

BUDAPESTI MŰSZAKI ÉS GAZDASÁGTUDOMÁNYI EGYETEM
GÉPÉSZMÉRNÖKI KAR

KOCSIS KENDE ÉS KÖVENDI JÓZSEF

**MÁSODIK HULLADÉKHASZHOSÍTÓ ERŐMŰ IN-
TEGRÁLÁSA BUDAPEST TÁVHŐELLÁTÁSÁBA**

TDK dolgozat

Konzulensek:

Dr. Bokor Balázs
adjunktus, BME ÉPGET

Dr. Szücs Botond
adjunktus, BME EGR

Budapest, 2023 november 5.

KÖSZÖNETNYILVÁNÍTÁS

Köszönettel tartozunk Orbán Tibornak, a FŐTÁV Zrt. műszaki vezérigazgató-helyettesének, aki segítségével nagyban elősegítette a tanulmány megvalósulását.

A kulturális és Innovációs Minisztérium ÚNKP-22-2-I-BME-66 kódszámú Új Nemzeti Kiválóság Programjának a Nemzeti Kutatási, Fejlesztési és Innovációs Alapból finanszírozott szakmai támogatásával készült.



TARTALOMJEGYZÉK

1. Bevezetés	3
1.1. Előzetes információk.....	4
1.2. A tanulmány célja	4
1.3. A tanulmány felépítése.....	5
2. Adatstruktúra	7
3. Távhőcsatlakozás elemzése	8
3.1. Hővesztések.....	8
3.2. Energetikai és gazdasági elemzés	14
3.2.1. Távhővezetékek lefektetése	14
3.2.2. Hőközetek energetikai elemzése.....	15
3.2.3. Gazdasági elemzés	20
4. Hőteljesítmény-igény előrejelzés.....	23
4.1. Gépi tanulás bevezetése	23

4.1.1. Gépi tanulás alkalmazási területei.....	24
4.1.2. Forward vs. adat alapú előrejelzés	24
4.2. Alkalmazott algoritmusok.....	25
4.2.1. Időbeliséget figyelembe nem vevő algoritmusok	26
4.2.2. Időbeliséget figyelembe vevő algoritmusok	29
4.3. Adatok elemzése	30
4.3.1. Jelentkező igények	30
4.3.2. Felhasznált adatok.....	33
4.4. Adatok előfeldolgozása.....	34
4.4.1. Train-test split	35
4.4.2. Hibás adatok szűrése.....	35
4.4.3. Adatok skálázása.....	39
4.5. Hiperparaméter optimalizáció.....	39
4.5.1. Sklearn	40
4.5.2. Hiperparaméterek optimalizáció megvalósítása	40
4.6. Eredmények kiértékelése	42
4.6.1. Hibaértékelő módszerek.....	42
4.6.2. A Főtáv empirikus modellje.....	43
4.6.3. Eredmények vizualizálása.....	44
5. Eredmények összegzése.....	55
6. Források	57
7. Mellékletek	59
7.1. Végző hiperparaméterek és jelentésük	59
7.1.1. Hiperparaméterek jelentése.....	59
7.1.2. Optimalizálási eredmények.....	61

1. BEVEZETÉS

Budapest kiterjedt távhőellátó rendszerrel rendelkezik, közel 240 ezer lakás hőellátásáért felelős, amely lakások száma a közeljövőben a közigazgatásban megfogalmazott tervek szerint tovább fog növekedni. A modern távhőszolgáltatás alkalmazása több szempontból is jelentős előnyökkel járhat a társadalomnak mind individuális, mind pedig társadalmi szempontból nézve. Egyéni szempontból nézve a lakossági fogyasztók számára szolgáltatott hőár sokkal kedvezőbb értéket tud felvenni, mint amelyet a földgáz alapú egyéni kazánokkal el lehet érni a rezsicsökkentés figyelembevételével. Természetesen a 2020 óta fennálló európai geopolitikai helyzet által előidézett instabil fosszilis energia-hordozó ellátásbiztonság és a megemelkedett árak miatt nem biztosított a rezsicsökkentett ár fenntartása, ezáltal tovább növelve a távhő helyzetelőnyét. Társadalmi szempontból előnyös a távhőrendszerek által nyújtott centralizált energiatermelés, amely magasabb hatásfokot, illetve jelentős mértékben csökkentett légszennyezést biztosít a városi lakóövezetek számára.

A távhőrendszerek harmadik nagy előnye, hogy az őket ellátó erőművek által alkalmazott tüzelőanyagok anyagi minősége nagy mértékben változhat az erőmű típusának függvényében. Jelenleg a város távhőrendszerének több, mint 80%-át földgáz alapú erőművekkel látják el, amely a fentebb említett problémák fényében jelentős mértékben megemeli a szolgáltatási költséget, lévén, hogy a lakossági fogyasztóknak szolgáltatott, rendszerileg befagyasztott távhőárakon generált veszteségeket a szolgáltató kénytelen a nemlakossági fogyasztókra ráterhelni. Ezen probléma kézenfekvő megoldása a hulladékégető kombinált ciklusú, villamos- és hőenergiát egyaránt előállító erőművek integrálása a hálózatba. Már több évtizede léteznek tervek a HUHA 2 második budapesti hulladékhasznosító mű létesítésére, amely megvalósulásával további 230 000 tonna települési szilárd hulladékot (TSZH) lenne képes energetikailag hasznosítani, ezáltal megoldást nyújtva a távhőrendszer problémáira, miközben az Európai Unió által meghatározott – a hulladéklerakást megelőzni igyekvő – irányelveknek megfelel. [1.]

1.1. Előzetes információk

A jelenlegi tanulmány keretein belül azt feltételezzük, hogy a tavalyi évben végzett, jelen tanulmány előzménye során tárgyalt HUHA 2 hulladékhasznosító erőmű megépült és üzemképes állapotban van. A megvalósulás indoklásának alapjául szolgál a 2022-ben készült tanulmány, amely gazdaságilag elemezte az erőmű beillesztését a távhőrendszerbe és alátámasztja a létesítést a gyors megtérülési adataival. [2.] Az erőmű egyes távhőkörzetekre való csatlakozása jelentős befolyással bír a rendszer paramétereire, mint a távhővezeték telepítésének beruházási költségeire, ezen vezeték szállítási hőveszteségi értékeire, az egyes hőközrzetek által támasztott hőtéljesítmény igényekre az év során és ezen hőigények által befolyásolt üzemi paraméterekre, mint a kombinált hatásfok és a kihasználtság. Az erőmű az eredeti tervek szerint képes csatlakozni egy egyesített budapesti távhő körvezetékre, amely összekötötte volna a város távhőkörzeteinek döntő többségét, ezáltal előnyös üzemi paramétereket és biztos felvevőpiacot biztosítva a termelőknek. Ezen beruházás a tanulmány elkészültének pillanatában nem valósult meg és magas beruházási költsége miatt valószínűsíthetően ez a helyzet továbbra is fenn fog maradni az előrelátható jövőben. Ennek fényében érdemben az Ócsai út és az M0-ás körgyűrű kereszteződésénél elhelyezkedő HUHA 2 erőműhöz legközelebb eső három távhőkörzetre való csatlakozás elemzése mellett döntöttünk.

1.2. A tanulmány célja

Jelen tanulmány célja a HUHA 2 hulladékégető mű létesülésének elősegítése, amely során a HUHA 2 hulladékerőmű távhőcsatlakozásának lehetőségeit vizsgáltuk meg a Dél-budapesti távhőkörzetek esetén a kelenföldi távhőkörzetre, csepeli távhőkörzetre és a kispesti távhőkörzetre. Célunk volt az energetikai, illetve gazdasági szempontból optimális csatlakozási struktúra meghatározása.

Ezt követően a csatlakoztatott hulladékhasznosító erőmű megfelelő szabályozhatóságának érdekében célunk volt egy – a jelenleg alkalmazott empirikus - hőtéljesítmény-igény elemző rendszer megalkotása, amely lehetővé tenné a HUHA 2 és az ellátott körzetek között húzódó hosszú összekötő távhővezeték ellenére a pontos kiadott hőtéljesítmény szabályozást. Ezen célt számos közelmúltbeli tudományos közleményhez

haszonlóan gépi tanulási módszerekkel közelítettük meg a bizonyítottan pontos előrejelzési képességük miatt.

A hőteljesítmény-igények pontos előrejelzésének segítségével csökkenthető a teljesítmény-igény, a visszatérő távhővezetékben mért hőmérséklet, ezáltal növelve az erőmű kombinált hatásfokát, illetve pontosan illeszthető a termelés a fogyasztók igényeihez az év minden évszakában, növelve a komfort mértékét és csökkentve a hirtelen fel és leterhelések szükségességét.

1.3. A tanulmány felépítése

Az elemzés két fázisból áll össze: az első lépésben az optimális távhőcsatlakozási struktúrát határoztuk meg. Ennek oka, hogy a fent említett hőkörizetek mind méretben, mind pedig hőigényben változó értékekkel rendelkeznek, ezáltal fontos tudni, hogy energetikailag, illetve gazdaságilag mely hőkörizetekre éri meg csatlakozni. Az elemzés során kiemelkedő fontossággal bír az R1 energy efficiency formula figyelembevétele, amely meghatározza egy hulladéksemleresítő létesítmény hulladékégető, avagy hulladékhasznosító státuszát. Az előbbi státusz mindenképpen elkerülendő, mivel a rendeletek fényében ennek besorolása minimálisan jobb csak a hulladéklerakáshoz képest és ezáltal nagymértékű büntetést von maga után minden tonna hulladék elégetése, amely legfőképpen a teljes CO₂ kvóta kifizetési köteleességben nyilvánul meg. [3.] Ezen érték a külső hőmérséklet emelkedésével járó hőteljesítmény-igény csökkenésével párhuzamosan csökkenni fog, emiatt ügyelni kell a csatlakozások elemzése során a nyári hónapokban támasztott hőigényekre az egyes csatlakozási konstrukciók esetén.

A távhőcsatlakozások elemzéséhez elengedhetetlen a korábban említett három közeli hőkörizet, a kelenföldi, a csepeli- és a kispesti távhőkörizet hőteljesítmény igényének előrejelzése, mivel a város déli határára telepített HUHA 2 erőmű összeköttetési távolsága miatt fokozottan érzékeny a szabályozási jel előrejelezhetőségére a megfelelő igények fedezéséhez. Ezen terület kiemelkedő jelentőséggel bír a termelők és a szolgáltatók számára is, mivel a fejlett előrejelzési rendszerek alkalmazásával nagymértékben pontosítani lehet a termelési adatokat, akár órás bontásban, amely segítségével elérhetővé válnak a célkitűzések között tárgyalt eredmények.

Ennek megvalósítására a tanulmány második részében gépi tanulás alapú módszert dolgozunk ki, amely segítségével a 6, illetve 24 órás előrejelzési adatokat tudunk biztosítani

a távhőszolgáltató részére, mindezt magas pontossággal. Jelenleg a Főtáv Zrt. által kezelt budapesti távhőszolgáltatási rendszer nem alkalmaz hasonló technológiát, a predikciós módszerük empirikus módszereken alapul, ahol a hőigényeket egy – a külső hőmérséklet alakulásához kötött – lineáris függvény szerint határozzák meg, illetve üzem közben módosításokat végeznek rajta az igények változásának függvényében. A jelenlegi módszer alkalmazásával azonban fennáll az a probléma, hogy a hőigények nem csupán a külső hőmérséklet függésében állnak, hanem befolyásoló tényezőként hat rájuk több externális paraméter, mint az időjárás és fogyasztói szokások, illetve internális paraméter, mint a korábbi tömegáram, előremenő hőmérséklet és teljesítmény adatok.

A paraméterek függésének komplexitása miatt nagyon összetett feladat a hőteljesítmény igények pontos előrejelzése akár félnapos időtávban is. Habár a Főtáv Zrt. által alkalmazott módszer sok évtizednyi tapasztalaton alapszik, a gépi tanulás alapú módszerek sokkal nagyobb pontosságú előrejelzési lehetőséget ígérnek, emiatt a munkánk során ezen megközelítést alkalmaztuk. A gépi tanulást több korábbi tanulmány során is alkalmazták hőigények előrejelzésére, többek között Idowu et al., és Johansson et al. [4.][5.]

A tanulmány során Python programnyelvet használtunk az algoritmusok megírásához és futtatásához.

2. ADATSTRUKTÚRA

A tanulmány alapját képző adatok nagy részét a Főtáv Zrt. szolgáltatta. Ezen információ csomagok három évnnyi mérési adatot tartalmaznak a három vizsgált hőkörizet mind-egyikére. A mért értékek az alábbiak voltak: előremenő hőmérséklet ($T_{e,i}$) és visszatérő hőmérséklet ($T_{v,i}$) a hőkörizetek minden ágában, továbbá tömegáram (\dot{m}_l) és hőteljesítmény értékek (\dot{Q}_l) az összes ág esetén mindhárom körizetben. A mérési intervallum órás volt, amely így a négy év esetében 35 040 adatsort jelent. Az adatokat az termelés helyszínen, az erőművekben, avagy fűtőművekben mérték. Ezen felül rendelkezésünkre bocsátották a jelenleg használatban lévő vezetékhalózati korát, szigeteltségi állapotát, átmérőit és átmérőfüggő veszteségét. Ezen felül ismerjük a vezetékek fektetési hosszát a fektetés- és szigetelés módját az átmérő függvényében is.

Továbbá lényeges bemenő információ az aggregált hálózati veszteségek százalékos ismerete havi bontásban.

A Főtáv adatsorain kívül szükségünk volt az elemzések elvégzéséhez egy – az azonos időtávra vonatkozó pontos, órás osztású – meteorológiai adatsorra, amely lehetőleg tartalmazza az összes olyan paramétert az időjárással kapcsolatban, amely hatással lehet a fogyasztói viselkedésre, fűtési igényekre és a vezetékszakaszok hőveszteségére. Ezen adatokat Meteostat weboldaláról kértük le, ahol órás osztásban rendelkezésre álltak ezen információk. [6.]

- Külső hőmérséklet ($T_{k,i}$) [°C]
- Harmatpont ($T_{h,i}$) [°C]
- Relatív nedvességtartalom (ϕ_l) [%]
- Szélirány ($wdir_i$) [°]
- Szélsebesség ($v_{szél,i}$) [m/s]
- Légnyomás (p_i) [Pa]
- és Időjárás kód ($coco_i$) [-]

Ezen adatokon felül meghatároztuk a talajhőmérsékletet fél méteres mélységben - ami megfelel a vezetékek fektetési mélyégének – egy átlagos év során. [7.]

3. TÁVHŐCSATLAKOZÁS ELEMZÉSE

A tanulmány első részében a korábban említett HUHA 2 erőműhöz való távhőcsatlakozás optimális struktúrájával foglalkozunk. Első lépésként a távhő szállítási hőveszteségeit értékeljük.

3.1. Hőveszteségek

Az **Adatstruktúra** fejezetben foglaltak szerint a hőveszteségi elemzés alapjául a Főtáv Zrt. által szolgáltatott adatrendszer szolgált, amely megadta az egyes vezetékszakaszok átmérőjét, fektetési módját (földben fektetett vagy szabadon vezetett), szigetelési fajtáját, illetve a vezetékszakaszok hosszát. Ezen adatokon felül meghatároztuk az egyes vezeték hőátbocsátási tényezőjét az átmérő függvényében 1"-os mérettől DN1000-es átmérőig. [8.] Ennek struktúrájára az alábbi, **1. táblázat** ad példát.

1. táblázat: Vezetékek hőátbocsátási tényezője az átmérő és a szigeteltség függvényében, részlet [8.]

Átmérő	ISOPLUS		ELŐSZIGETELÉS NÉLKÜL	
	Szabadon vezetett U [W/mK]	Földben vezetett U [W/mK]	Szabadon vezetett U [W/mK]	Földben vezetett U [W/mK]
DN 20	0,1285	0,1337	0,2951	0,3071
DN 25	0,1550	0,1625	0,3094	0,3244
DN 32	0,1597	0,1661	0,3421	0,3558
DN 40	0,1820	0,1904	0,3720	0,3891
DN 50	0,2030	0,2122	0,4033	0,4216
DN 63	0,2203	0,2306	0,4312	0,4513
DN 65	0,2376	0,2489	0,4592	0,4811
DN 75	0,2419	0,2534	0,4755	0,4981
DN 80	0,2462	0,2579	0,4919	0,5152
DN 90	0,2525	0,2639	0,5156	0,5390

A táblázatban foglalt, előszigetetlen csövek hőátbocsátási tényezőit a **2. táblázat** értékei alapján arányosítás alapján határoztuk meg. A táblázat a terjedelmi követelmények betartásának érdekében az adatok részletét tartalmazza csak.

$$U_{előszig.nélküli} = U_{ISOPLUS} \cdot a$$

Ahol:

- $U_{előszig.nélküli}$ = előszigetelés nélküli vezeték hőveszteségi tényezője [W/mK]
- $U_{ISOPLUS}$ = előszigetelt vezeték hőveszteségi tényezője [W/mK]
- a = arányossági tényező a **2. táblázatban** foglalt értékek alapján [1]

2. táblázat: Fajlagos hőveszteségi értékek és arányszámuk, részlet

Fajlagos hőveszteség			
	Hagyományos	Előszigetelt	Arányszám
DN	kWh/nyvm/év	kWh/nyvm/év	[-]
DN 20	266,28	115,94	2,30
DN 25	286,63	143,60	2,00
DN 32	314,66	146,88	2,14
DN 40	346,23	169,40	2,04
DN 50	376,96	189,74	1,99
DN 63	405,49	207,15	1,96
DN 65	434,02	224,56	1,93
DN 75	448,48	228,14	1,97
DN 80	462,95	231,72	2,00
DN 90	486,98	238,43	2,04
DN 100	511,003	245,142	2,085

Az ismertett adatok segítségével meghatároztuk az összes hőkörzetben található vezetékek számára az $U \cdot l$ értéket, amely az adott vezetékszakasz hőátbocsátási tényezőjének és a vezeték hosszának szorzata. Ez a lépés azért volt szükséges, mert így a szorzat additív lett, ami nem mondható el a hőátbocsátási tényezőről. Ezek alapján meghatároztuk az összes hőkörzetben található $\sum_1^n U \cdot l$ értéket. A kapott adatokat a **3. táblázatban** láthatóak összegezve, a számunkra lényeges körzetek adatai sárga színnel kiemelten láthatóak.

$$U \cdot l = \sum_1^n U_n \cdot l_n$$

Ahol:

- $U \cdot l$ = az adott hőkörzetre jellemző hőveszteségi érték [W/m]
- U_n = az „n”-edik vezetékszakasz hőátbocsátási tényezője [W/mK]
- l_n = az „n”-edik vezetékszakasz hossza [m]

3. táblázat: Az egyes hőkörzetek $U \cdot l$ értékei

Összegző tábla	
Hőkörzet	$U \cdot l$ [W/K]
Füredi	43 474,67
Kispest	108 600,81
Rákoskeresztúri	12 116,51
Újpalota	28 504,37
Bimbó utca	-
Donáti utca	-
Észak-budai	100 386,41
Észak-pest	143 805,29
Keleti K. utca	-
Pincszer utca	-
Toboz utca	-
Budafok	-
Csepel	54 301,13
Kelenföld	162 339,46
Rózsakert	2 535,90

Az egyes körzetekre jellemző értékek felhasználásával meghatározható az adott körzetben található összes távhővezeték tapasztalható hővesztesége.

$$\dot{Q}'_{ker,i} = U \cdot l \cdot \left(\frac{T_{e,i} - T_{v,i}}{2} - T_{talaj,i} \right)$$

Ahol:

- $\dot{Q}'_{ker,i}$ = az adott területben tapasztalt hőteljesítmény veszteség [W]
- $T_{e,i}$ = távhő előremenő hőmérséklet [°C]
- $T_{v,i}$ = távhő visszatérő hőmérséklet [°C]
- $T_{talaj,i}$ = talajhőmérséklet [°C]

A számítás elvégzésével eredményekhez jutunk a vezeték hőveszteségével kapcsolatban, ám észrevehető, hogy a számítás alapját képező hőveszteség képlet nem veszi figyelembe a tömegáram értékét, ennek folytán a képlet nem tükrözi a valóságban tapasztalható működést. A modernebb távhőrendszerek mind frekvenciaváltós főköri szivattyúkkal vannak felszerelve, amely lehetővé – és elvárt üzemállapottá teszi – a változó tömegáramú hőszolgáltatást. Ez jelentős előnyt jelent mind energetikailag, mind gazdaságilag, azonban ezen üzemállapottal együtt jár az alacsonyabb hőigényű időszakokban – névlegesen a nyári évszakban – tapasztalható alacsony tömegáram. Ezen tömegáram csökkenés nagymértékben megnöveli a relatív hőveszteségek értékét, mivel a hőveszteségek csak kismértékben csökkennek le, ellenben a szállított és kiadott hőmennyiség sok esetben töredékére esik. Ezen jelenséget alátámasztja a **1. ábrában** foglalt adatsor, amelyek 2020-ban lettek meghatározva az egész éves mérések alapján.

