 1-30.
- Bengio, Y. (2000). Gradient-based optimization of hyperparameters. *Neural computation*, 12(8), 1889-1900.
- Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. *Advances in neural information processing systems*, 24.
- Bibri, S. E. (2021). Data-driven smart eco-cities of the future: an empirically informed integrated model for strategic sustainable urban development. *World Futures*, 1-44.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.
- Boisen, M., Terlouw, K., Groote, P., & Couwenberg, O. (2018). Reframing place promotion, place marketing, and place branding-moving beyond conceptual confusion. *Cities*, 80, 4-11.
- Bonakdar, A., & Audirac, I. (2020). City branding and the link to urban planning: Theories, practices, and challenges. *Journal of Planning Literature*, 35(2), 147-160.
- Braun, E. (2008). *City Marketing: Towards an integrated approach* (No. EPS-2008-142-ORG).
- Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (2017). *Classification and regression trees*. CRC press.
- Búzás, N., Kállay, L., & Lengyel, I. (2003). *Kis-és középvállalkozások a változó gazdaságban*. JATEPress.
- Cassel, S. H. (2008). Trying to be attractive: Image building and identity formation in small industrial municipalities in Sweden. *Place Branding and Public Diplomacy*, 4, 102-114.
- Cevallos, S. (2022). Urban marketing for smart cities. *ORAȘE INTELIGENTE ȘI DEZVOLTARE REGIONALĂ*, 6(02), 91-96.

- Deffner, A., Karachalis, N., Psatha, E., Metaxas, T., & Sirakoulis, K. (2020). City marketing and planning in two Greek cities: plurality or constraints?. *European Planning Studies*, 28(7), 1333-1354.
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14, 241-258.
- Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. *Automated machine learning: Methods, systems, challenges*, 3-33.
- Frazier, P. I. (2018). A tutorial on Bayesian optimization. *arXiv preprint arXiv:1807.02811*.
- Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In *icml* (Vol. 96, pp. 148-156).
- Gajzágó, G. (2019). A stratégiai tervezés alakulása Dunaújváros térségében. *Területfejlesztés és Innováció*, 12(2), 19-32.
- Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, 105151.
- Garamhegyi, Á., & Révész, B. (2000). A területi marketing, mint az önkormányzatok versenyképességének egy lehetséges eszköze. In B. Farkas & I. Lengyel (Szerk.), *Versenyképesség - regionális versenyképesség* (124-114. o.). Szeged: JATEPress.
- Ghojogh, B., & Crowley, M. (2019). The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. *arXiv preprint arXiv:1905.12787*.
- Goovaerts, P., Van Biesbroeck, H., & Van Tilt, T. (2014). Measuring the effect and efficiency of city marketing. *Procedia Economics and Finance*, 12, 191-198.
- Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology*, 12(4), 1243-1257.
- HBH Stratégia és Fejlesztés Kft., & Collective-Intelligence Kft. (2018). Összefoglaló Kutatási Jelentés: Az önkormányzati humán kompetenciák és képességek. KÖFOP-2.3.3-VEKOP-16-2016-00001 azonosító jelű "Helyi versenyképesség-fejlesztési kutatási program" című kiemelt projekt kapcsán.
- Harmaakorpi, V., Kari, K., & Parjanen, S. (2008). City design management as a local competitiveness factor. *Place Branding and Public Diplomacy*, 4, 169-181.
- Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362.
- Hegyi, F. B. (2007). Marketingorientált városfejlesztés [Marketing Oriented City Development]. *Tér és Társadalom*, 21(3), 129-141.
- Herrschel, T. (2013). Competitiveness and sustainability: can 'smart city regionalism'square the circle?. *Urban Studies*, 50(11), 2332-2348.
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(03), 90-95.
- Jeney, L., Kulcsár, D., & Tózsá, I. (2013). Gazdaságföldrajzi tanulmányok közgazdászoknak. BCE Gazdaságföldrajzi és Jövőkutató Tanszék.
- Kadiyala, A., & Kumar, A. (2018). Applications of python to evaluate the performance of bagging methods. *Environmental Progress & Sustainable Energy*, 37(5), 1555-1559.
- Kavaratzis, M. (2007). City marketing: The past, the present and some unresolved issues. *Geography compass*, 1(3), 695-712.