Megnevezés	me.	jan.	feb.	márc.	ápr.	máj.	jún.	júl.	aug.	szep.	okt.	nov.	dec.	Összesen
Hálózatba betáplált hő	GJ	454 080	312 469	283 713	157 224	83 260	52 806	47 516	45 261	62 056	200 268	317 003	374 262	2 389 916
Hálózatból fogyasztónak kiadott hő	GJ	424 741	282 257	261 130	138 514	65 612	35 798	31 383	28 493	44 943	175 484	287 084	345 752	2 121 191
Hálózati veszteség	GJ	29 339	30 211	22 582	18 710	17 648	17 008	16 133	16 768	17 113	24 784	29 919	28 511	268 726
Hálózati veszteség betáplált hőre	%	6.46%	9.67%	7.96%	11.90%	21.20%	32.21%	33.95%	37.05%	27.58%	12.38%	9.44%	7.62%	11.24%

1. ábra: A 2020-as év hőveszteségi értékei a kelenföldi távhőrendszerre

Ezen információk ismeretében egyértelműen látszik, hogy jelen esetben a tömegáram értékének figyelmen kívül hagyása nagy mértékben rontja a hőveszteségi számítás pontosságát. Ennek kompenzálására kidolgoztunk egy arányossági módszert, amely segítségével figyelembe tudjuk venni a tömegáram változásokat is. A rendelkezésünkre álló éves adatokból meghatároztuk a havi összes tömegáram értéket, amelyet utána hozzárendelünk az adott körzet százalékos hőveszteségeihez ugyanazon évben. Ennek ábrázolása látható a **2. a) ábrán**. Az értékek egymáshoz való rendelése egyenes arányossággal nem

lenne megoldható, lévén, hogy nem-lineáris függőség tapasztalható köztük, továbbá szükséges kiszűrni az esetleges mérési hibákat az adatsorokban.

Ennek megoldására egy függvényt alkottunk az adatpárokból, amely alapján meghatároztuk a pontthalmazra legjobban illő, a tömegáram-hővesztesség függőség fizikai jellegét hűten visszaadó trendvonalat. Ezen trendvonal egyenletének meghatározását követően felhasználható az egyenlet az arányosított hővesztességi értékek meghatározására minden tömegáram adat esetén. Ennek számítási módját az alább látható egyenletben mutattuk be. Az eredményeket a **2. a) ábra** tartalmazza.

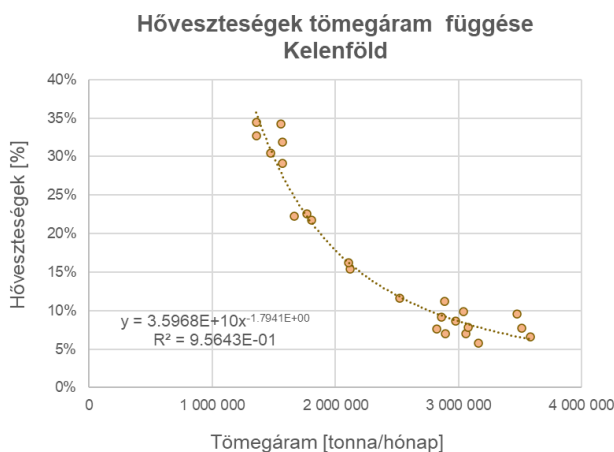
$$f_{x,i} = \dot{m}_i \cdot y$$

Ahol:

- $f_{x,i}$ = tömegáram-hővesztességi arányossági tényező [%]
- \dot{m}_i = a havi átlagos tömegáram érték [tonna/h]
- y = függvénykapcsolat egyenlete

Hónapok	Tömegáram [tonna/h]	Hővesztességek	t_i [-]	
2021	1	3 475 355	9,52%	6,62%
	2	3 062 037	6,98%	8,31%
	3	2 977 767	8,62%	8,73%
	4	2 860 531	9,17%	9,38%
	5	2 122 227	15,41%	16,03%
	6	1 567 809	29,10%	27,60%
	7	1 567 809	31,88%	27,60%
	8	1 555 204	34,27%	28,00%
	9	1 770 718	22,56%	22,19%
	10	2 886 349	11,22%	9,23%
	11	3 037 824	9,89%	8,42%
	12	3 516 233	7,70%	6,48%
2022	1	3 585 258	6,62%	6,26%
	2	3 078 817	7,85%	8,22%
	3	3 162 690	5,72%	7,84%
	4	2 892 580	7,00%	9,20%
	5	1 804 931	21,76%	21,44%
	6	1 475 042	30,44%	30,79%
	7	1 359 095	34,44%	35,67%
	8	1 358 294	32,66%	35,70%
	9	1 664 014	22,30%	24,80%
	10	2 108 426	16,24%	16,22%
	11	2 520 633	11,61%	11,78%
	12	2 822 878	7,65%	9,61%

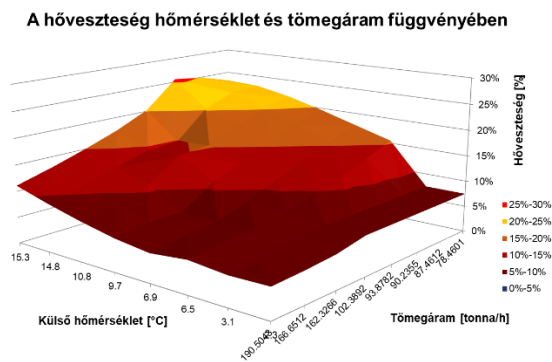
a)



b)

2. ábra: a) Tömegáramok és a százalékos hővesztességek kapcsolata és b) a hővesztességek tömegáram függése Kelenföld hőközrztben

A **2. b) ábrán** látható az illesztett trendvonal, illetve a hozzá tartozó egyenlet, amely hiperbola függvényjellegűt vesz fel. Az adatpontok trendvonalra való illeszkedésének mértékét az R^2 determinációs együttható adja meg, amely estünkben 0,956 értéket vett fel, ami elfogadható mértékű egyezést jelent. A távhővezetékek hővesztességének hőmérséklet- és tömegáram-függése a **3. ábrán** látható.

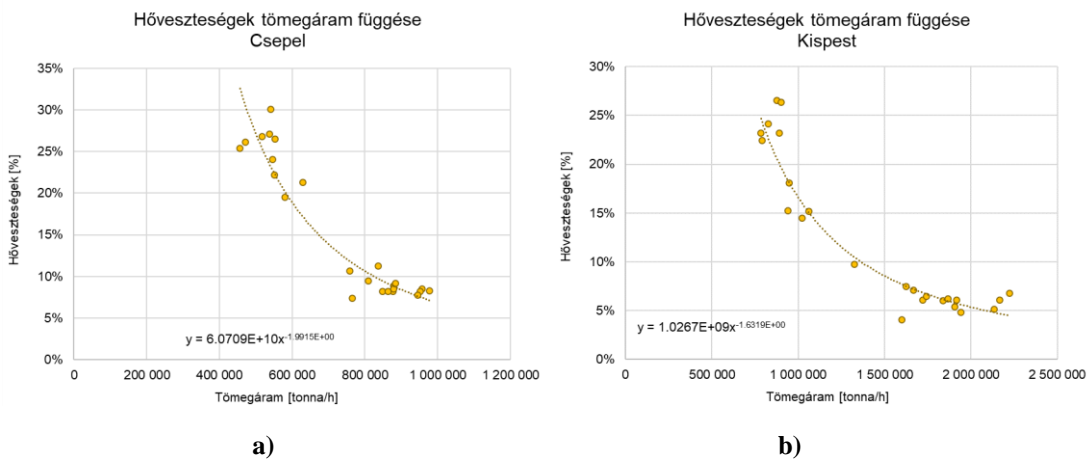


3. ábra: A hőveszteség külső hőmérséklet és tömegáramfüggése

A fent bemutatott táblázatban foglalt függvénykapcsolati értékek segítségével arányosítani lehet az adott órai hőveszteséget az adott hónapra jellemző kiszámított tömegáram-hőveszteség arányossági tényezővel, így egy közeli becslést kapva a valódi veszteségek mértékéről. Ezen számítások alapján a korábban felírt hőveszteségi képlet a következőképpen módosul:

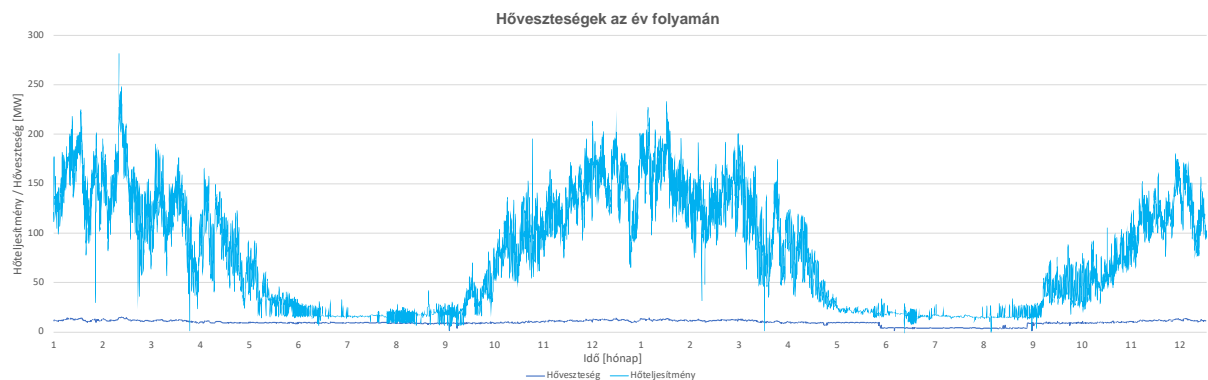
$$\dot{Q}'_{ker,i} = U \cdot l \cdot \left(\frac{T_{e,i} - T_{v,i}}{2} - T_{talaj,i} \right) \cdot (1 + f_x)$$

Ezen számításokat mindhárom vizsgált hőkörizet esetén elvégeztük, amelyeket a **4. a) ábra** és **4. b) ábra** mutatja be.



4. ábra: Hőveszteségek tömegáram függése Csepel **a)** és Kispest esetén **b)**

A hőveszteségek kiadott hőteljesítményfüggését jól ábrázolja a **5. ábra** és **6. ábra**, amelyeken jól megfigyelhető a nyári és téli időszakokban tapasztalható hőveszteség arány különbsége.



5. ábra: A hővesztések és a hőteljesítmény alakulása kétéves időtartamra Kelenföld hőközterben



6. ábra: A hőteljesítményhez viszonyított relatív hővesztések Kelenföld hőközterben

Az **6. ábrán** látható relatív hővesztés a hővesztések és a kiadott hőmennyiség hányadával meghatározott fizikai mennyiség. A téli üzemi állapotban 10% alatti értéket vesz fel, ám a nyári állapot esetén helyenként az 50%-ot is meghaladó értéket tapasztalunk. A kapott adatok számos esetben tartalmaztak hamis mérési adatokat, amelyek szűrési módszerét a későbbi fejezetekben részletezzük. Jelen esetben az adatsor elemzése során a 60% feletti veszteségi értékeket helyettesítettünk a következő 20 órás veszteségi adatok átlagértékével, mivel ezek valótlan értékeket mutattak volna. Ezen logikai rendszer matematikai megfogalmazása az alábbi képletben látható.

$$x_{h\acute{o}veszt,i} = HA \left(\frac{\dot{Q}'_{ker,i}}{\dot{Q}_{ker,i}} > 0,6; \sum_{i+1}^{i+20} \frac{\dot{Q}'_{ker,i}}{\dot{Q}_{ker,i}}; \frac{\dot{Q}'_{ker,i}}{\dot{Q}_{ker,i}} \right)$$

Ahol:

- $x_{h\acute{o}veszt,i}$ = Hővesztés aránya a kiadott hőmennyiséghez képest [-]
- $\dot{Q}_{ker,i}$ = A terület hőteljesítmény igénye [W]

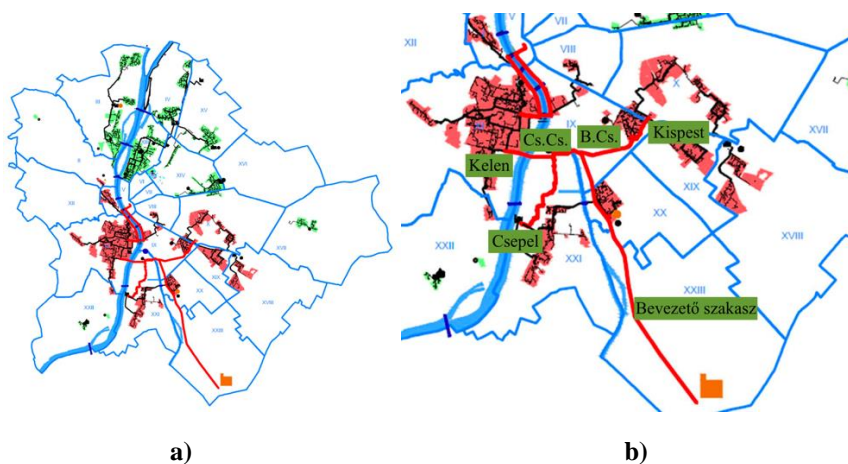
A fent látható egyenlet a pár sorral feljebb leírt logikai vizsgálat elvégzéséért felel, az első szekcióban a vizsgált feltétel van, a másodikban az igaz vizsgálat esetén alkalmazandó érték, a harmadikban pedig a hamis vizsgálat esetén alkalmazandó képlet.

3.2. Energetikai és gazdasági elemzés

3.2.1. TÁVHŐVEZETÉKEK LEFEKTETÉSE

Az egyes távhőközrzetekre való csatlakozás elemzése során a HUHA 2 hulladékégető mű létesítését vizsgáló, 2017-ben a Főtáv Zrt. által készített műszaki tanulmányra támaszkodtunk. [9.]

A távhőcsatlakozások nyomvonalvezetéséhez és ezek beruházási költségeinek meghatározásához a műszaki tanulmány adatait alkalmaztuk, a Dél-budapesti (Kelenföld, Csepel és Kispest) hőközrzetekhez vezető vezetékek nyomvonala a ábrán látható. Az **7.ábrán** pirossal vannak jelezve a számunkra releváns hőközrzetek.



7. ábra: a) Tervezett vezetékszakaszok az egyesített budapesti távhővezeték létrehozásához és b) az egyes vezetékszakaszokat határoló csomópontok

A dokumentációban meghatározott adatok, illetve a **7.ábrán** jelölt nyomvonalak segítségével meghatároztuk az egyes szakaszok hosszát.

- Bevezető szakasz: 11 500 m
- Bevezető csatlakozás (B.Cs) – Kispest : 4 000 m
- Bevezető csatlakozás (B.Cs) – Csepeli csatlakozás (Cs.Cs): 1 100 m
- Csepeli csatlakozás (Cs.Cs) – Kelenföld: 2 500 m

- Csepeli csatlakozás (Cs.Cs) – Csepel: 4 100 m

Ezen adatokon felül rendelkezésünkre állt információ a bevezető szakasz és a kispesti csatlakozás megépültének költségéről, amely 14 260 millió Ft volt 2017-ben. Ezen adatok segítségével meghatároztuk a vezeték fajlagos beruházási költségét, amely megadta, hogy egységnyi vezeték hossz lefektetése mennyibe kerül.

$$p_{vez} = \frac{C_{beruh,vez}}{l_{vez}}$$

Ahol:

- p_{vez} = vezeték fajlagos beruházási költsége [Ft/m]
- $C_{beruh,vez}$ = vezeték beruházási költsége [Ft]
- l_{vez} = vezeték hossz [m]

A fajlagos vezeték hossz értéke ezen számítás alapján $p_{vez} = 920\,000$ Ft/m lett, természetesen ez az érték 2017-es adatokkal volt számítva, így ennek kompenzálására a tanulmány készültének időpontjáig vett öt éves időtartam alatt felhalmozódott inflációs értékkel növeltük az értéket. Ez az inflációs érték 32,16%-ra adódott, amely felemelte a fajlagos értéket $p_{vez,inf} = 1\,215\,872$ Ft/m-re, amellyet a további számítások során alkalmaztunk. A beruházási költséget az adott szakasz hosszának függvényében határoztuk meg az alábbi számítási módszer szerint.

$$C_{beruh,vez} = l_{vez} \cdot p_{vez,inf}$$

3.2.2. HŐKÖRZETEK ENERGETIKAI ELEMZÉSE

A távhőcsatlakozások elemzése során megvizsgáltuk, hogy a három Dél-budapesti távhőközrtek esetén mely hőközrtekre, vagy azok milyen kombinációjára érdemes csatlakozni leginkább energetikai és gazdasági szempontból. Ennek meghatározásához felvettünk pár alapadatot, amelyek értékei fixek maradnak az összes közrtek számítása során. Ezen adatok a következők voltak:

- csúcskihasználási óraszám (τ_{cs}): 8100 órában lett meghatározva feltételezve, hogy a HUHA 2 az egész év folyamán maximális kapacitáson üzemel a 230 ezer tonna TSZH semlegesítésének érdekében. Az év során a kiadott hőmennyiség és villamosenergia-mennyiség mértéke változik a hőközrtek teljesítményigényének függvényében.

- tüzelési teljesítmény ($P_{\text{tüz}}$): 74,93 MW értékben lett meghatározva. Megadja a tüzelőanyag elégetéséből elméletileg kinyerhető maximális teljesítményt a belső hatások figyelembevétele nélkül.

$$P_{\text{tüz}} = \frac{m_{\text{TSZH}} \cdot F_{\text{TSZH}}}{3,6 \cdot \tau_{\text{CS}}}$$

Ahol:

- $P_{\text{tüz}}$ = tüzelési teljesítmény [MW]
- m_{TSZH} = az éves elégetett TSZH mennyiség [tonna]
- F_{TSZH} = TSZH fűtőértéke [MJ/kg]
- τ_{CS} = csúcskihasználási óraszám [h]

A körzetek vizsgálata során hat konstrukciót vizsgáltunk meg: Kelenföld, Csepel, illetve Kispest csatlakoztatása egyedüli hőközvetként, illetve a Kelenföld-Csepel, Csepel-Kispest és a Kelenföld-Csepel-Kispest körzetek kombinált csatlakoztatása. Az összes esetben meghatároztuk a rendelkezésünkre álló óras adatok segítségével a hőteljesítmény igény-, illetve hőveszteség lefutást az éves adatok átlagolásával és több körzet összekapcsolása esetén szummázásával.

$$\dot{Q}_{\text{ker},i} = \sum_1^n \left(\frac{\sum_1^m \dot{Q}_{\text{ker},i}}{2} \right)_n \quad \text{és} \quad \dot{Q}'_{\text{ker},i} = \sum_1^n \left(\frac{\sum_1^m \dot{Q}'_{\text{ker},i}}{2} \right)_n$$

Ahol:

- $\frac{\sum_1^m \dot{Q}_{\text{ker},i}}{2}$ = a hőteljesítmény igények adott órára vett átlagértéke [MW]
- $\frac{\sum_1^m \dot{Q}'_{\text{ker},i}}{2}$ = a hőveszteség igények adott órára vett átlagértéke [MW]
- m = az átlagolt évek száma [-]
- n = az összegzett körzetek száma [-]

Ezt követően a HUHA 2 hőteljesítmény kiadását rögzítettük az ellátott körzetek mindenkorai hőteljesítmény igényéhez, az elektromos termelést pedig a műszaki dokumentációban foglalt adatokra alapozva egy lineáris függvény alapján változtattuk a hőkiadás függvényében.

$$\dot{Q}_{\text{HUHA},i} = HA(\dot{Q}_{\text{ker},i} > \dot{Q}_{\text{HUHA},\text{max}}; \dot{Q}_{\text{HUHA},\text{max}}; \dot{Q}_{\text{ker},i})$$

$$P_{\text{HUHA}} = HA(\dot{Q}_{\text{HUHA}} = \dot{Q}_{\text{HUHA},\text{max}}; P_{\text{HUHA},\text{min}}; \dot{Q}_{\text{HUHA}} \cdot (-0,1972) + 25,2)$$

Ahol:

- $\dot{Q}_{HUHA,i}$ = a HUHA 2 által kiadott pillanatnyi hőteljesítmény [W]
- $\dot{Q}_{HUHA,max}$ = a HUHA 2 maximális hőteljesítménye [W]
- P_{HUHA} = a HUHA 2 által kiadott pillanatnyi villamos teljesítmény [W]
- $P_{HUHA,min}$ = a HUHA 2 minimális villamos teljesítménye [W]

A korábbi számítások eredményeképp már birtokában vagyunk az egyes hőkörtetek mindenkor órás hőveszteségi adatainak, ám a pontos értékeléshez szükségünk van az ellátott hőkörtet/hőkörteteket az erőművel összekötő vezetékszakasz veszteségeinek ismeretére is. Ennek érdekében az összekötő vezetéki hőveszteségeket meghatározásuk után összegeztük az alábbi módon.

$$\dot{Q}'_{vez,i} = U_{vez} \cdot l_{vez} \cdot \left(\frac{T_{e,i} - T_{v,i}}{2} - T_{talaj,i} \right) \cdot (1 + f_x)$$
$$\dot{Q}'_{össz,i} = \frac{\dot{Q}_{HUHA,i} - \dot{Q}'_{vez,i}}{\dot{Q}'_{ker,i}} \cdot \dot{Q}'_{ker,i} + \dot{Q}'_{vez,i}$$

Ahol:

- $\dot{Q}'_{vez,i}$ = összekötő vezeték hővesztesége [W]
- $\dot{Q}'_{össz,i}$ = összesített hőveszteség [W]

Ezen számítások alapján meg tudjuk adni a pontos hőveszteségét az adott csatlakoztatott rendszer függvényében, amely direkt módon fogja befolyásolni a bevételi értékeket a fogyasztókhöz valójában eljuttatott hőmennyiség tekintetében.

Az egyes hőkörtetekre való csatlakozás energetikai elemzésén kívül egy fontos tényezőt figyelembe kell venni, **A tanulmány felépítése** c. fejezetben megemlített R1 energy efficiency formulát. A formula fontossága abban jelentkezik, hogy a 0,65-ös értéket el nem érő hulladékhasznosító erőművek nem sorolhatók hulladékhasznosító kategóriába, hanem hulladékegető kategóriát kapnak. Ennek értelmében alig lesznek kedvezőbb helyzetben környezetkárosítási, illetve jogi szempontból, mint egy hulladéklerakó, avagy egy hulladékegetést végző személy.

A formula képlete alább látható.

$$R1 = \frac{E_p - (E_f + E_i)}{(0,97 \cdot (E_w + E_f))}$$

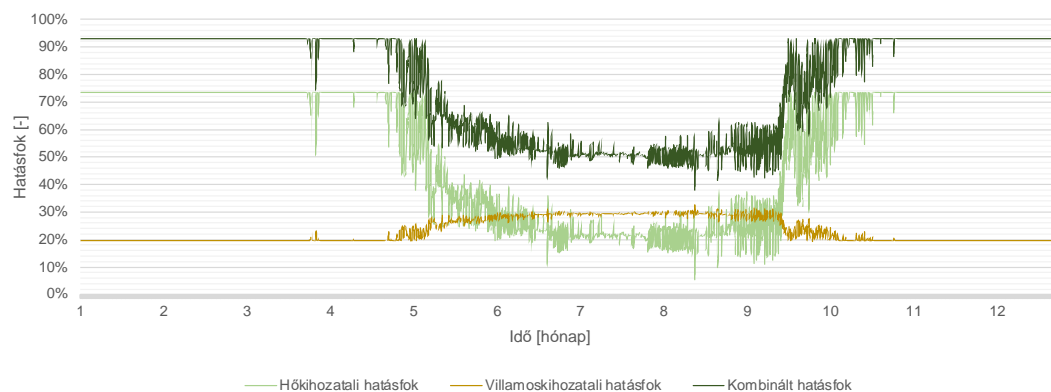
Ahol:

- E_p = hő vagy villamos energia formájában termelt éves energia [GJ]. A termelt villamosenergia értékét 2,6-tal, a hő értékét 1,1-gyel szorozzuk meg
- E_f = támasztó tűz és indító tüzelés energiaigénye [GJ]
- E_w = A tüzelőanyag energiataralma [GJ]
- E_i = az E_f és E_w értékeken kívül bevitt energiamennyiség [GJ]

Az összes hőközret esetén meghatároztuk az R1 formulát. Az értéken tartásának érdekében tanácsos növelni a kiadott hőmennyiséget, emiatt a kis hőteljesítményű közretetek esetén veszélyben lehet a határérték teljesítésének megvalósulása.

Következő lépésként meghatároztuk az erőmű hőkihozatali, illetve villamoskihozatali hatásfokát órás bontásban, továbbá ezek alapján az erőmű kombinált hatásfokát, amelyek alakulását az alább látható **8. ábrán** jelöltük. Megfigyelhető, hogy az év során a maximális hőteljesítmény igények fennállása mellett konstans értéken marad mindhárom hatásfok, ám az emelkedő külső hőmérsékletek mellett a hőkiadás hatásfoka drasztikusan csökkenni kezd, amelyet nem tud kompenzálni a lassabb ütemben növekvő villamoskihozatali hatásfok. Ennek eredményeképp a kombinált hatásfok is le fog csökkenni. Ez a csökkenés veszélyt jelent az R1 formula által támasztott követelmények szempontjából.

Hatásfokok alakulása az év során Kelenföldön

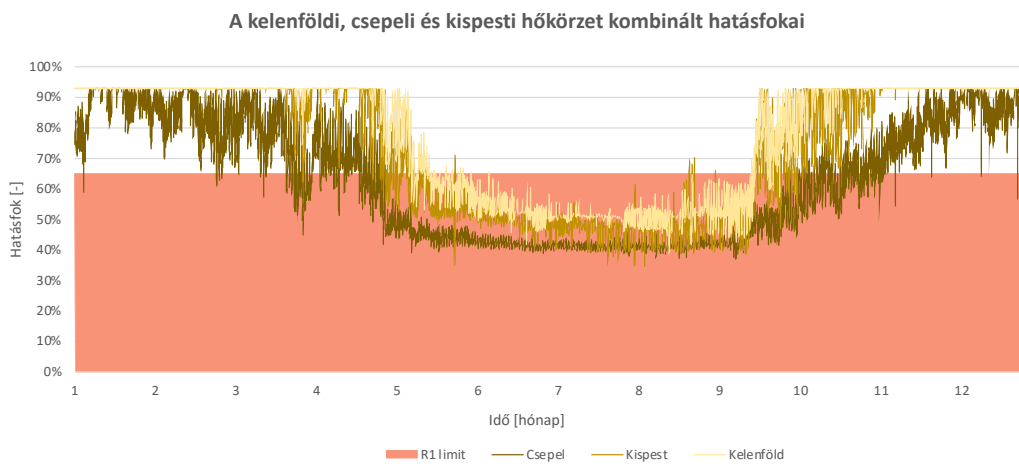


8. ábra: A hőkihozatali, villamoskihozatali-, illetve a kombinált hatásfok alakulása az átlagolt év során Kelenföldön

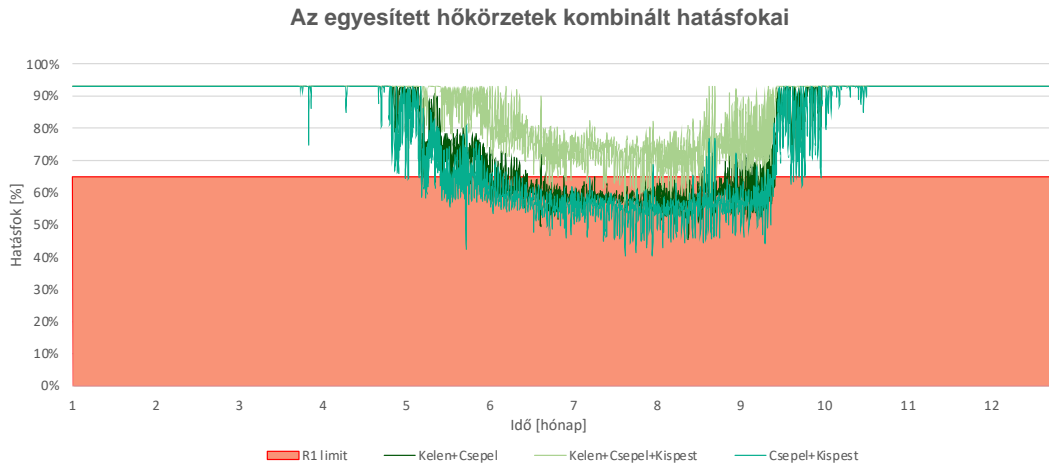
A következő, **9. ábrán**, illetve a **10. ábrán** lévő két diagram tartalmazza az egyes hőközretetekre való csatlakozás esetén kialakuló átlagolt hatásfok értékeket, illetve halvány piros színnel jelezve van az R1 formula által felállított határérték alatt található zóna, amely elkerülendő. Látható, hogy a világos színnel jelölt kelenföldi közret érintette a

legkisebb mértékben a „tiltott zónát”, mivel ezen körzet rendelkezik a legnagyobb hőigénnyel.

Érdeemes megjegyezni, hogy az R1 formula értéke magasabb értéket fog felvenni, mint az év során tapasztalt valós átlagos kombinált hatásfok, de a jelleg szemléltetésére tökéletesen alkalmasak az alábbi diagramok.



9. ábra: A kelenföldi, csepeli és kispesti hőközvetek kombinált hatásfokainak alakulása az átlagolt év során



10. ábra: A kelenföldi és csepeli, a csepeli és kispesti, illetve mindhárom hőközvet összekapcsolása során tapasztalt kombinált hatásfokának alakulása

A **10. ábra** jól szemlélteti, hogy a nagyobb hőigénnyel rendelkező kapcsolási struktúra kedvezőbb hatásfok értéket fog eredményezni az egész év során, különös tekintettel a nyári időszakra. A világoszöld színnel jelölt kelenföldi, csepeli és kispesti összekapcsolt csatlakozási állapot gyakorlatilag az év teljes szakaszában elkerüli az R1 szempontjából „tiltott zónát”.

3.2.3. GAZDASÁGI ELEMZÉS

Az energetikai eredmények fényében meghatároztuk az egyes zónák csatlakoztatásával realizálható órás, illetve ezek alapján az egész éves összegzett bevételt. Ennek kiszámításához figyelembe vettük az egyes, jelenleg üzemelő erőművek értékesítési árának, illetve az alternatívaként kínált HUHA 2 hőértékesítési árának különbségét, amelyet aztán az eladott hőmennyiség mértékében összegeztünk. A díjazási értékeket havi rendszerességgel a Magyar Közlönyben közölt adatok adják meg.

4. táblázat: Az egyes körzeteket ellátó erőművek és kazánok értékesítési díjai a 2023. szeptember 29-ei Magyar Közlöny adatai alapján [10.]

	Kelenföldi erőmű	Csepel erőmű	Csepeli fűtőmű	Kispesti erőmű
Értékesítési díj [Ft/GJ]	12871	4807	12703	12580
HUHA 2 értékesítési díj [Ft/GJ]	2000			

A számítás alapján meg tudjuk határozni, hogy a szolgáltató számára mely körzet csatlakoztatásával érhető el a legmagasabb bevétel és ezáltal a leggyorsabb megtérülési érték. Ennek fényében nem az erőmű szemszögéből végezzük a számítást, így a villamos energia értékesítéséből származó bevételt nem számoltuk bele a megtérülési adatokba.

$$p_{bevétel,i} = (p_{gáz} - p_{HUHA}) \cdot (\dot{Q}_{HUHA,i} - \dot{Q}'_{össz,i})$$

Ahol:

- $C_{bevétel}$ = a mindenkori órás bevétel értéke [Ft]
- $p_{gáz}$ = a csatlakoztatott hőközvetet alapból ellátó erő-, avagy fűtőmű hőértékesítési díja [Ft/GJ]
- p_{HUHA} = a HUHA 2 hőértékesítési díja [Ft/GJ]

A csepeli hőközvetet esetében egyidejűleg üzemel mind erőmű, mind pedig fűtőmű. Ennek okán a számítás során a két egység árának átlagolt értékesítési árával számoltunk.

Az órás bevételi adatok összesítésével meghatároztuk az átlagos évre vetített bevételi költségeket, amelyek segítségével kiszámítottuk az egyes beruházások megtérülési idejét. A számítások során 12%-os inflációs rátával számoltunk, mivel a bevételek alapján rövid megtérülési időre lehet számítani, így tanácsos a jelenlegi infláció értékéhez közelebb eső rátát alkalmazni.

$$NPV = \sum_n \frac{\sum_i p_{bevétel,i}}{(1+r)^n} - C_{beruházás}$$

Ahol:

- NPV = nettó jelenérték [Ft]
- $\sum_i p_{bevétel,i}$ = éves aggregált bevétel [Ft]
- r = infláció [-]
- $C_{beruházás}$ = beruházási költség [Ft]

A számítás eredményeit az **5. táblázat** tartalmazza.

5. táblázat: A távhőcsatlakozások gazdasági értékelése

Távhőcsatlakozások eredményei összegezve						
	Kelenföld	Csepel	Kispest	Kelenföld+Csepel	Csepel+Kispest	Kelenföld+Csepel+Kispest
Összekötő vezetékszakas [m]	14800	16700	15500	17800	15000	22900
Beépítési költség [mFt]	17995	20305	18846	21643	18238	27843
Gáz értékesítési díj [mFt/MWh]	0.046	0.032	0.045	0.039	0.038	0.041
HUHA2 értékesítési díj [mFt/MWh]	0.007	0.007	0.007	0.007	0.007	0.007
Éves bevétel [mFt]	17790	11893	17268	18461	18185	19271
R1 érték	1.222	1.115	1.208	1.252	1.236	1.295
Megtérülési idő [hónap]	21	25	15	17	14	21

Látható, hogy az összes beruházás kiemelkedően gyors megtérülési mutatókkal rendelkezik, ezen felül a Csepel + Kispest, illetve a Kispest konstrukció volt a legkedvezőbb gazdasági szempontból, amelyek 14, illetve 15 hónapos dinamikus megtérüléssel rendelkeznek. A korábban említett R1 értéknek minden hőkörizet megfelelt, így elegendő csupán a gazdasági megfontolások alapján döntést hozni a megvalósításról.

A gyors megtérülési idő magyarázata a gázalapú jelenlegi erőművek és fűtőművek által alkalmazott értékesítési árak, illetve a HUHA 2 által nyújtott alacsony, hulladék alapú hőenergia közötti magas különbség. Természetesen a gázárak csökkenésével növekedhet a beruházások megtérülése, ám a jelenlegi Európai Unió környezettudatossági trendek és a fosszilis energiahordozók évtizedek óta tartó emelkedő ár tendenciája mellett ennek alacsony esélye van.

Annak tudatában, hogy a HUHA 2 távhőcsatlakozása a Dél-budapesti hőkörizetek mindegyikére példátlanul rövid megtérülési idővel rendelkezik - főleg egy több tízmilliárdos beruházás esetén -, figyelmünket hőkörizetek igényei felé fordítottuk. Az elvégzett energetikai és gazdasági elemzés értékeinek megtartásához elengedhetetlen a távhőszolgáltató számára az egyes hőkörizetekben jelentkező igények pontos ismerete, lehetőleg több órás előrejelzési időtartamban az erőművek fel-és leterhelésének időben történő elvégzéséhez. Ez az előrejelzési idő még fontosabb szerepet kap a HUHA 2 csatlakoztatása esetén, mivel a több, mint 13 km-es összekötő vezetékszakasok drasztikusan megnövelik

a szabályozási jel holtidejét. Ha az igény jelentkezésének pillanatában változtatná meg a HUHA 2 a teljesítménykiadását, akkor a nagy távolság miatt kezelhetetlen lenne a hőkiadás szabályozása.

Ezen problémát némileg kompenzálja a távhőrendszerekben tárolt hőkapacitásból eredő tehetetlenség, ám ekkora táv áthidalása esetén már ez nem lenne megoldás. Ennek érdekében a megfelelő csatlakoztatási lehetőségek biztosításáért a következő fejezetekben a hőkörzetek hőteljesítmény igényének előrejelzésével fogunk foglalkozni, amelyet gépi tanulási módszerekkel közelítettünk meg. A gépi tanulás kifejezetten alkalmas a nagy adatmennyiségekkel dolgozó problémák megoldására, kifejezetten az időjárás-, esztünkben pedig a hőteljesítmény predikcióra.

4. HŐTELJESÍTMÉNY-IGÉNY ELŐREJELZÉS

A tanulmány célja c. fejezetében leírtak szerint a további fejezetek során a kelenföldi-, csepeli és kistérségi távhőköri körzetben jelentkező hőteljesítmény-igények előrejelzésével foglalkoztunk. Az előrejelzés pontosításával csökkenthető az erőművek teljesítményigénye és növelhető a hatásfoka, mivel nem szükséges magasabb teljesítménnyel üzemelni a fogyasztói igények biztos ellátásának érdekében, továbbá a hirtelen jelentkező hőigények pontos előrejelzése esetén nem szükséges nagy tüzelési teljesítmény alkalmazása a gyors felfűtéshez. Emellett a pontos teljesítmény predikciók segítségével az erőművi, illetve a távhőszolgáltatási szabályozás pontosságát is növeljük, amely lehetővé teszi az alacsonyabb visszatérő hőmérséklet elérését a túltermelés elkerülésével.

Ezek alapján a hőteljesítmény előrejelzés pontosságának növelésével direkt módon csökkentjük mind a termelők, mind pedig a távhőszolgáltatók költségeit, miközben párhuzamosan növeljük a távhőszolgáltatás ellátásbiztonságát és a jelenlegi, illetve a jövőben csatlakozó fogyasztók komfortérzetét.

Az előrejelzést **A tanulmány felépítése** c. fejezetben részletezett okok alapján gépi tanulási módszerek alkalmazásával végeztük el a jelenlegi empirikus hőteljesítmény előrejelzési módszer pontosságának növelésének reményében.

4.1. Gépi tanulás bevezetése

A gépi tanulás alapja, hogy a kezdeti bemeneti adatok és az eredmények (=kimeneti adatok) segítségével az algoritmus megtanulja a kapott információban található mintázatokat és kapcsolatokat, amely segítségével a jövőbeli eseteknél hasonló állapot fennállásakor már rendelkezik előzetes ismerettel az eset pontos kiértékeléséhez. A tanulási fázis sikerességének ellenőrzésére nem az egész adatsort bocsájtjuk az algoritmus rendelkezésére, hanem egy részét – általában 25%-át – fenntartjuk tesztelésre, avagy másnéven kiértékelésre. A tesztelés során ugyanúgy megkapja az algoritmus az adatokat, ám ezúttal nem adjuk meg számára a kimeneti adatokat, így azokat saját magának kell meghatároznia csupán a bemenő adatok alapján, csakúgy, mint ahogy a való életben fog működni. Ezek után összevetjük az általa létrehozott kimeneti adatokat a valódi kimeneti adatokkal, amelyek alapján meg tudjuk határozni az algoritmus predikciós pontosságát.

Számtalan gépi tanulást alkalmazó algoritmus létezik, ezek közül az általunk alkalmazottakat a későbbiekben mutatjuk be.

4.1.1. GÉPI TANULÁS ALKALMAZÁSI TERÜLETEI

A különböző statisztikai problémák két nagy csoportba sorolhatók, klasszifikációs és regressziós problémába, amelyek közötti különbségek megértése fontos a gépi tanulási metódusok működésének értelmezéséhez.

A klasszifikáció, avagy másnéven osztályozás során a vizsgált adatokban csoportosítási lehetőségeket keresünk, majd ezek alapján próbáljuk meg az új megfigyeléseket elhelyezni valamely csoportban. Erre példa lehet egy ember vércsoportjának meghatározása. A kategóriák/csoportok között az esetek nagy részében egyértelmű megkülönböztetés van.

A regresszió ezzel ellenben olyan folyamatos adatsorok elemzése során alkalmazandó, amelyek nem vesznek fel pontosan azonosítható csoport-értékeket, hanem egy skálán helyezkednek el. Az elemzése során a függő változó – másnéven eredmény, avagy kimenő adat – és egy vagy több független változó között keresünk valamilyen függvénykapcsolatot a folytonos értékek meghatározásához.

4.1.2. FORWARD VS. ADAT ALAPÚ ELŐREJELZÉS

Az előrejelzési módszerek két kategóriába sorolhatók: forward és adatalapú módszerek. Az épületek, épületkomplexumok stb. fizikai tulajdonságainak minél pontosabb modellezésére törekvő szimulációs rendszereket jobb magyar szó híján „forward” módszernek hívjuk.