- Kavaratzis, M., & Hatch, M. J. (2021). The elusive destination brand and the ATLAS wheel of place brand management. *Journal of travel research*, 60(1), 3-15.
- Kleijnen, J. P. (2017). Design and analysis of simulation experiments: Tutorial. In *Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences* (pp. 135-158). Springer International Publishing.
- Kokol, P., Kokol, M., & Zagoranski, S. (2022). Machine learning on small size samples: A synthetic knowledge synthesis. *Science Progress*, 105(1), 00368504211029777.
- Komasi, H., Hashemkhani Zolfani, S., Prentkovskis, O., & Skačkauskas, P. (2022). Urban Competitiveness: Identification and Analysis of Sustainable Key Drivers (A Case Study in Iran). *Sustainability*, 14(13), 7844.
- Kuzior, A., Sobotka, B., Filipenko, A., & Kuzior, P. (2019). Marketing communications of administrative organs of local governance and local community. *Marketing and Management of Innovations*, (2), 314-325.
- Lafuente, E., Araya, M., & Leiva, J. C. (2022). Assessment of local competitiveness: A composite indicator analysis of Costa Rican counties using the 'Benefit of the Doubt' model. *Socio-Economic Planning Sciences*, 81, 100864.
- Ma, W., de Jong, M., Hoppe, T., & de Bruijne, M. (2021). From city promotion via city marketing to city branding: Examining urban strategies in 23 Chinese cities. *Cities*, 116, 103269.
- Majláth, M., Reicher, R. Z., Pecze, K., & Kolnhofer-Derecskei, A. (2022). Helymarketing, mint kutatási téma-szekunder kutatási eredmények a hazai publikációkra vonatkozóan.
- Malecki, E. J. (2017). Economic Competitiveness and regional development dynamics. *Handbook of Regions and Competitiveness*; Thompson, P., Huggins, R., Eds, 136-152.
- Mayr, A., Binder, H., Gefeller, O., & Schmid, M. (2014). The evolution of boosting algorithms. *Methods of information in medicine*, 53(06), 419-427.
- McKinney, W. (2010, June). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, No. 1, pp. 51-56).
- Nagy, S., Piskóti, I., Dankó, L., Molnár, L., & Marien, A. (2020). A társadalmi marketing paradigmái-elméleti-módszertani alapoató kutatás: Az Országos Tudományos Kutatási Alap által támogatott kétéves kutatási program beszámolója (OTKA száma: K 81718).
- Nagy, S., Piskóti, I., Molnár, L., Gulyásné Kerekes, R., & Karajz, S. (2015). Termékek és szolgáltatások piaci potenciáljának komplex vizsgálati modellje
- Naimi, A. I., & Balzer, L. B. (2018). Stacked generalization: an introduction to super learning. *European journal of epidemiology*, 33, 459-464.
- Panyor, Á., Győriné Kiss, E., & Gulyás, L. (Szerk.). (2011). Regionális turisztikai menedzsment modul: Bevezetés a desztináció menedzsmentbe. Keszthely: Szegedi Tudományegyetem, Debreceni Egyetem, Pannon Egyetem.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Piskóti, I. (2017). A desztinációmárketing hatékonysága – modellek, mérések, trendek és eszközök. *Prosperitas*, 4(2), 7-43.
- Piskóti, I. (2022). A coopetition-modell új dimenziói a helymarketingben. *Magyar Gazdaságföldrajzi És Településmarketing Műhely*, 1(1), 01.002. e-ISSN: 2939-516X.

- Piskóti, I., Hidvéginé Molnár, J., Pataki, S., Schumpler, H., & Gulyás, I. (2007). Desztináció-menedzsment lépésről-lépésre. Módszertani füzet desztináció-menedzsment szervezetek létrehozásához és működtetéséhez az Észak-magyarországi Turisztikai Régió példáján–munkaanyag. Eger–Miskolc.
- Puczkó, L., & Tózsá, I. (2017). Településmarketing V. tankönyv (3. pillér).
- Rechnitzer, J. (1995). Some sketch points for the contemplation and implementation of settlement marketing. *Tér és Társadalom*, 9(1-2), 5-16.
- Resch, B. (2010). Mixtures of Gaussians: A tutorial for the course computational intelligence. Signal Processing and Speech Communication Laboratory. <http://www.igi.tugraz.at/lehre/CI/tutorials/MixtGaussian/MixtGaussian.html>, Inffeldgasse 16c, last updated, 24.
- Segal, M. R. (2004). Machine learning benchmarks and random forest regression.
- Sgambati, S., & Carvalho, L. (2023). Benchmarking urban competitiveness for economic recovery: an application to Porto and Lisbon. *Journal of Place Management and Development*.
- Simon, D. (2013). *Evolutionary Optimization Algorithms*, John Wiley & Sons.
- Stănciulescu, G. C. (2009). The role of urban marketing in the local economic development. *Theoretical and Empirical Researches in Urban Management*, 4(1 (10)), 114-135.
- Štetić, S., Šimičević, D., Pavlović, S., & Stanić, S. (2014). Business tourism competitiveness model: Competitiveness of Serbia as a business tourism destination. *Calitatea*, 15(S5), 176-191.
- Tózsá, I. (2011). *City on Sale an Introduction to Urban Regional Marketing*.
- Tózsá, I. (2014). A településmarketing elmélete. Előadás a Kassai Műszaki Egyetem Gazdaságtudományi Karának Első Nyári Egyetemén, 2011. június 9 - július 3, Kassa, Szlovákia.
- Van Rossum, G., & Drake, F. L. (2009). *Introduction to python 3: python documentation manual part 1*. CreateSpace.
- Vanolo, A. (2017). *City branding: The ghostly politics of representation in globalising cities*. Taylor & Francis.
- Wang, D., Miwa, T., & Morikawa, T. (2020). Big trajectory data mining: a survey of methods, applications, and services. *Sensors*, 20(16), 4571.
- Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern recognition*, 48(9), 2839-2846.
- Ye, X., Fan, Y., Miao, J., & He, Z. (2022). The competitiveness of regional urban system in Hubei province of China. *Land*, 11(6), 879.
- Yezhova, A. (2020). Theoretical Concept and Definition of City Marketing. *Public Administration and Society*, 21, 121-130.
- Yezhova, A. (2021). Vision and Market Segmentation in Urban Strategy through Marketing Approach: A Case Study of Sumy City. *Acta Universitatis Sapientiae, European and Regional Studies*, (20), 48-62