A forward módszerek komplexitása nagy mértékben változhat a szimulálni kívánt rendszerek bonyolultságától függően egy pontos előrejelzésre képes modell felépítése pedig időigényes, bonyolult egyenletrendszereket alkalmazó számításokat kíván és magas hozzáértési szintet. Az ilyen rendszerek közé tartozik többek között a TRNSYS, illetve az EnergyPlus szoftver, amelyek alkalmazásához hosszú tanulási időre van szükség. Ezen problémákon felül a szoftverek megfelelő működéséhez magas precizitással meghatározott adatok szükségesek a szimulálni kívánt rendszer minden részletéről, amelyek

gyakran nem elérhetőek. A nagymennyiségű befektetett munka ellenére a tapasztalatok azt mutatják, hogy a forward alapú módszerek sokszor alulmaradnak a gépi tanulás alapú rendszerekhez képest.

A második módszer a tanulmányban is alkalmazott gépi tanulási módszerek alapját képező adat alapú előrejelzés, amely a vizsgált rendszer historikus adatainak vizsgálatával kívánja meghatározni a jövőbeli igényeket. Ezen módszer hátránya, hogy hatalmas mennyiségű hibátlan adatra van szükség a rendszer megfelelő szintű betanításához. A valóságban hibátlan adatok gyűjtése szinte lehetetlen, így a bennük található hibákat és hiányosságokat kezelni szükséges az elvárt pontosságú eredmények érdekében. Ezzel szemben előnye, hogy ezen műveletek elvégzése után a rendszer a korábbi, historikus valós adatok alapján dolgozik, amelyek tartalmazzák már az externális (hőmérséklet, időjárás stb.) és az internális (tömegáram, előremenő hőmérséklet stb.) paraméterek, illetve a fogyasztói szokások hatását. Az algoritmus feladata az, hogy a nagymennyiségű adathalmazok között megtalálja a megfelelő függvénykapcsolatokat és súlyozásokat, amely alapján a jövőben előforduló hasonló körülmények esetén meg tudja állapítani a várható igényeket nagy pontossággal.

Ezen információk értelmében a tanulmány további részében gépi tanulási modellekkel fogunk foglalkozni.

4.2. Alkalmazott algoritmusok

A szimulációk során számos algoritmust alkalmaztunk, amelyek közül a legjobb eredményt produkáló metódusokat tartottuk meg és mutatjuk be a következő alfejezetekben. Az algoritmusok két fő csoportra oszthatóak: időbeli sorrendet figyelembe vevő, illetve időbeli sorrendet figyelembe nem vevő metódusok.

Az időbeliséget figyelembe nem vevő algoritmusok kizárólag az éppen aktív adatsorban található partaméter értékeket alkalmazzák a predikciók létrehozásához, tehát nincs ismeretük a már korábban lejátszódott folyamatokról.

Ezzel szemben az időbeliséget figyelembe vevő algoritmusok rendelkeznek beprogramozott memóriával, amely segítségével a beállításuk függvényében tetszőleges adatsor távolságra „emlékeznek”, ezáltal a jelenlegi predikciók létrehozásához fel tudják használni a korábbi adatokat is, illetve azok gradiens értékét is, másnéven a változás mértékét. Az ilyen algoritmusok kifejezetten alkalmasak az időbeli mintázatok és trendek

felismerésére, mint például a fogyasztók napi rutinja, amely előnyt biztosít számukra a teljesítmény-igény előrejelzés során. A tanulmány során hat időbeliséget figyelembe nem vevő metódust vizsgáltunk, amelyek között található statisztikai alapú algoritmus és neurális háló is. A vizsgált algoritmusok individuális eredményeinek javításának érdekében egy „ensemble”, azaz együttes értékelési módszert alkalmaztunk, amely során az összes gépi tanulási metódus együttes prediktált értékét vesszük figyelembe.

A tanulmány során kitekintésként a kelenföldi hőközretre alkalmaztunk egy hetedik, időbeliséget figyelembe vevő neurális hálót is, amely segítségével megvizsgáltuk az időfüggő metódusok potenciális előrejelzés pontosság növelő képességét.

4.2.1. IDŐBELISÉGET FIGYELEMBE NEM VEVŐ ALGORITMUSOK

4.2.1.1. *K nearest neighbour*

A *K nearest neighbours* (KNN) algoritmus egy nem-parametrikus gépi tanulás alapú metódus, amely a nevéből adódó jelentés értelmében az adott adatpont értékét az általa azonosított adatszoportokhoz történő hozzárendelés alapján dönti el. A nem-parametrikus kifejezés arra utal, hogy az algoritmus nem támaszkodik előre ismert eloszlási adatokra az eredmény meghatározásának érdekében, ami előnyös előre nem ismert adathalmazok értékelése során. Alkalmazható igaz/hamis (=boolean) jellegű értékmeghatározásra és regresszióra is, amely a mi esetünket fedi.

Az algoritmus intuitív módon, az azonos szomszédságban lévő adatpontok csoportosításával, majd ezen csoportokhoz való közelség alapján közelíti a független paraméterek és a folytonos kimenetel közötti kapcsolatot. A szomszédság méretét az elemzőnek kell meghatároznia.

Az algoritmus problémája az, hogy a szimuláció számítási igénye erősen függ a független (=bemeneti) paraméterek számától, így sok változó alkalmazása esetén megnő a futási idő.

Jelen tanulmány során a KNN-t csak az **AdaBoost** algoritmus keretein belül alkalmazzuk.

4.2.1.2. Support vector regressor

A Support Vector Regressor (SVR) a Support Vector Machine (SVM) algoritmus regresszióhoz alkalmazható verziója. Működési elve az, hogy a rendelkezésére bocsátott adatsort egy, avagy több lineáris hipersíkkal kívánja elválasztani úgy, hogy az egy csoportosulásban található adatok mind ugyanazon csoportba tartozzanak. A hipersík a szimulált tér dimenziószámától függ, névlegesen mindig egy dimenzióval a tér alatti elem lesz. Ez gyakorlatban pontot, vonalat, síkot, alteret stb. jelent a növekvő dimenziók esetén.

Az algoritmus polinomiális, avagy radiális kernelt alkalmaz, a „kernel” az algoritmus matematikai operátor egysége (=magja), amely a modell számításainak elvégzéséért felel. A polinomiális kernel az adatok csoportosítása érdekében magasabb dimenziókba konvertálja az adatokat egészen addig, ameddig fent említett csoportosítás nem valósul egy adott ε hibahatáron belül. A radiális kernel „végtelen dimenzióban” minden adatpont közvetlen környezetében található, már azonosított csoportú pontok alapján határozza meg a predikciós értéket. Az algoritmus működését a következő egyenletekkel lehet leírni:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i + \xi_i^*$$

Ahol a következő követelményeknek kell teljesülni:

$$y_i - w^T \phi(\bar{x}_i) - b \leq \varepsilon + \xi_i$$

$$w^T \phi(\bar{x}_i) + b - b \leq \varepsilon + \xi_i^*$$

Ahol ε az összes pontra vonatkozó kívánt hibatarományt jelöli. A ξ_i és ξ_i^* slack, azaz „laza” változók, amelyek garantálják, hogy minden ε esetén létezik megoldás. C egy büntető kifejezés, amelyet az adatok illesztése és simasága közötti egyensúly megteremtésére használunk. ϕ egy kernelfüggvényt jelent, amely a bemeneti teret egy magasabb dimenziós jellemzőtérre képezi le.

Alapvető a probléma az SVR algoritmusnál, hogy a magasdimenziós terek leképezéséhez nagyfokú polinomokat használ, amelyek hatására a futási számítás igény több, mint kvadratikusan (n^{4+}) kapcsolatban van a bemeneti adatok mennyiségével, ami miatt több 10 000 kezdeti adat alkalmazása esetén drasztikusan megnő a futási idő.

4.2.1.3. *Random forest regressor*

A Random forest regressor algoritmus a fák jellegét utánzó Decision tree learning, azaz „döntési fa” alapú gépi tanulási metódus. A decision tree a fa ágainak elágazását utánozva határozza meg a vizsgált értékek eredményét, hol minden elágazás során egy „if” feltételt igazságértékét vizsgáljuk, majd az eredmény függvényében haladunk tovább a fa ágain. A fák komplexitását növelni lehet a mélységükkel, ami korrelál a fánként elvégzett „if” feltételek számával, tehát egy szofisztikáltabb döntéssel. A Random forest regressor nagymennyiségű fát alkalmaz egyszerre, amelyek által kiadott eredményeket szavazásos alapon kiértékeli egy meggyőző eredménnyé.

4.2.1.4. *AdaBoost*

Az AdaBoost, teljes nevén Adaptive Boosting egy algoritmus, amely több másik algoritmussal szimbiózisban növelni tudja azon algoritmusok pontosságát, innen ered a „boosting” kifejezés. Az algoritmus a **Random forest regressor** algoritmus fejezetében ismertetett módhoz hasonlóan több algoritmus darabot alkalmaz. Az általa alkalmazott algoritmus darabok egyszerűségük miatt „weak learner”-nek, avagy gyenge tanulóknak vannak elnevezve.

A metódus alapja, hogy nagyon egyszerű elemeket alkalmaz, a Random forest algoritmus alkalmazása esetén a fák helyett csak „tönköket” használ, olyan fákat, amelyek csak egy döntésből állnak. A gyenge tanulók tanulnak egymás eredményeiből a paraméterek súlyozása lévén, amely folyamatosan változik a futás során. A kiértékelésnél itt is szavazás alapján történik, ám jelen esetben minden gyenge tanuló rendelkezik egy súlyozási értékkel, ami alapján a helyes előrejelzést felmutató tanulók nagyobb beleszólást kapnak a végeredménybe, mint a rossz eredményt adók.

4.2.1.5. *Multilayer perceptron*

A Multilayer, avagy többrétegű perceptron (MLP) egy modern feedforward neurális hálózat (FFNN), amely teljesen összekapcsolt neuronokból áll, nemlineáris aktiválási függvényrel, legalább három rétegbe szervezve, és arról nevezetes, hogy képes megkülönböztetni a nem lineárisan szeparálható adatokat. A neurális hálók az emberi agy működését utánozni kívánó gépi tanulási metódusok, amelyek neuron rétegekből állnak. A rétegek, illetve a neuronok számának növelésével növelni lehet a modell komplexitását.

Jelen háló minden rétegén található neuronok össze vannak kötve a szomszédos réteg neuronjaival és ezen összeköttetések súlyozási értékkel rendelkeznek, amely meghatározza a bemenet és kimenet közötti kapcsolatot. A háló a tanulási fázis során a súlyozások változtatásával tanul „backpropagation”, avagy másnéven visszaterjedés lévén, amely során a háló által prediktált, és a valós érték közti tévedés mértékétől függően változtatja visszamenőlegesen a szükséges neuronok súlyozását.

A neurális hálók működési elvébe jelen tanulmány során nem fogunk mélyebben belemerülni.

4.2.1.6. Voting regressor

A Voting regressor algoritmus matematikai koncepciója meglehetősen egyszerű. Az algoritmus alapja, hogy több gépi tanulás alapú algoritmust foglal magába, amelyek minden pontra vett eredményét összeveti, majd az általunk beállított kezdeti paraméterek szerint súlyozza azokat. Ez alapján előállít egy együttes, úgynevezett „ensemble” előrejelzést az összes algoritmus adatának segítségével, amely így statisztikai alapon kevésbé fogja tartalmazni azt egyes metódusok belső felépítéséből következő elkerülhetetlen jellemző hibákat.

$$P_{voting} = \sum_{i=1}^n (wt_i \cdot P_i)$$

Ahol P_{voting} jelzi a végső eredményt, wt_i jelzi a megadott súlyozási értékeket, illetve P_i az egyes algoritmusok adott pontra kiadott eredményét.

4.2.2. IDŐBELISÉGET FIGYELEMBE VEVŐ ALGORITMUSOK

Számos olyan algoritmus létezik, amely figyelembe veszi az adatok időbeli lefutását, ám mi a Long short-term memory nevű hálóstruktúrát alkalmaztuk.

4.2.2.1. Long short-term memory

A Long short-term memory, vagy másnéven LSTM egy visszatérő neurális háló (recursive neural network, azaz RNN), amely az időben korábban bekövetkezett adatok, vagy a saját maga által alkotott predikciókat a memóriájában eltárolva fel tudja ismerni

az egymást követő adatok közötti mintázatokat és trendeket és ezeket beépíteni az előrejelzésbe.

Az LSTM megoldást kínálja az RNN-ek esetében domináns eltűnő/elszálló gradiens problémára. A probléma alapja az, hogy a visszatérő háló felhasználja az előző prediktált értékeket a következő érték előrejelzésekor, ám a műveletek során nem változtatja meg a súlyozási értékeit annak érdekében, hogy nagymennyiségű adatot tudjon feldolgozni. A feldolgozott adatok mindegyike esetén a súlyozás szorzásra kerül, így $i_n = i \cdot w^n$ képlet alapján (ahol i a bemeneti paraméter, w pedig az előző súlyérték) hatványozódik a következő bemeneti értéke. Ennek hatására kismennyiségű adatsor esetén is lehetetlenné válik az algoritmus tanítása, mert a gradiens értéke drasztikusan megnő ($w > 1$ esetében), avagy szinte megszűnik ($0 < w < 1$) esetében. A gradiens direkt befolyással bír a lépéstávolságra, mivel a meredekebb szakaszokon a kis „ x ” tengely lépésköz is nagymértékű távolság ugrást eredményez a függvényen, ezáltal megnehezítve a globális minimum érték azonosítását és elérését. Ezt a problémát az LSTM egy hosszútávú és egy rövidtávú memória elkülönítésével oldja meg, amely ezáltal nem fog hatványozottan változni és alkalmassá teszi a neurális hálót nagymennyiségű adat feldolgozására.

4.3. Adatok elemzése

4.3.1. JELENTKEZŐ IGÉNYEK

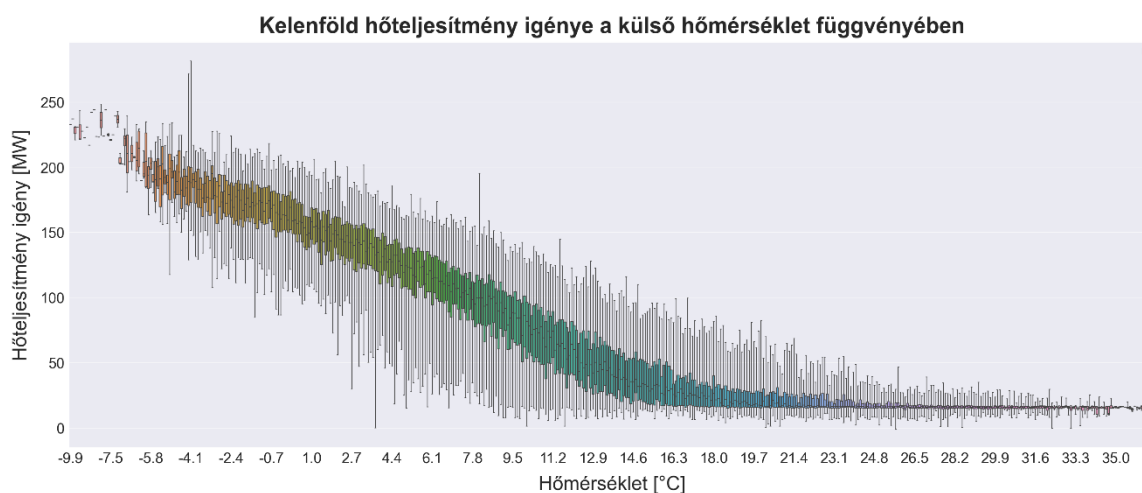
A jelentkező hűteljesítmény-igények előrejelzése bonyolult feladat, sok esetben nehezen megvalósítható pár órás viszonylatban egyedülálló épületek esetén is. Esetünkben három különálló – kerület nagyságú területeken elhelyezkedő – távhőközvet pontossági hőigény előrejelzését kíséreltük meg a rendelkezésre álló externális és internális adatok alapján 6 óra, 24 óra, vagy akár meghatározatlan hosszúságú esetén.

A **Gépi tanulás bevezetése** fejezetben említésre került a paraméterek komplexitása, ezt kívánjuk egy kicsit részletezni. Externális befolyásoló tényezőknek nevezzük azon befolyásoló hatásokat, amelyek egyáltalán nincsenek az irányításunk alatt. Ezen értékek többek között például a levegő relatív nedvességtartalma, a szél erőssége és bizonyos esetekben az iránya, az időjárás – a jelenlegi tanulmányban időjárásköddal jelezve -, illetve a felhasználók napi rutinjából keletkező fluktuációk. Ezen lengésekre példa a reggeli

kisebb és az esti nagy hőigénycsúcs, a hétköznap és hétvége közötti teljesítmény különbség, illetve az évszak.

Internális paramétereknek nevezzük az általunk befolyásolható értékeket, amelyek szintúgy hatással vannak a jelenlegi üzemállapokra. Ezek lehetnek többek között a korábban szolgáltatott hőmennyiség és tömegáram értékek, illetve a távhővezetékben található előremenő és visszatérő közeghőmérséklet. A korábbi internális adatok hatást gyakorolnak a jelenlegi üzemre is, amelyet bemutatót Davy G. et al. [11.]

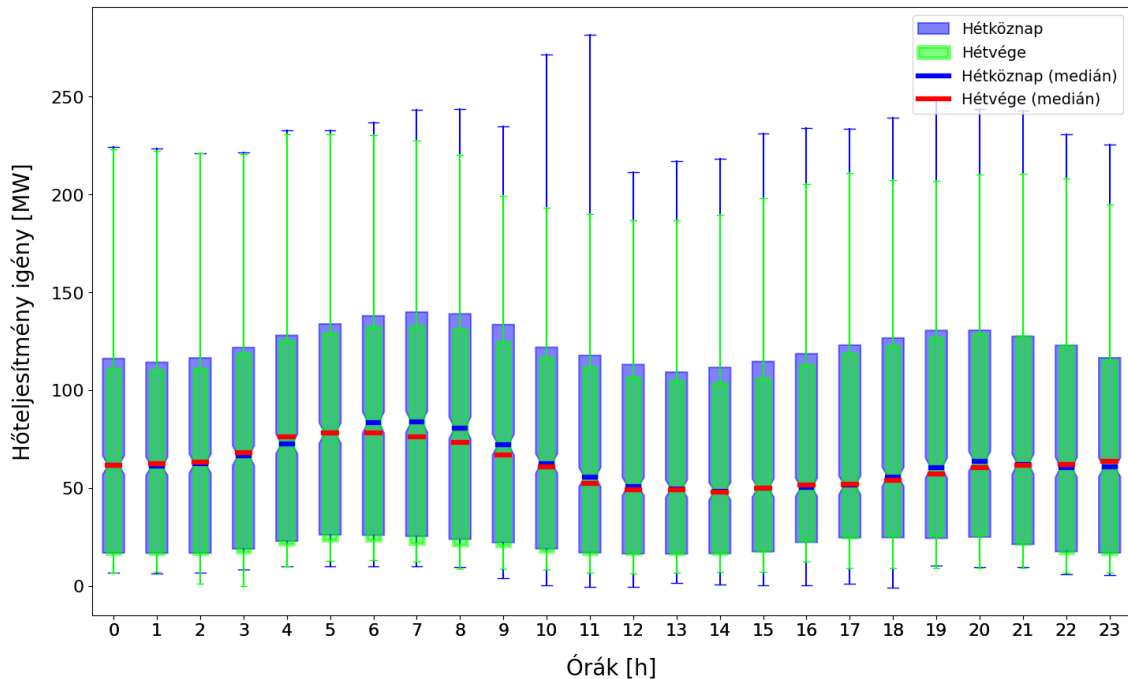
Ezek figyelembevételével a hőteljesítmény-igény egyik legerősebben a külső hőmérséklettől függ, amely egy kvázi-lineáris jelleget kölcsönöz az igényeknek, amelyet a **11. ábrán** láthatunk. Ezen felül erős függés tapasztalható még a napon belüli óra, illetve a hét napjának függvényében, amelyet a **12. ábrán** mutattunk be.



11. ábra: A hőteljesítmény-igény a külső hőmérséklet alakulásának függvényében a kelenföldi távhőköri zóna esetén

A **11. ábrán** megfigyelhető, hogy kb. 15 °C-os külső hőmérséklet felett az igények egy közel konstans értéket vesznek fel, ennek oka a fűtési és a használati melegvíz (HMV) igények közötti különbségeket jelzi. A hidegebb időszakok alatt szükséges fűtési teljesítmény megszűnésével csak az egész év során folyamatosan fennálló HMV igények maradnak meg, amelyek fennállnak az egész nyári időszak alatt is. A szürke színnel jelölt vonalak az adott hőmérsékletre tartozó szórás értéket határozzák meg.

Hőteljesítmény igény a hétköznap és hétvégén



12. ábra: Hőteljesítmény-igények órás lefutása hétköznap, illetve hétvégén box-plot diagramban ábrázolva

A **12. ábrán** bemutatott box-plot diagramon kék színnel jelöltük a hétköznapok, és zölddel a hétvégék során mért átlagolt hőteljesítmény-igényeket, amelyek mediánértékét a dobozokon található színes vonalakkal jelöltük: kék vonallal a hétköznapok és piros vonallal a hétvégék medián értékét. Az ábrán szintén feltüntettük az adott órához tartozó maximum és minimum értékeket is, amelyek a dobozokhoz tartozó vékony vonalakkal vannak jelölve.

Megfigyelhető, hogy a hétköznapok mind medián, mint pedig maximum pont szempontjából magasabb értéken helyezkednek el, mint a hétvégi esetek. A diagram jól ábrázolja a napi teljesítményingadozásokat, illetve a hétvégén tapasztalható alacsonyabb hőigényeket a hétköznapok esetekhez képest. Ennek oka az irodaházak és munkahelyek üzemszünete, illetve a pihenésre távozó lakosság nagyobb számában keresendő. Ezen információ biznyságul szolgál azon állításnak, miszerint mind a napon belüli órák, mind pedig a héten belüli nap elhelyezkedése komoly befolyással bír a fogyasztói szokásokra, ezek által pedig a hőteljesítmény-igény mértékére.

Ennek értelmében az elemzésünk során figyelembe vettük a nap óráját, a hét napját, illetve a hónap napját is a fogyasztói trendek pontos lekövetéséért.

4.3.2. FELHASZNÁLT ADATOK

Az elemzés első lépése a rendelkezésünkre álló adatstruktúrában tárolt információ megértése és a benne tárolt alapvető mintázatok felismerése. Ezt követően a következő lépéseken kell végig haladni:

- célok meghatározása: esetünkben a hőigények előrejelzése
- kiválasztás, csökkentés és szűrés
- modellek kiválasztása
- értelmezés

Ezt követően azonosítottuk az előrejelzést befolyásoló tényezőket, amelyeket bemeneti adatként alkalmaztunk a modellek során. Ezen bemeneti paramétereket a **Jelentkező igények** fejezetben tárgyalt kategóriákba soroltuk, illetve meghatároztuk még az idő paramétereket is, amelyeket az alábbi, **6. táblázatban** foglaltunk össze.

6. táblázat: A bemeneti paraméterek adatszoportba, illetve szettekbe foglalva

Paraméter szettek	Bemeneti paraméterek																	
	Idő				Externális						Internális adatok							
	Month	DoM	DoW	HoD	$T_{k\ddot{u}}ls\ddot{o}$	$\varphi_{k\ddot{u}}ls\ddot{o}$	T_{harmat}	$wdir$	$v_{sz\ddot{e}l}$	ρ	$T_{e, t-6}$	$T_{e, t-24}$	ΔT_{t-6}	ΔT_{t-24}	m_{t-6}	m_{t-24}	Q_{t-6}	Q_{t-24}
Teljes szett	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
24 órás szett	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗	✗	✗
Externális szett	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗

A **Jelentkező igények** fejezetben említettük, hogy a korábbi internális adatok befolyással vannak a jelenlegi hőtéljesítmény-igényekre, emiatt integráltuk az előremenő hőmérséklet, az előremenő és visszatérő hőmérséklet közötti hőmérséklet különbség, a tömegáram és a kiadott hőtéljesítmény 6, illetve 24 órával elcsúsztatott értékeit is minden adatsorhoz, amelyek segítségével az éppen alkalmazott csúsztatás óraszámával egyező időtávban tudunk előrejelezni.

A bemeneti paramétereket idő, externális és internális paraméter csoportokba osztottuk, illetve készítettünk három szettet. A szettek segítségével meg tudjuk határozni, hogy egyes paraméterek elhagyásával vagy integrálásával miként változik a modellünk előrejelzési pontossága, és mely paraméterek milyen súlyozási értéket kapnak. A súlyozási érték adja meg, hogy egy paraméter (pl.: hőmérséklet) mekkora hatással van a kimenet eredményére, tehát a nagy súly erős korrelációt jelent, míg az alacsony súly arra enged következtetni, hogy a paraméter kevésbé befolyásolja a kimeneti adatot. A szetteket következőképpen alakítottuk ki:

- Teljes szett: tartalmazza az összes rendelkezésre álló kezdeti paramétert és az időjárás előrejelzés adatait is, ennek segítségével 6 órás időtávban tudunk előre jelezni.
- 24 órás szett: nem tartalmazza a 6 órával korábbi előremenő hőmérséklet, az előremenő és visszatérő hőmérséklet közötti hőmérséklet különbség, tömegáram és hőteljesítmény adatokat, ennek segítségével 24 órás időtávban tudunk előre jelezni.
- Externális szett: csak az externális adatokat tartalmazza, ezáltal nincsenek jelen az internális paraméterek által okozott megkötések, ám jóval alacsonyabb pontosság érhető el vele.

Ezen csoportosítások elvégzését követően a rendelkezésünkre álló összes adatsort felosztottuk a négy évszak szerint tavaszi, nyári, őszi és téli szakaszra, amely szakaszokat ezt követően egyesével adtunk be a különböző algoritmusoknak tanulási és tesztelési feladatra. Ezt a műveletet elvégeztük az összes paraméter szett és az összes hőkörzet esetén, amely így több, mint 180 individuális futtatást tett szükségessé. Ezen lépés azért volt szükséges, mert a gépi tanulási metódusok jobban meg tudják tanulni az ugyanazon évszak jellegzetes mintázatait és trendjeit, mintha az egész év rendszerére kéne megtenniük ugyanezt, ezáltal jelentősen növelni tudtuk az előrejelzési pontosságot.

4.4. Adatok előfeldolgozása

A különböző algoritmusok és modellek használata előtt több előfeldolgozási lépést kell végrehajtani az adaton a hibák kiszűrésének és az algoritmus optimális működési körülményeinek megteremtéséért. Ezen lépések alkalmazása során különös figyelmet kell fordítani az úgynevezett adatszivárgás elkerülésére. Az adatszivárgás azt az előfeldolgozási hibát foglalja magába, amikor tanítás és a kiértékelés során használt adatot nem különítjük el egymástól megfelelően, így előfordulhat, hogy a gépi tanulás során alkalmazott modell olyan információkhoz jut hozzá, amelyek a valóságban nem álltak volna a rendelkezésre, ezáltal torzítva a modellezés eredményét.

Az így előforduló hiba több szempontból is komoly kockázattal járhat, hiszen az adatok pontosságát és a modell megbízhatóságát veszélyeztetheti. Az adatszivárgás elkerülése végett fontos szem előtt tartani a valós szituációban lefolyó kiértékelési menet módját, mikor nem áll rendelkezésünkre a várt eredmény az egyes példányokhoz, csak azon

bemeneti adatok – másnéven attribútumok – melyek alapján az előrejelzést el kívánjuk végezni. Ezt figyelembe véve tehát fontos az előfeldolgozás megkezdése előtt szétválasztanunk a tanításra és kiértékelésre használt adatokat, ezzel garantálva a tanító és a teszt adatok egymástól való teljes elszigetelődését.

4.4.1. TRAIN-TEST SPLIT

A fentebb említett adatszivárgás elkerülésének leghatékonyabb módja az adatok két részre választása a további feldolgozás megkezdése előtt. Ezen lépést train-test split-nek nevezik, ilyenkor a tanításhoz és a teszteléshez szükséges adatokat különválasztjuk egymástól. Amennyiben ezzel a lépéssel kezdjük az előfeldolgozást, könnyen elkerülhető a tanításra és tesztelésre szánt adatok keveredése.

Idősoros adatok elemzése során kiemelkedő fontosságú elkerülni az adatok olyan keverését, aminek okán a tesztadatok közé – időrendi szempontból nézve – későbbi adatok kerülnek, mint maguk a tesztadatok. Ezen hibára azért fontos felhívni a figyelmet, mert a gépi tanulás alapú modellek esetén alap esetben tanácsos adatkeverést alkalmazni az adatsor robusztusságának növeléséért.

A szimulációink során a rendelkezésünkre álló 4 évnyi, azaz 48 havi adat a 75-25% train-test bontását követően 36 havi tanító adattal és 12 havi teszt adattal rendelkezünk. A bontás után az adat négy részletben áll rendelkezésünkre: tanító adat és ezen adatokhoz tartozó megoldások, illetve a teszt adat és a teszt kiértékeléséhez szükséges megoldások.

A tanító adatok alatt azon adathalmazokat értjük, amelyek alapján az előrejelzést végezzük, a tanító adatra való megoldások pedig a megoldásokat tartalmazzák. Fontos, hogy a következő alfejezetekben taglalt lépéseket mind a négy adathalmazra el kell végezni.

4.4.2. HIBÁS ADATOK SZŰRÉSE

A korábbiakban említettük, hogy a gépi tanulási modellek alkalmazása esetén kiemelkedő fontosságú a hibátlan adatsorok alkalmazása, mert kis mérési hibák, hamis értékek, avagy adatszolgáltatás-kimaradás esetén is teljesen eltérítheti az algoritmusokat a pontos tanulástól. Ennek érdekében hibaszűrést alkalmaztunk a teljes adatsoron.

A rendelkezésre álló értékek több szempontból történő vizsgálata során több helyen is tapasztaltunk rendellenes adatokat és mintázatokat, többek között az egyes ágak tömegáram, előremenő/visszatérő hőmérséklet és legfőképpen a prediktálni kívánt hőmennyiség értékei esetén. A megfigyelt hibák javítását két különálló lépésre bontottuk.

4.4.2.1. Az egész évre vonatkoztatott szűrés

Elsőnek az egész adatsort egy egységként vizsgálva kiszűrtük a hibásan mért hőmennyiségeket, melyek létezésére a hőmennyiség adatok időbeni ábrázolása vetett fényt. A **13. ábrán** jól látható, hogy egyes esetekben az aktuális trendtől nagymértékben eltérő, ugrásszerűen változó értékek vannak jelen, amelyek pár esetben negatív értékeket vettek fel.

Ezen kiugrások figyelmen kívül hagyása nagy mértékben ronthatja a predikció eredményét, ugyanis a pontok a valós esetekhez képest mutatott drasztikus eltérései nagymértékben befolyásolni tudják az egyes modellek döntési lépéseit mind a tanulási, mind pedig később a tesztfázis alatt.

Ezen hibák kijavítására a következő módszert választottuk: amennyiben a hőmennyiség adatok egészének átvizsgálása során olyan adatokat észleltünk, amelyek az általunk meghatározott határérték abszolútértékénél nagyobb mértékben eltértek a korábbi két órányi adat átlagértékétől, akkor azon adatokat hibásnak vettük. Az általunk alkalmazott határérték az elmúlt két óra átlagának 30%-aként volt meghatározva. A hibás adatokat ezt követően az előző két adatpont értékének átlagával helyettesítettük.

$$\dot{Q}_i = HA \left(\left| \dot{Q}_i - \frac{\dot{Q}_{i-1} + \dot{Q}_{i-2}}{2} \right| > \frac{\dot{Q}_{i-1} + \dot{Q}_{i-2}}{2} \cdot 0,3; \frac{\dot{Q}_{i-1} + \dot{Q}_{i-2}}{2}; \dot{Q}_i \right)$$

Ahol:

- \dot{Q}_i = a vizsgált adatpont értéke [W]

4.4.2.2. A nyári időszakok szűrése

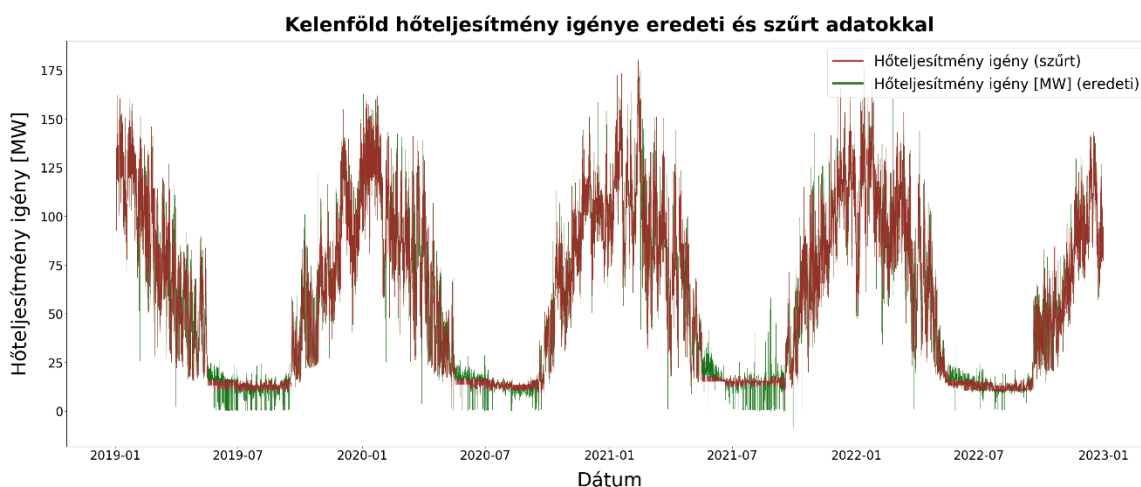
A második fajta tipikus hibát a nyári időszakokban tapasztaltuk. Ezen időszakokban a kiadott hőmennyiség lecsökken a HMV igények fedezésének szintjére, ami töredéke a fűtési időszakban jelentkező igényeknek. Ennek okán két probléma is jelentkezik: az első az, hogy a mérési pontatlanság relatív aránya jelentősen megnő az alacsony hőkiadás miatt, továbbá valamiért az alacsony értékek esetén a mérőrendszerek érzékenysége nem

tűnik mindenhol megfelelőnek a kis térfogatú áramok és hőmennyiségek mérésére, ami miatt fluktuációt tapasztalunk az adatokban.

A harmadik probléma a hőköztereket ellátó erőművek modulálási képességéből adódik, mivel a rendszerek nem képesek megfelelően alacsony hőteljesítmény kiadására, így ezt a rendszerek taktálásával, azaz periodikus ki-bekapcsolásával oldják meg. Ez sem az erőművek szempontjából, sem pedig az adatszolgáltatás szempontjából nem kedvező állapot, mert torzítja a valóban fellépő igények értékeit.

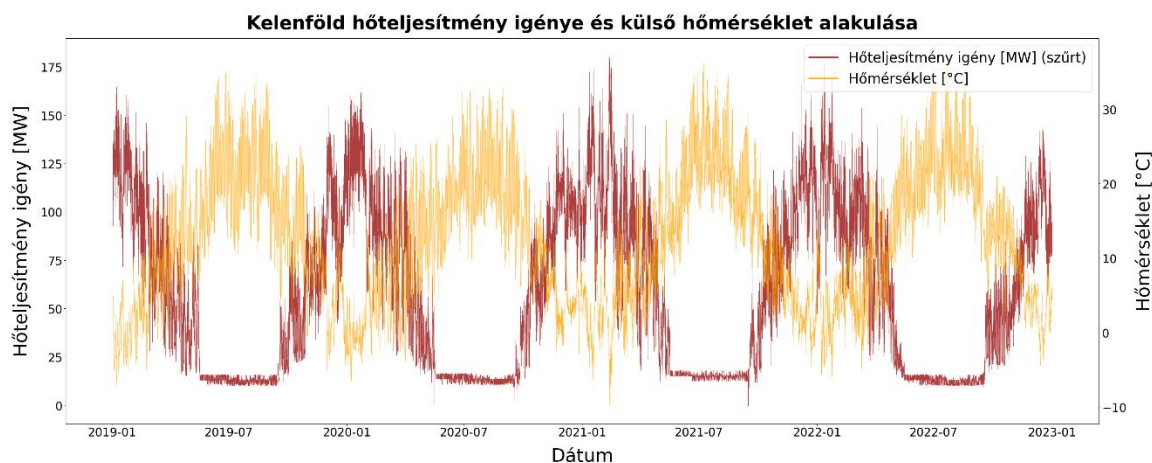
Ezen problémák megoldására egy második szűrési lépést hoztunk létre, amely során az egyes nyári szakaszokat különválasztva egyesével korrigáltuk azokat. A korrekció módja megegyezik **Az egész évre vonatkoztatott szűrés** c. fejezetben leírt korrekciós módszerrel, csak a két órával korábbi adatokhoz való hasonlítás helyett a teljes nyári szakaszra számolt átlaghoz viszonyítottuk az adatpontokat. A két szűrési mechanizmus összevont eredményét a **13. ábrán** jelenítettük meg. Az ábrán zöld színnel jelöltük az eredeti adatszolgáltatás adatait és barna színnel a már megszürt, csökkentett hibamennyiséget tartalmazó adatsort.

Jól látható, hogy az első szűrés egy lágyabb, kevés változtatással járó hibajavítást eredményez, míg a második, kizárólag a nyári szakaszok esetén alkalmazott korrekció során előfordult, hogy egymást követően több értéket is felül kellett írni a pontosabb eredmények érdekében. Érdekes megfigyelni ugyanakkor a korábban említett magasan kiugró értékek kiszűrését is.



13. ábra: A Főtáv Zrt.-től kapott eredeti hőteljesítmény adatok a kelenföldi hőközterre (zöld szín), illetve a szűrés utáni adatsor (barna szín)

A **14. ábrán** látható az immár megszürt hőteljesítmény-igény és a külső hőmérséklet-lefutás ábrázolása, ahol egyértelműen megfigyelhető a negatív korreláció a két paraméter között.



14. ábra: A kelenföldi hőközvet szűrt hőteljesítmény igényei és a külső hőmérsékletek alakulása

4.4.2.3. Hiányzó értékek helyettesítése

A hibás adatokon kívül sajnos jelentős mennyiségű részlegesen, avagy teljesen üres értékeket is tartalmaztak az adatsorok. A teljesen üres adatsorokat töröltük a rendszerből. Abban az esetben, ha az adatok részben hiányoztak, akkor a következő módszert alkalmaztunk: a sok elemből álló átlagértékek esetén megvizsgáltuk, hogy az átlagértéket összerakó értékek közül hiányzik-e elem és amennyiben igen, akkor helyettesítettük az értéket egy helyes adatsor értékéből. Ez legfőképpen az előremenő, illetve a visszatérő hőmérsékletek esetén volt alkalmazott művelet.

$$T_i = HA \left(DARAB.ÜRES \left(\frac{\sum_1^n T_{i,n}}{n} \right) > 0; T_i = z_i; \frac{\sum_1^n T_{i,n}}{n} \right)$$

Ahol:

- T_i = a vizsgált hőmérsékletpont értéke [°C]
- z_i = helyes hőmérsékletpont [°C]

A DARAB.ÜRES függvény megadja, hogy hány üres cella van a vizsgált adattartományban.

4.4.3. ADATOK SKÁLÁZÁSA

Az egyes modellek pontossága sok esetben növelhető pusztán az adatok megfelelő skálázásával. Ezen lépés szükségessége abban jelentkezik, hogy - főleg több különböző paraméter alkalmazása esetén - előfordulhat olyan adatsor, amiben a különböző paraméter értékek értelmezési tartománya nagyban eltér. Erre példa lehetnek a tömegáram és a hőmérséklet értékek, amelyek 10^4 -es és 10^1 -es nagyságrend körül mozognak.

A több nagyságrendbeli különbségek a modell szempontjából problémát jelentenek, mert sokszor szinte lehetetlenné válik a különböző léptékű adatok egységes felhasználása pontos eredmény elérésének érdekében. Emellett az adatok megfelelő skálázása felgyorsíthatja a tanítási folyamatot, mivel nem szükséges nagy számokkal számolnia az algoritmusnak.

A probléma megoldására a későbbiekben bemutatásra kerülő sklearn könyvtárcsomag MinMaxScaler függvénye nyújt megoldást. [12.] Ez a függvény a megadott adatokat az általunk megadott tartomány közé skálázza minden paraméter esetén, ezzel elérve az egységes értelmezési tartomány használatát.

A művelet alapja, hogy elsőnek standardizálja a skálázni kívánt adatokat, azaz oly módon transzformálja őket, hogy az átlagértékük 0, a szórásuk pedig 1 értékű legyen, majd átskálázza őket a kívánt tartományra. Fontos megemlíteni, hogy a skálázás ugyan lecsökkenti a kiugró adatpontok és a várható értékek közötti abszolút különbséget, az adatsor továbbra is változatlanul tartalmazza azokat.

4.5. Hiperparaméter optimalizáció

Az **Alkalmazott algoritmusok** fejezetben bemutatott algoritmusokat eddig csak általános működési szempontból vizsgáltuk, azonban az algoritmusok minél pontosabb betanításához elengedhetetlen a korábban is említett hiperparaméterek finomhangolása. A hiperparaméterek az adott algoritmus működését jellemző specifikus beállítások, amelyeket a futás előtt kell meghatározni. Az optimalizálás során az a cél, hogy megtaláljuk azt a hiperparaméter kombinációt, amely segítségével az algoritmus a legjobb eredményt tudja produkálni az adott adatsorra.

Mivel a hiperparaméterek algoritmusra jellemző kezdeti értékek, ezért az optimalizáció folyamatát algoritmusonként kell elvégezni olyan paraméterekkel, amelyek megfelelnek az adott algoritmusnak.

Az egységes, jól optimalizált és kontrolálható keretrendszer kialakítása érdekében a korábban említett regressziós algoritmusokat nem magunk készítettük el, hanem az sci-kit-learn könyvtárcsomagra támaszkodtunk. [13.]

4.5.1. SKLEARN

A könyvtárcsomag célja, hogy a tudományos és ipari területen egyaránt elterjedt Python programozási nyelvhez egy könnyen használható felhasználóbarát felületet nyisson a legkorszerűbb és legismertebb gépi tanuláshoz kapcsolódó algoritmusokhoz. Nagy előnye, hogy kiterjedt és nagy részletességű dokumentációval rendelkezik, amely megalkotása során törekedtek a szakzsargon alkalmazásának csökkentésére.

Munkánk szempontjából fontos volt, hogy az általunk prezentált eredmények reprodukálhatóak legyenek, mely alapesetben nem lehetséges, mivel ezen algoritmusok nem determinisztikusak. Ezen problémára is megoldást kínál a sickit-learn hiszen lehetséges olyan - minden algoritmushoz elérhető - beállítást megadni, mellyel rögzíteni lehet a kiindulási feltételeket. Ezen beállításnak a neve a „random_state”, mely alapértelmezetten nincs megadva - azaz nem determinisztikus a végkimenet - viszont amennyiben egy egész számot rendelünk hozzá, úgy a kiindulási bizonytalanságot meg lehet szüntetni. A modellek felállítása során minden algoritmus esetén alkalmaztuk a random_state kezdeti paramétert.

4.5.2. HIPERPARAMÉTEREK OPTIMALIZÁCIÓ MEGVALÓSÍTÁSA

Az optimális hiperparaméterek meghatározását nehezíti a tény, hogy sokváltozós problémával állunk szemben. Az egyik fajta megközelítés során manuálisan módosíthatjuk az egyes paramétereket egészen addig, míg elérjük a kívánt pontossági értéket, ám ez egy igen fárasztó és hosszadalmas megoldás lenne.

Ennek elkerülése végett az általunk választott alternatíva során meghatároztuk az egyes hiperparaméterek csoportjait, majd ezek felhasználásával a GridSearch nevű algoritmus

végighaladt a megadott értékek az összes lehetséges kombinációján, majd ezek alapján meghatározta a legkedvezőbb esetet. [14.]

Ez a gyakorlatban azt jelenti, hogy az összes bemutatott algoritmus esetén a GridSearch segítségével elvégeztük a hőteljesítmény előrejelzést, az előrejelzés és a valós mért teljesítmény adatok között tapasztalható hibaszázalék értéket (későbbiekben MAPE, lásd: **Átlagos abszolút százalékos hiba**) kiértékeljük, majd ezt a folyamatot elvégeztük az összes lehetséges hiperparaméterre, ezáltal meghatározva az optimális hiperparaméterek kombinációját. Az optimalizációt elvégeztük minden hőköri és minden algoritmus kombinációjának esetén a hőköri specifikus tulajdonságainak pontos modellezésének érdekében.

Az optimalizáció során lényeges figyelembe venni a számítógép hardvere által támasztott fizikai megkötéseket, nevezetesen a memória, illetve a számítási kapacitás terén. A futtatások időbeli szükségletei teljes mértékben a számítások elvégzésének sebességétől függ, melyet a tanulmány során használt számítógépek CPU-i (Central Processing Unit, avagy processzor) limitálnak. További korlátozást jelent a futtatások során tapasztalható memóriaigény, amely a nem megfelelő paraméterek megadása esetén könnyedén túlterhelte a RAM-ot (Random Access Memory, avagy tetszőleges hozzáférésű memória), ami ezáltal a teljes rendszer összeomlásához vezet. Egyes modellek RAM igénye meghaladta a több, mint 24 GB-ot.

A jelenleg rendelkezésünkre álló időkeret nem tette lehetővé az összes lehetséges hiperparaméter kombináció vizsgálatát, amelynek okát a kombinációk számával lehet szemléletesen bemutatni.

Vegyünk egy példát, ahol 5 különböző hiperparamétert szeretnénk változtatni és hiperparaméterenként 3 lehetőséget megvizsgálni. Ezek alapján $3^5=81$ különböző kombinációt szükséges kivizsgálni, amely – annak figyelembevételével, hogy egy tanítási szakasz 5-10 percet is igénybe – összesítve több, mint 13 órát vesz igénybe. Ha figyelembe vesszük, hogy ez egyetlen algoritmus potenciális futási idejét jelenti a hét alkalmazott modell közül, amelyeket mindhárom körzet esetén le kell futtatni, akkor látható lesz az összes lehetséges opció megvizsgálásának exponenciálisan növekvő időigénye.

Ennek fényében előzetes mérések és tapasztalatok alapján előre meghatároztuk, azon kezdeti paraméter értékek csoportját, amelyek potenciálisan az optimális, avagy az ahhoz közel eső pontossági értékekhez tartoznak.

Az optimalizálási eredményeket és az egyes hiperparaméter rendszerek szerepét és jelentését a mellékletben található **Végső hiperparaméterek és jelentésük** c. fejezetben mutattuk be.

4.6. Eredmények kiértékelése

4.6.1. HIBAÉRTÉKELŐ MÓDSZEREK

Az eredmények kiértékelése során kiemelkedő fontosságú a megfelelő hibaérték mutató kiválasztása a megfelelő pontosság elérésének érdekében.

4.6.1.1. Négyzetes középhiba

A négyzetes középhiba, vagy másnéven RMSE (round mean square error) a maradékok (előrejelzési hibák) szórása. A hibaértékek azt mérik, hogy az adatpontok milyen messze vannak a regressziós görbétől, az RMSE pedig meghatározza ezen hibaértékek szórását. Más szóval azt mutatja meg, hogy az adatok mennyire koncentráltak a legjobb illeszkedés görbéje körül. A négyzetes hiba középértékét gyakran használják a klimatológiában, az előrejelzésben és a regresszióelemzésben a kísérleti eredmények ellenőrzésére.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$$

Ahol:

- \hat{y}_t = az előrejelzett hőteljesítmény-igény [W]
- y_t = a valós hőteljesítmény-igény [W]

4.6.1.2. Átlagos abszolút hiba

Az átlagos abszolút hiba, vagy másnéven MAE (mean absolute error) A statisztikában az azonos jelenséget kifejező páros megfigyelések közötti hibák mérőszáma. Az MAE-t az abszolút hibák összegének és a minta méretének hányadosaként számítják ki.

$$MAE = \frac{\sum_{t=1}^n |\hat{y}_t - y_t|}{n}$$

4.6.1.3. Átlagos abszolút százalékos hiba

Az átlagos abszolút százalékos hiba, vagy másnéven MAPE (mean absolute percentage error) a MAE értékek százalékosot értéke a vizsgált pontok valós értékéhez viszonyítva.

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{\hat{y}_t} \right|$$

A jelen esetben ez a mérőszám megadja a mért hőteljesítmények és az előre jelzett hőteljesítmények közötti százalékos hibahatárt, amely általánosan értelmezendő az egész adatsorra. Fontos megjegyezni, hogy a MAPE százalékban kifejezett értéke több információval szolgál a nagy mértékben változó igényű távhőszolgáltatási szektor számára.

Gyakorlatiasan kifejezve a százalékos hibaértéket össze lehet hasonlítani egy kis hőteljesítmény-igényű és egy nagy teljesítmény igényű hőkörizet esetén is, mivel az érték relatív az igényekkel, ellenben a MAE és az RMSE mérőszámokkal, amelyek MW dimenzióban fejezik ki a hibát és nem engednek következtetni a tényleges értéktől való eltérés mértékére.

Ezen megfontolások alapján az algoritmusok pontosságának kiértékelése során a MAPE értékeket fogjuk alkalmazni.

Ezen műveletet követően 4+4, azaz nyolc évnyi adatsor állt rendelkezésünkre, amelyet alkalmaztunk az LSTM futtatásához. Ennek okán a **22. ábrán** lévő diagramon látható, hogy 2026-os - azaz az újonnan létrehozott - adatokra történt a tesztelés, ez az értékek hitelességét nem befolyásolja.

4.6.2. A FŐTÁV EMPIRIKUS MODELLJE

Az eredményeink megfelelő kiértékeléséhez stabil alapot biztosít a Főtáv Zrt. által jelenleg alkalmazott lineáris előrejelzési függvény, amelyet a rendelkezésünkre bocsátottak. A függvénynek van pár limitációja, nevezetesen az, hogy nem alkalmazható:

- 15 °C-os külső hőmérséklet felett,
- illetve a fűtési időszakon kívül, azaz május 16. – szeptember 15. között

A három vizsgált hőkörizetre három, kevés eltérést tartalmazó képlet van, amelyeket az alábbi egyenletben ábrázolunk.

$$\hat{y}_t = HA(T_{k,i} \leq 12; T_{k,i} \cdot m_1 + b_1; T_{k,i} \cdot m_2 + b_2)$$

Ahol:

- $m_1 = 12$ °C-os külső hőmérséklet alatti meredekség [-]
- $m_2 = 12$ °C-os külső hőmérséklet feletti meredekség [-]
- $b_1 = 12$ °C-os külső hőmérséklet alatti eltolás [-]
- $b_2 = 12$ °C-os külső hőmérséklet feletti eltolás [-]

Az egyes hőközrzetek más-más paraméter értékeket kaptak a képlet alapján, így ezeket a **7. táblázatban** jelenítettük meg.

7. táblázat: A Főtáv által alkalmazott előrejelző függvény paraméterei hőközrzetek szerint

	T_k	12 °C alatt	12 °C felett
Kelenföld	m	-6.893	-6.875
	b	145.184	145.040
Csepel	m	-2.407	-2.374
	b	54.078	53.944
Kispest	m	-5.264	-5.313
	b	115.645	115.781

A szükséges számítások elvégzését követően meghatároztuk az összes hőközrzetre a jelenlegi hőtéljesítmény predikciós hibaértéket MAPE metrikában:

- kelenföldi távhőközrzet: 26,5%
- csepeli távhőközrzet: 28,87%
- kiscesti távhőközrzet: 27,9%

Ezen értékek figyelembevételénél érdemes megjegyezni, hogy ezen MAPE értékek csak a korábbi előfeltételeknek megfelelő időpillanatokban érvényesek, a 15 °C-os külső hőmérsékletek feletti, illetve a fűtési szezonon kívül eső időpillanatokban nem tud megfelelő eredményt garantálni, aminek értelmében az év megközelítőleg 35-45% során nem használható. Ezzel szemben az általunk alkotott modell az év minden szakára képes előrejelzést alkotni.

4.6.3. EREDMÉNYEK VIZUALIZÁLÁSA

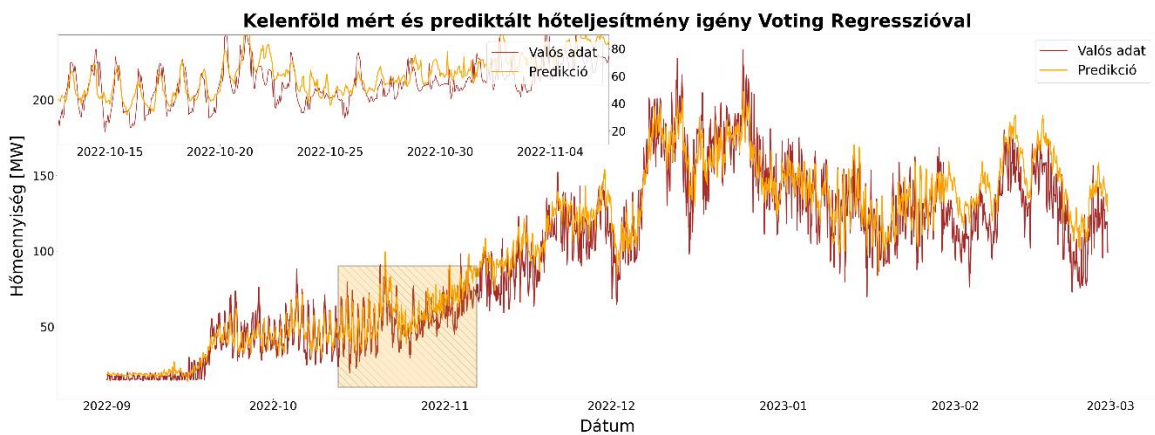
A különböző modellek alkalmazásával kapott eredményeinket elsősorban a MAPE értékük által értékeltük, mivel a százalékban történő kifejezése lévén ezen hibaértékelő módszer határozza meg legjobban az nagy skálán változó üzemállapotok során tapasztalt pontosságot. A predikciós metódusok eredményeinek értékeléséhez lényeges alapot

biztosít a **15. ábra**, amely meghatározza az előrejelzések minőségi kategóriáit a MAPE érték alapján.

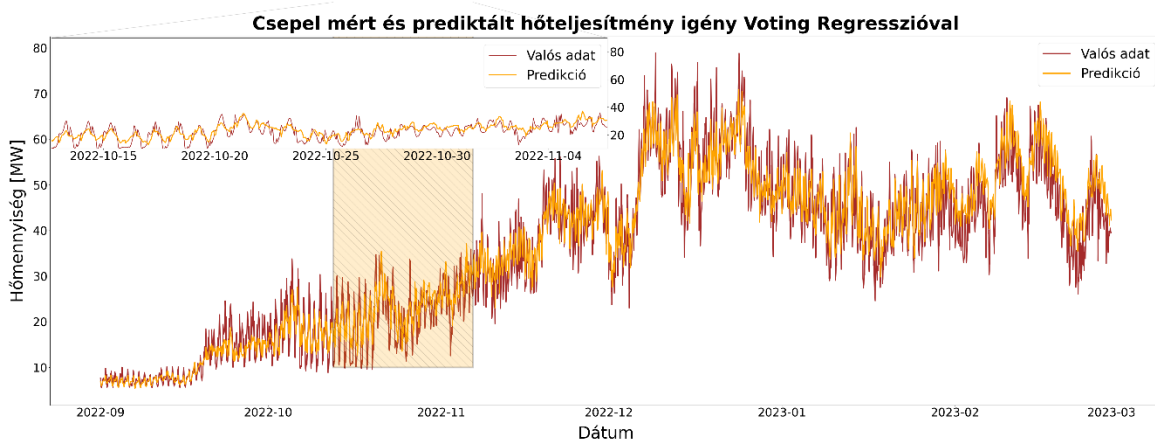
Átlagos abszolút százalékos hiba (MAPE)	Előrejelzés erőssége
<10%	Kiemelkedően pontos előrejelzés
10~20%	Jó előrejelzés
20~50%	Elfogadható előrejelzés
>50%	Gyenge és pontatlan előrejelzés

15. ábra: Előrejelzési pontosság minőségi kategóriái [15.]

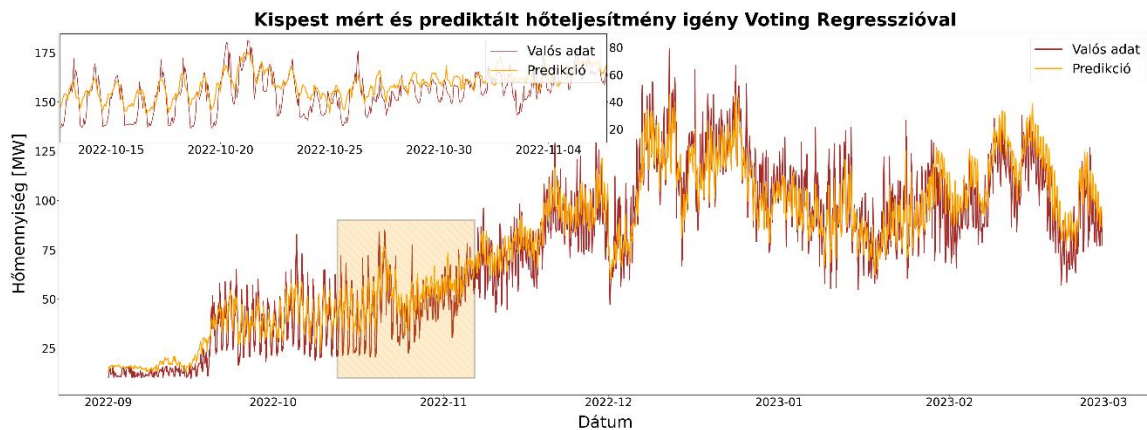
A következőkben bemutatjuk az egyes algoritmikus metódusokkal elért előrejelzéseinket diagram, illetve táblázat formában is. Az alább látható **16. ábrán**, **17. ábrán** és **18. ábrán** vizualizáltuk ugyanazon szakaszát a teszt fázisnak.



16. ábra: Kelenföld mért és prediktált hőteljesítmény-igénye Voting Regresszorral



17. ábra: Csepel mért és prediktált hőteljesítmény-igénye Voting Regresszorral



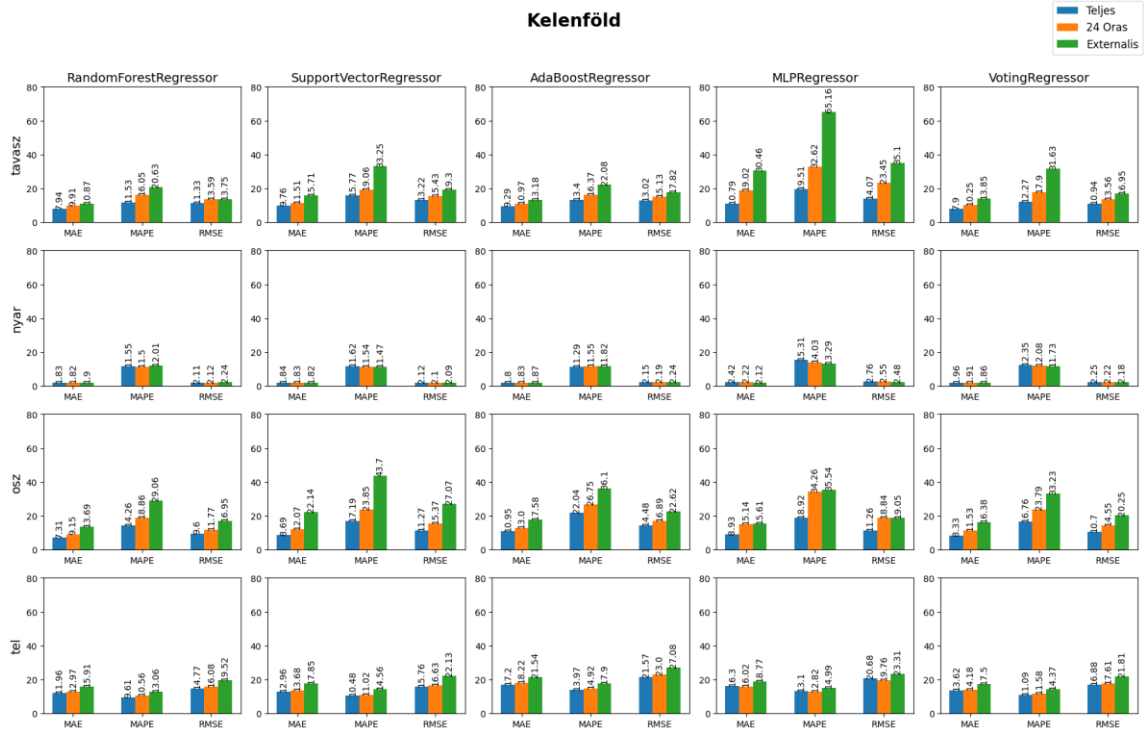
18. ábra: Kispest mért és prediktált hőteljesítmény-igénye Voting Regresszorral

Az ábrákon narancssárga színnel jelöltük a prediktált értékeket, és barna színnel a mért adatokat. A diagramok bal felső sarkában kiemeltük a diagram területén sárga sraffozott területben található függvény szakaszt a jobb láthatóság érdekében.

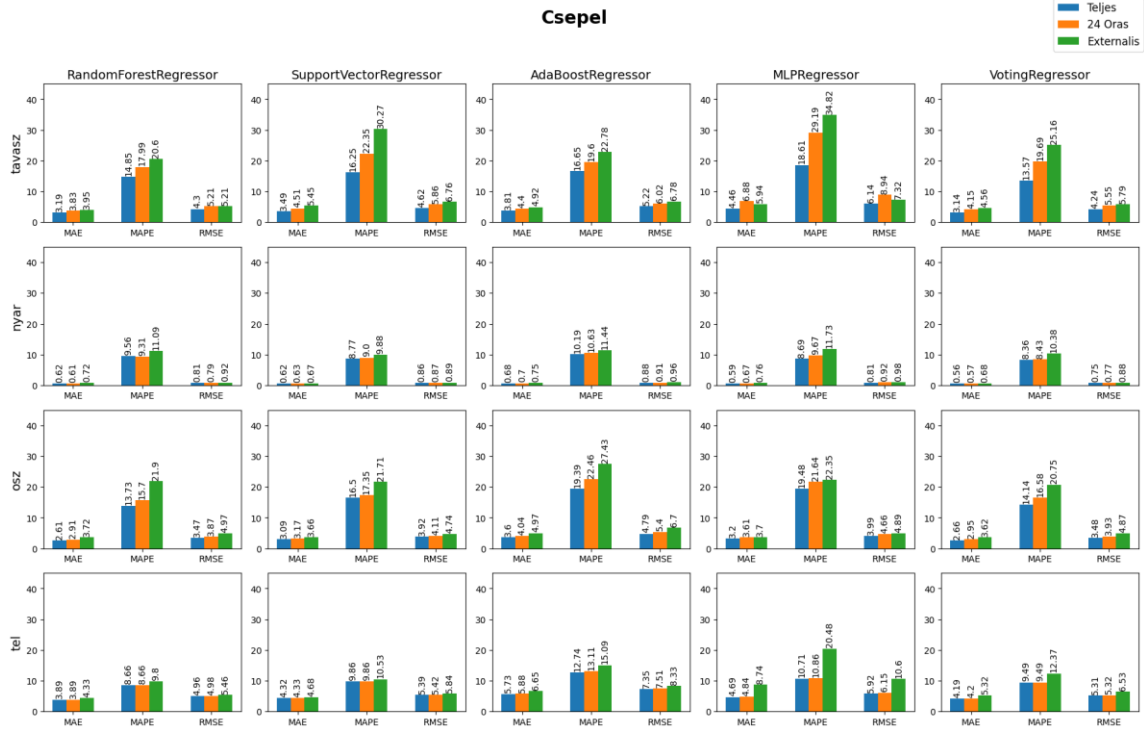
A korábbiakban említetteknek megfelelően a MAPE értékeket alkalmaztuk a fő hibaérték indikátorként, amely összegyűjtött eredményeit a következő, **19. ábrán**, **20. ábrán** és **21. ábrán** mutatunk be, kiegészítve a MAE és RMSE értékekkel egyetemben.

A következő három ábrán kék színnel jelöltük a teljes paraméterszett, narancssárgával a 24 órás paraméterszett és zölddel az externális paraméterszett alkalmazásával elért MAPE értékeket, amelyeket meghatároztunk a korábban részletezett évszakos bontás minden szakaszára külön-külön. A bontás eredményeképpen minden hőközvet minden évszakának predikciójáért egy külön algoritmus felel, amelyek az év megfelelő szakaszában aktiválódnak.

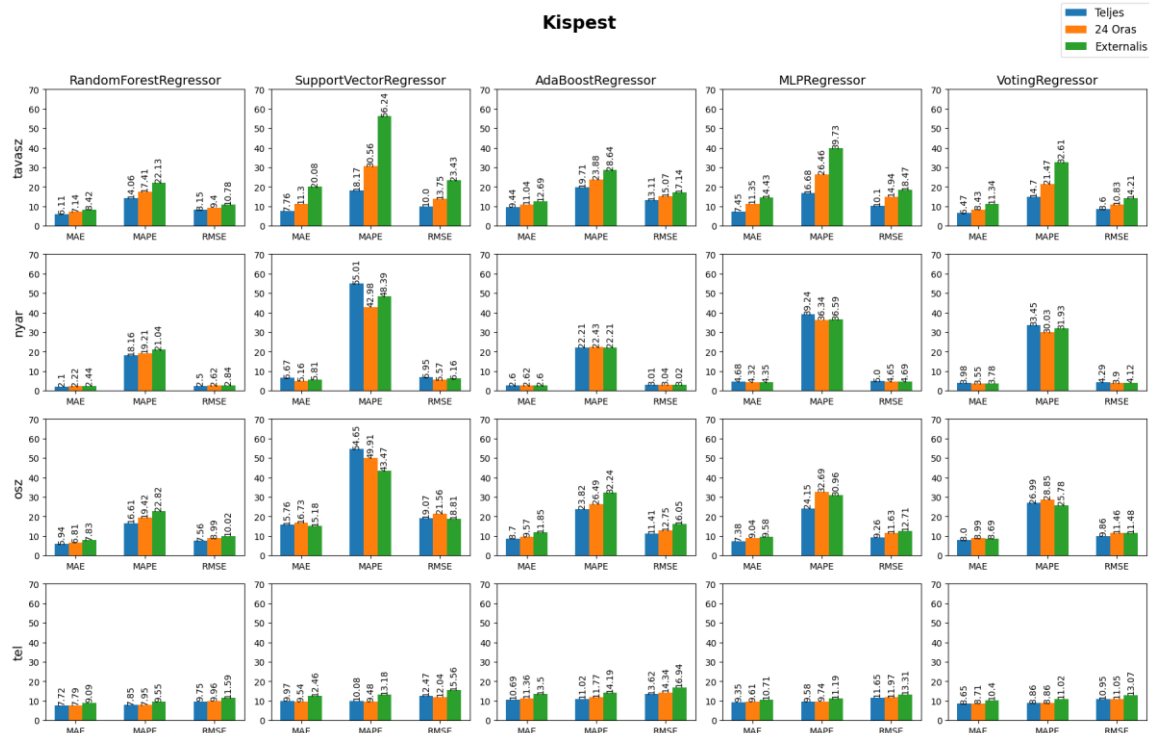
A MAE értékek és az RMSE értékek egyaránt megawatt [MW], a MAPE értékek pedig százalék [%] dimenzióban értelmezendők.



19. ábra: MAPE értékek Kelenföld hőközret esetén a teljes, a 24 órás és az externális paraméterszetre évszakok szerint



20. ábra: MAPE értékek Csepel hőközret esetén a teljes, a 24 órás és az externális paraméterszetre évszakok szerint



21. ábra: MAPE értékek Kispest hőközret esetén a teljes, a 24 órás és az externális paraméterszetre évszakok szerint

Az előző három oszlopdiagramban ábrázolt MAPE eredmények értékeit a **8. táblázatban** foglaltuk össze az egyszerűbb értelmezhetőség érdekében.

8. táblázat: MAPE értékek a három vizsgált hőközret és paraméterszett esetén

		Kelenföld					Csepel					Kispest				
		Tavaszi	Nyári	Őszi	Téli	Átlag	Tavaszi	Nyári	Őszi	Téli	Átlag	Tavaszi	Nyári	Őszi	Téli	Átlag
Teljes szett	RF	11.53	11.55	14.26	9.61	11.74	14.85	9.56	13.73	8.66	11.70	14.06	18.16	16.61	7.85	14.17
	SVR	15.77	11.62	17.19	10.48	13.77	16.25	8.77	16.50	9.86	12.85	18.17	55.01	54.65	10.08	34.48
	ADA	13.40	11.29	22.04	13.97	15.18	16.65	10.19	19.39	12.74	14.74	19.71	22.21	23.82	11.02	19.19
	MLPR	19.51	15.31	18.92	13.10	16.71	18.61	8.69	19.48	10.71	14.37	16.68	39.24	24.15	9.58	22.41
	Voting	12.27	12.35	16.76	11.09	13.12	13.57	8.36	14.14	9.49	11.39	14.70	33.45	26.99	8.86	21.00
24 órás szett	RF	16.05	11.50	18.86	10.56	14.24	17.99	9.31	15.70	8.66	12.92	17.41	19.21	19.42	7.95	16.00
	SVR	19.06	11.54	23.85	11.02	16.37	22.35	9.00	17.35	9.86	14.64	30.56	40.98	49.91	9.48	32.73
	ADA	16.37	11.55	26.75	14.92	17.40	19.60	10.63	22.46	13.11	16.45	23.88	22.43	26.49	11.77	21.14
	MLPR	32.62	14.03	34.26	12.82	23.43	29.19	9.67	21.64	10.86	17.84	26.46	36.34	32.69	9.74	26.31
	Voting	17.90	12.08	23.79	11.58	16.34	19.69	8.43	16.58	9.49	13.55	21.47	30.03	28.85	8.86	22.30
Externális szett	RF	20.63	12.01	29.06	13.06	18.69	20.60	11.09	21.90	9.80	15.85	22.13	21.04	22.82	9.55	18.89
	SVR	33.25	11.47	43.70	14.56	25.75	30.27	9.88	21.71	10.53	18.10	56.24	48.39	43.47	13.18	40.32
	ADA	22.08	11.82	36.10	17.90	21.98	22.78	11.44	27.43	15.09	19.19	28.64	22.21	32.24	14.19	24.32
	MLPR	65.16	13.29	35.54	14.99	32.25	34.82	11.73	22.35	20.48	22.35	39.73	36.59	30.96	11.19	29.62
	Voting	31.63	11.73	33.23	14.37	22.74	25.16	10.38	20.75	12.37	17.17	32.61	31.93	25.78	11.02	25.34

A táblázatban megtalálható az összes algoritmus eredménye minden évszakra külön meghatározva, amelyeket ezt követően a színezett oszlopokban átlagoltunk. Ezt a folyamatot mindhárom hőközre elvégeztük.

Mind az oszlopdiagramok, mind pedig a táblázat alapján megfigyelhető, hogy a SVR algoritmus és az MLP Regressor neurális háló sok esetben kiemelkedően magas hibaértéket produkált. Az MLP esetén ez a hiba betudható az évek felbontásával okozott adatmennyiség csökkenésnek, aminek hatására a kiterjedt hálóstruktúrájú modell nem jutott megfelelő mennyiségű adathoz. Ezt a jövőben orvosolni kívánjuk a hálóstruktúra méretének és összetételének módosításával. Az SVR modell a tapasztalatok alapján kevésbé volt alkalmas a teljesítmény előrejelzésre, mint a többi vizsgált algoritmus.

A legjobb eredményt a Random Forest Regressor érte el, amely annak tudható be, hogy a Voting Regressor algoritmus esetén nem alkalmaztunk módosított súlyozási paramétereket, amely a jövőben egy fejlesztési lehetőségként kínálkozik.

Észrevehető, hogy a paraméterszettek változtatásával fokozatosan romlottak a MAPE értékek. Ennek oka a **Felhasznált adatok** c. fejezetben foglaltaknak megfelelően az internális adatok mennyiségének változásához köthető: a kizárólag externális adatokkal dolgozó szettek esetén lettek a legmagasabb hibaértékek, a legjobb eredményeket pedig a 6, illetve 24 órás eltolt adatokat tartalmazó szett mutatta fel.

4.6.3.1. Időfüggetlen metódusok eredményeinek összehasonlítása a jelenlegi rendszerrel

A kiértékelés egyik legfontosabb lépése az eredmények összevetése a jelenleg alkalmazott empirikus lineáris modellel, ezek értékeit a **Főtáv empirikus modellje** c. fejezetben részleteztük. Esetünkben a legjobban teljesítő Voting-, illetve a Random Forest Regressor eredményeivel vetjük össze a MAPE értékeket.

- Voting Regressor:
 - Kelenföld: 13,12% → -13,38% MAPE csökkenés
 - Csepel: 11,39% → -17,48 MAPE csökkenés
 - Kispest: 21% → -7,87 MAPE csökkenés
- Random Forest Regressor:
 - Kelenföld: 11,74% → -14,76% MAPE csökkenés
 - Csepel: 11,7% → 16,2% MAPE csökkenés
 - Kispest: 14,17% → 13,73% MAPE csökkenés

Ezen adatok alapján a gépi tanulást alkalmazó metódusok, nevezetesen a Voting regressor alkalmazásával **60,5%-os** hőtéljesítmény-igény előrejelzés pontosság növekedést

tudtunk elérni a csepeli hőközretre, **55,7%**-os növekményt a kelenföldi hőközretre és **49,2%**-os a kispesti hőközretre.

A prezentált hőteljesítmény-igény előrejelzési pontosságának növelése direkt módon növeli az erőművek hatásfokát és szabályozhatóságát, csökkenti a hirtelen, előre nem látott igényváltozás esetén szükséges teljesítménycsúcsokat, illetve a visszatérő hőmérsékletet a túltermelés elkerülésével.

Ezen felül növeli a távhőszolgáltató ellátásminőségét és -biztonságát, továbbá mind az erőművek, mind pedig a szolgáltató bevételeit direkt módon emeli. Ezek alapján kijelenthető, hogy összetett rendszerek – mint esetünkben három, kerület nagyságú hőközret – ellátása során az egyik legfontosabb tényező a pontos hőteljesítmény-igény előrejelző rendszer alkalmazása, amelynek segítségével direkt módon tudjuk javítani mind az energetikai, mind pedig a gazdasági mutatókat.

A HUHA 2 létesítésének szempontjából továbbá kritikus jelentőséggel bírnak a jelen tanulmány során elért eredmények, mivel ezek lehetővé teszik a 11,5 km-es összekötőszakasszal rendelkező erőmű pontos teljesítmény szabályozását bármely távhőcsatlakozási konstrukció kialakítása esetén.

4.6.3.2. LSTM kiértékelése

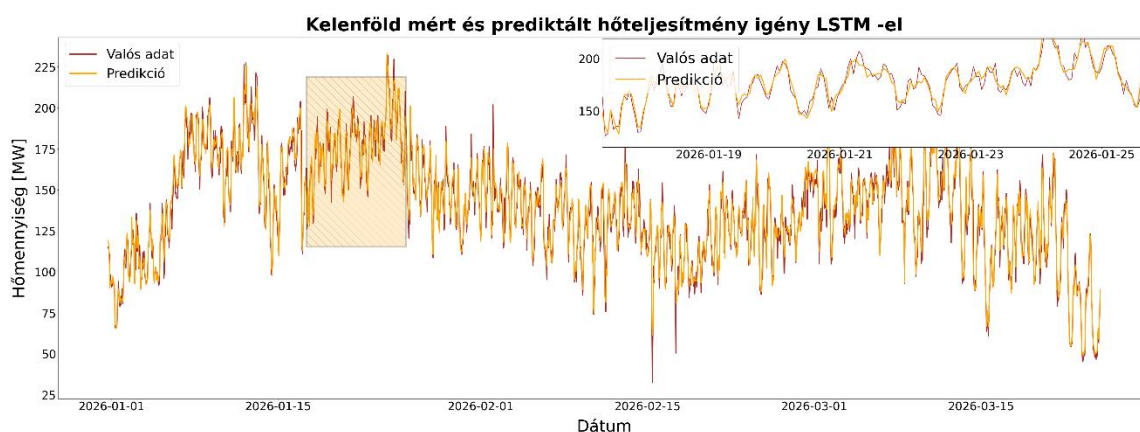
Az LSTM modellt az **Alkalmazott algoritmusok** c. fejezetben említetteknek megfelelően kitekintésként vizsgáltuk meg az időfüggetlen metódusok mellett, amelyekkel számos tanulmány foglalkozott már.

Az általunk alkalmazott LSTM-et használó komplex neurális háló öt rétegből áll, amelynek második rétegén alkalmaztuk az LSTM architektúrát. A háló az általunk alkalmazott legösszetettebb modell, ami több, mint 1 130 000 paramétert alkalmaz a számítás során. Az összetett és kiterjedt kialakítása lehetővé teszi a jóval pontosabb előrejelzést, ám emellett még a korábban bemutatott modellekhez képest is jóval megnövekedett bemeneti adatigénye van a kívánt hibahatár eléréséhez. Összehasonlításképpen Idowu et al. 4 havi, Davy et al. pedig 27 havi mérési adatsort alkalmazott a rendszerük felállításához, az általunk felépített modellekhez 4 évnyi, azaz 48 havi mérési adatsort alkalmaztunk, ám ez sem volt elegendő tanulási adat az LSTM modellnek. [4.][11.]

Ezen probléma megoldásának érdekében mesterséges adatokat hoztunk létre a rendelkezésünkre álló adatstruktúra alapján, amelyeket a korábbi évekhez hozzáillesztve

tanítottuk és teszteltük a neurális hálót. A mesterséges adatokat Gauss-i zaj hozzáadásával hoztuk létre, amely során az egész - négy éves – adatsor összes értékéhez a NumPy könyvtár `random.normal` programkódjának alkalmazásával egy gaussi eloszlás alapján számított véletlenszerű hibaértéket rendeltünk, amely így pozitív, avagy negatív irányba is eltolhatta az adott pont értékét. [16.] Ennek segítségével új adatokhoz jutottunk, amelyek, habár tartalmazták az alapból rendelkezésünkre álló információk karakterisztikáit, de nem egyeztek meg velük.

Az időfüggetlen algoritmusok mellett kitekintésként alkalmazott LSTM eredményei a **22. ábrán** láthatóak. Ezen neurális hálót csak Kelenföld hőkörzet esetén alkalmaztuk kísérleti jelleggel az időfüggést figyelembe vevő metódusok potenciális hibacsökkentési lehetőségeinek vizsgálatáért.



22. ábra: Az LSTM neurális háló kizárólag externális bemeneti adatok mellett elért előrejelzésének részlete

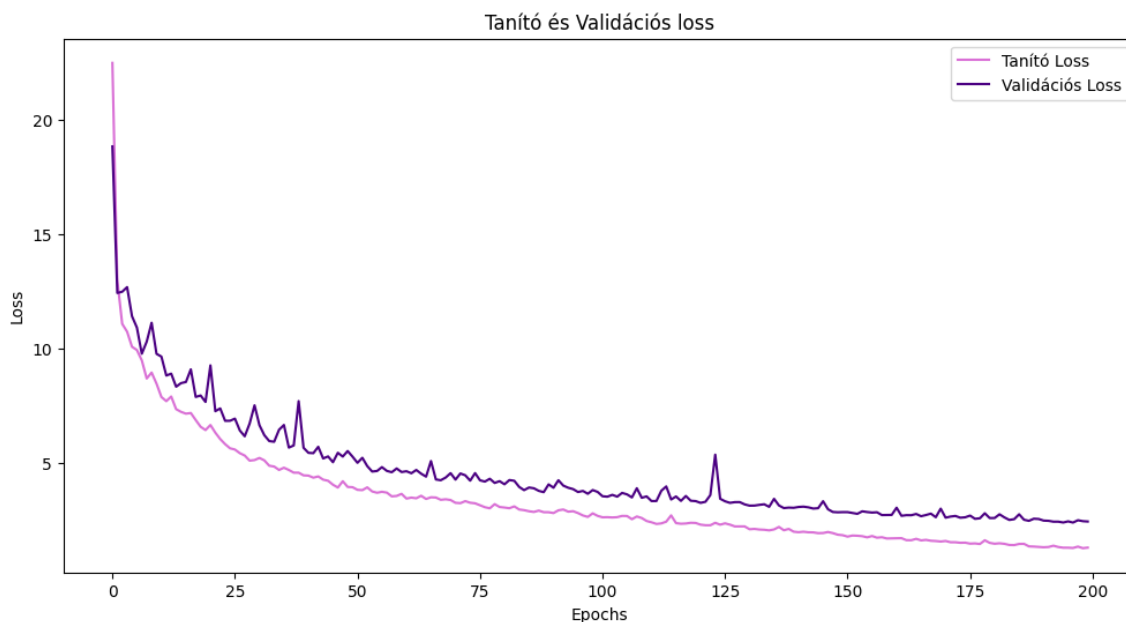
A diagramon barna színnel jelöltük a mérési adatokat és narancssárga színnel jelöltük az LSTM által végzett predikciós értékeket. A diagram jobb felső sarkába kiemeltük a diagram területén sárga sraffozott területben található függvény szakaszt a jobb láthatóság érdekében.

A neurális háló kiértékelését követően jelentős mértékű javulást tapasztaltunk a MAPE értékek tekintetében, amely még inkább kiemelkedő eredmény annak tekintetében, hogy az általunk végzett futtatások során az LSTM kizárólag az externális paraméterszett adatai alapján végezte el a predikciókat, tehát nem rendelkezett az összes többi algoritmus számára rendelkezésre álló 6, illetve 24 órával elcsúsztatott előremenő hőmérséklet, hőmérsékletkülönbség, tömegáram, és hőteljesítmény adatsorral sem. Ez nem csupán arra

enged következtetni, hogy a trendek felismerése kiemelkedően fontos szerepet játszik a hőteljesítmény-igények előrejelzése során, hanem arra is, hogy további adatok megadásával tovább lehet pontosítani a modellt.

Mivel az algoritmus nem rendelkezett internális adatokkal, az egyetlen fizikai korlátja az előrejelzési időtartamának a bemeneti időjárás adatok előre mutatása volt, ellentétben az internális adatokat alkalmazó többi algoritmussal, amelyek maximális előrejelzési határidejének gátat szab az internális adatok eltolásának mértéke.

A tanulási és validációs folyamat során tapasztalt hibaértékeket a **23. ábrán** mutattuk be. A validáció a tanulási fázissal párhuzamosan futtatott lépés, ami lehetővé teszi, hogy a módszer a tanulási fázis folyamán minden egyes epochot követően – amikor a tanító adathalmaz összes elemén végig ért egyszer – kiértékelje a predikciós képességét ezáltal elkerülve az esetleges „overfit”-et (az algoritmus túlzott illeszkedése a tanító adathalmazra, lásd **Multilayer perceptron** fejezet). Validáció segítségével felismerhető az overfit jelensége, amikor a validációs fázis során mért hibaértékek csökkenő tendenciája stagnáló, avagy növekvő jellegűt vesz fel, míg a vele azonos epochban mért tanítási adaton mért hibaértékek továbbra is csökkennek. Az alábbi ábrán látható, hogy esetünkben nem következett be overfit jelenség.



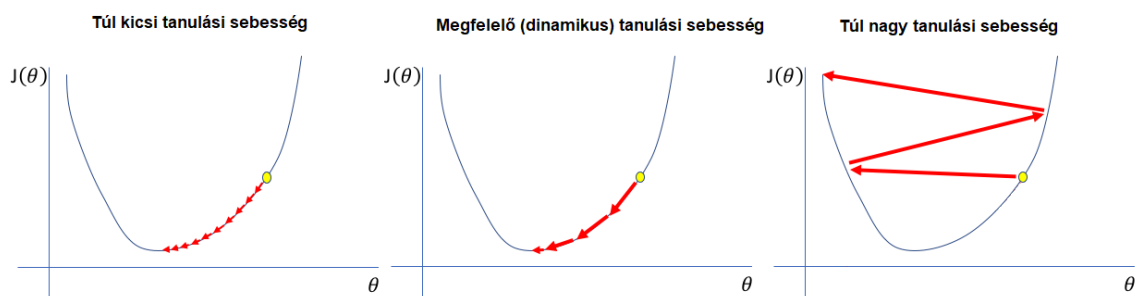
23. ábra: Az LSTM neurális háló hibaértékeinek alakulása a tanulási és validációs fázis alatt

A fenti diagramon rózsaszín színnel jelöltük a tanulási fázisban, illetve lila színnel a validációs fázisban tapasztalt hibaértéket, amelyek pozitív korrelációban vannak a MAPE értékek változásával. A függvények jellegén jól látható a kezdeti epochokban

nagymértékű tanulás volt tapasztalható, amely meredek eséssel jelenik meg, majd kb. a 15. epochot követően egyenletes mértékben csökkenő jelleget figyelhetünk meg.

A lassabban csökkenő jelleg főként két tényező miatt tapasztalható: az algoritmus pár epoch alatt rátanul az adatsorra, amelyet követően főként kisebb javulási értékeket tud elérni, a második tényező pedig abban található, hogy az epochok számának növekedésével a háló tanulási sebességét (lásd „learning_rate”, **Multilayer perceptron** fejezet) fokozatosan csökkentettük, ami elősegítette a globális minimum megtalálását, ám lelassította a javulás mértékét.

Ennek magyarázatára szemléletes példa, hogy a nagyobb lépésközök alkalmazásával gyorsabban tudjuk vizsgálni az adatsort, ám nagy esélyünk van arra, hogy a minimum-pontokat átugorjuk. Ezzel szemben a kisebb lépésközök alkalmazása lassabb futást és több iterációt eredményez, talán el sem jutva a minimum pontig, ám ha egyszer már megtaláltuk a közelítőleges elhelyezkedését, akkor sokkal alkalmasabb annak pontosítására. Ennek szemléltetése a **24. ábrán** található.



24. ábra: A tanulási sebesség (learning rate) mértékének hatása [17.]

Ezen két módszer kombinálásával tudjuk elérni az optimális megoldást, így 15 epochonként csökkentjük a tanulási sebességet. A tanulási sebesség iteratív csökkentése hozzájárul az overfit jelenség elkerüléséhez is.

4.6.3.2.1 Az LSTM eredményeinek összehasonlítása

Az **Időfüggetlen metódusok eredményeinek összehasonlítása a jelenlegi rendszerrel** c. fejezetben foglaltakhoz hasonlóan ezen alfejezetben az LSTM neurális háló eredményeit mutatjuk be. A **22. ábrán** és a **23. ábrán** bemutatott jellegek fényében a következő hőtéljesítmény predikciós átlagos abszolút százalékos hibaértékeket kaptuk:

- MAPE: 4,79 % \rightarrow 21,71% MAPE csökkenés a jelenlegi rendszerhez képest, és 6,95% hibacsökkenés a legjobb időfüggetlen metódus által elért eredményhez képest (RF: 11,71%)

- MAE: 2,69 MW
- RMSE: 4,1 MW

Ezen eredmények értelmében tehát az LSTM externális paraméterszett alkalmazása mellett **81,9%**-os MAPE csökkenést, és ezáltal pontosságnövekményt ért el a jelenleg alkalmazott rendszerhez képest, továbbá **40,9%**-os hibaszázalék csökkenést ért el a legjobb időfüggő gépi tanulási módszerhez képest (Random forest regressor). Ez figyelemre méltó javulás annak fényében, hogy az RF számára elérhető volt az összes internális adat is, míg az LSTM kizárólag az externális, tehát időjárás előrejelzési adatok alapján dolgozott.

A bemutatott eredmények kiválóan mutatják az időfüggést figyelembe vevő algoritmusokban rejlő potenciált. A **Időfüggetlen módszerek eredményeinek összehasonlítása a jelenlegi rendszerrel** c. fejezetben foglalt értékeléssel egyezésben ezen eredmények nagymértékű profit és potenciálisan - az ellátásbiztonság és komfortnövelés révén – fogyasztómennyég növekedést fognak eredményezni, továbbá felül reményeink szerint elősegítik a HUHA 2 hulladékhasznosító mű telepítésének megvalósulását.

5. EREDMÉNYEK ÖSSZEGZÉSE

A tanulmány során a Dél-budapesti távhőközrzeteken belül elhelyezkedő kelenföldi-, csepeli- és kispesti hőközrzet a HUHA 2 hulladékéroműhöz való távhőcsatlakozásának energetikai és gazdasági elemzését, illetve az említett hőközrzetek hőtéljesítmény igényének gépi tanulási algoritmusok alkalmazásával történt hőtéljesítmény-igény előrejelzését végeztük el. A távhőcsatlakozási elemzés során meghatároztuk az közrzeteken belül, illetve azokon kívül található vezetékszszakaszok hosszát és ezek alapján a rendelkezésre álló adatok alapján egy éves külső hőmérséklet és hőtéljesítmény esetén tapasztalt hőveszteségi értékeket óras bontásban. Megadtuk az egyes hőközrzetekre való csatlakozás esetén tapasztalt energetikai és gazdasági jellemzőket az R1 formula figyelembevételével, illetve az egyes kapcsolódási opciókhoz tartozó távhővezeték beruházási költséget, amelyek segítségével meghatároztuk a dinamikus megtérülési időtartamot. A Kispeszt, illetve a Kispeszt + Csepel konstrukciók bizonyultak a legjobb választásnak gazdasági szempontból tekintve, 15 és 14 havi megtérülési idővel, ám az összes vizsgált eset gyors megtérülési idővel rendelkezik, a legnagyobb érték 25 hónapos volt.

Ezt követően ezen hőközrzetek hőtéljesítmény-igényének gépi tanulás alapú előrejelzésével foglalkoztunk, amely esetén az RF, SVR, AdaBoost, MLP és a Voting regressor algoritmusokat és neurális hálókat alkalmaztuk. A bemeneti adataink megfelelő minőségének garantálásáért kidolgoztunk egy kétlépéses szűrési mechanizmust, amelyet sikeresen alkalmaztunk az egész adattartományra. A bemeneti adatokat három paraméterszetre osztottuk az eredmények bemeneti adatfüggésének meghatározásáért, majd pedig évszakokra bontottuk a megnövelt predikációs pontosság elérésének érdekében, amelyeket minden közrzet és algoritmus során alkalmaztunk.

Az egyes algoritmusok optimális futási eredményeinek garantálásáért hiperparaméter optimalizációt végeztünk az összes algoritmus és közrzet esetében, amelyek eredményei alapján a valós előrejelzési számításokat elvégeztük. A predikációs hiba fő mérőszámának a MAPE-t választottuk, de emellett feljegyeztük mind a MAE, mind pedig az RMSE értékeket is a futások során.

A kapott eredményeink kiértékelése során összevetettük az általunk alkotott modellek predikációs mérőszámait a Főtáv Zrt. által alkalmazott empirikus modellel. A legjobb idősoros adatokat figyelembe nem vevő modell, a Random Forest Regressor 11,7%-os egész

évre vetített MAPE értéket ért el, amely 17,48 %-kal alacsonyabb, mint a jelenleg alkalmazott rendszer, amely 60,5 %-os pontosságnövelési eredményt jelent az egész évre.

Az korábban bemutatott modellek mellett kitekintésképpen megvizsgáltunk egy időbeliséget figyelembe vevő neurális háló modellt, az LSTM-et, amelyet Kelenföld hőköri-
térére alkalmaztunk. A modell kellő pontossági értékének eléréséért gaussi zaj alkalmazásával mesterségesen létrehoztunk négy évnyi további adatot, amelyeket szintén beadunk az hálónak tanulás és validáció céljából. Az modell futtatásakor kizárólag az externális paraméterszettet alkalmaztuk.

A kapott eredmények messze felülmúlták mind a támasztott elvárásainkat, mind pedig az imént bemutatott időbeliséget figyelembe nem vevő metódusok által prezentált eredményeket. Az LSTM MAPE értéke az egész évre 4,79%-os értéket vett fel, amely a **15. ábrán** prezentált kategóriák szerint a kiváló predikciók között is magas értéket képviselő eredmény. A jelenleg alkalmazott rendszerhez képest ez 21,71%-os MAPE értékcsökkenést, és ezáltal 81,9%-os pontosság javulást értünk el, amely kiemelkedő eredménynek mondható.

Az LSTM sikerességét növeli, hogy ezen eredményt azon paraméterszeten teljesítette, amely a legrosszabb eredményeket produkálta az összes többi vizsgált algoritmus esetén. Ezek alapján az algoritmus alkalmazható jelen formájában kötetlen távolságú előrejelzés létrehozására – ameddig rendelkezésére áll időjárás előrejelzési adat -, avagy kiaknázható az internális paraméterek bevitelével elérhető további MAPE értékcsökkenés.

Összesítésképpen kijelenthető, hogy a jelen tanulmányban prezentált eredmények sikeresen meghaladták a beruházási, illetve az előrejelzési határértékeket. Tapasztalataink alapján mind a HUHA 2 távhőcsatlakoztatása, mind pedig a kidolgozott gépi tanulási módszerek kiemelkedően nagy bevétel-, szabályozási pontosság és ezek által felhasználói komfort növekedést tud elérni alacsony beruházási költségen, avagy teljesen ingyenesen. Az időbeliséget figyelembe vevő algoritmusok esetén kijelenthető, hogy alkalmasabbak a hasonló jellegű predikciós számítások pontos elvégzésére, mint az időbeliséget figyelembe nem vevő társaik. Ennek értelmében a jövőben ezen alternatívák tanulmányozását tervezzük folytatni.

6. FORRÁSOK

- [1.] Hulladékkezelés az Európai Unióban <https://www.europarl.europa.eu/news/hu/headlines/society/20180328STO00751/hulladekkezeles-az-eu-ban-trendek-es-statisztikak-infografika>
- [2.] Kocsis Kende: HUHA 2 hulladékégető mű létesítésének vizsgálata Budapest távhőellátásában, Energiagazdálkodás folyóirat, 64. évfolyam Klenen különszám, 2023. március.
- [3.] R1 energy efficiency formula <https://www.ciwm.co.uk/ciwm/knowledge/the-r1-energy-efficiency-formula.aspx>
- [4.] Samuel O. Idowu, Saguna Saguna és Olov Schelén: Applied Machine Learning: Forecasting Heat Load in District Heating System, Energy and Buildings, Energy and Buildings 00 (2015) 1–17
- [5.] Christian Johansson, Markus Bergkvist, Davy Geysen, Oscar De Somer, Niklas Lavesson és Dirk Vanhoudt: Operational Demand Forecasting In District Heating Systems Using Ensembles of Online Machine Learning Algorithms, Energy Procedia 116 (2017) 208–216
- [6.] Meteostat meteorológiai adatbázis <https://meteostat.net/en/>
- [7.] A talaj hőmérséklete az év során https://mersz.hu/hivatkozas/m514aea_10_p1#m514aea_10_p1
- [8.] ISOPlus előszigetelt vezetékek hőátbocsátási tényezői https://www.isoplus.hu/leadmin/data/downloads/documents/hungary/manuals/Merev_csovek.pdf
- [9.] HUHA II. Új Iszapégető és Hulladék-hasznosító Erőmű Műszaki Döntés Előkészítő Tanulmány Rev.1. (S-6 Mérnök Kft, koordinátor: Orbán Tibor, 2017. augusztus)
- [10.] Magyar Közlöny 2023. szeptember 29., 139. szám https://l.facebook.com/l.php?u=https%3A%2F%2Fmagyarkozlony.hu%2Fhivatalos-lapok%2FHkemDS0ebvTK0sVaiHOx65132f260b683%2Fdokumentumok%2F4caaf1730eda6046852939889ac5906fb79659ac%2Fletoltes%3Ffbclid%3DIwAR0EgTB07JR0hTayW2_TfLWArZtvmpT_ZbPWwt1catTn0fbkDBB-LGjxPdE&h=AT2SDAMHr38zPJsVNdM8XHcWpvfk8FrkACkqldc-Ok6iEioiOqCzLqJ_Sv5UmmTgMoB_R9-Q1QNWVzQYTCCAP-VYjpw0x_irSRmRw_GZ-p5V8g05TaZk8AMvG6mryyWEFwZLH3g

- [11.] Davy G., Christian J. and D. Vanhoudt: Operational thermal load forecasting in district heating networks using machine learning and expert advice, Energy and Buildings, DOI: 10.1016/j.enbuild.2017.12.042
- [12.] Sklearn MinMaxScaler függvény: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>
- [13.] Scikit-learn könyvtárcsomag: <https://scikit-learn.org/stable/about.html>
- [14.] GridSearch hiperparaméter optimalizáló algoritmus https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
- [15.] Lewis, C. D.: Industrial and Business Forecasting Methods. London: Butterworths, 1982. ISBN 0408005599
- [16.] NumPy random.normal : https://numpy.org/doc/stable/reference/random/generated/numpy.random.normal.html?fbclid=IwAR14HEj1xv6l8xr6YgCZMi-LOdP547xy38Idk8FqdPP1Q8boZ_5X9WHPzgoU
- [17.] Jeremy Jordan: Setting the learning rate of your neural network <https://www.jeremyjordan.me/nn-learning-rate/>

7. MELLÉKLETEK

7.1. Végző hiperparaméterek és jelentésük

7.1.1. HIPERPARAMÉTEREK JELENTÉSE

A hiperparaméterek az egyes algoritmusokra jellemző, működésüket befolyásoló kezdeti paraméterek, amelyeket a futtatás előtt be kell állítani. Ezen értékek beállítása kulcsfontosságú egy megfelelő gépi tanulási módszer működéséhez, mert ezen paraméterek segítségével tudjuk finomhangolni az algoritmusokat a specifikus, általunk alkalmazott adatszetek legjobb megismerésének érdekében. A következőkben az egyes algoritmusokra és hálókra jellemző paraméterek jelentését fogjuk bemutatni a működésük jobb megértésének érdekében. A következő alfejezetekben bemutatott hiperparaméterek jobb megértésének érdekében javasoljuk az **Alkalmazott algoritmusok** fejezetben leírtak párhuzamos tanulmányozását.

7.1.1.1. *Random forest regressor*

- **criterion**: a fákon belüli elágazásokhoz alkalmazott minőségi kritérium
- **max_depth**: az egyes fák maximum mélysége
- **n_estimators**: az alkalmazott fák száma az „erdőben”
- **min_samples_split**: az elágazások létrejöttéhez szükséges minimum adatmennyiség
- **max_features**: A legjobb osztás keresésekor figyelembe veendő jellemzők száma

7.1.1.2. *Support vector regressor*

- **kernel**: az algoritmus „szívének”, a kernelnek (=mag) matematikai alakját megadó paraméter
- **degree**: a kernel egyenlet polinomiális foka
- **gamma**: a kernel együtthatója

- C: szabályozási paraméter, amelynek szabályozási erőssége fordítottan arányos a C paraméter értékével
- epsilon: keresési sugár, amely belül nem számol hibát az algoritmus

7.1.1.3. *K nearest neighbour*

- n_neighbors: a figyelembe vett szomszédok száma
- weights: a pontok súlyozási értéke közelség szempontjából
- algorithm: a számítási algoritmus
- leaf_size: a ball_tree és a kd_tree algoritmusoknak adott adatsomagok mérete
- metric: a pontok közötti távolság számítási módja
- p: a Minkowszky paraméter teljesítmény tényezője, amely 1 vagy 2 értéket vesz fel, majd ennek függvényében az euklideszi, avagy manhattan távolságmutatót alkalmazza.

7.1.1.4. *AdaBoost*

- base_estimator: az alap algoritmus, amelyet fellendít az AdaBoost (jelen esetben K nearest neighbours regressor)
- n_estimators: az alkalmazott becslő egységek száma
- loss: a hibafüggvény fajtája, amely alapján a súlyozások számítása történik
- learning_rate: Az egyes regresszorokra minden egyes boosting-iterációnál alkalmazott súly, nagysága befolyásolja a tanulási gyorsaságot, ám növeli annak esélyét, hogy a keresett globális minimumot nem találja meg az algoritmus.

7.1.1.5. *Multilayer perceptron*

- hidden_layer_sizes: az rétegek számának és a bennük található neuronok számának meghatározása
- activation: a rejtett rétegek aktivációs függvénye
- solver: az optimalizáló technikája
- alpha: súlyozási tényező

- `learning_rate`: a tanulási gyorsaság beállítása, esetünkben adaptív állapotra van állítva, amely a futás közben módosítja az értéket az elért hibacsökkenés csökkenés függvényében
- `batch_size`: az adatokat mekkora csomagméretben adjuk be az algoritmusba
- `max_iter`: a maximum iterációk száma, amelyek során az algoritmus az elejétől kezdve végigfut az egész adatsoron, közben tovább tanulva.
- `early_stopping`: beépített védelmi leállás, ami megakadályozza, hogy háló a túl sok iteráció következtében „overfit” legyen, másnéven nagy pontossággal rátanuljon a tanító adatsorra, ám további adatsorok prediktálására már nem lesz alkalmas.

7.1.1.6. Long short-term memory

Az LSTM esetén a hálóstruktúra módosításával igyekeztünk javítani az előrejelzési pontosságot. A háló rétegeinek felépítése a következő volt az elsőtől az utolsó réteggig sorban:

- 1 dimenziós konvolúció, a konvolúció különböző mintázatok felismerésére és azok kinyerésére alkalmazható matematikai műveletsor
- LSTM: a korábban leírt LSTM, amelyből 512 algoritmus egység található
- Dense (fully connected) layer 128, ahol 128 neuron található, amelyek rendelkeznek összeköttetéssel az összes szomszédos rétegben található neuronnal.
- Dense (fully connected) layer 32
- Dense (fully connected) layer 1

7.1.2. OPTIMALIZÁLÁSI EREDMÉNYEK

A **Hiperparaméterek optimalizáció megvalósítása** c. fejezetben kifejtett optimalizációs logika alapján elvégeztük a szükséges modellezéskét, amelyek az alább prezentált algoritmusok eredményeihez vezettek. Az eredményeket a **9. táblázatban** foglaltuk össze. A

táblázatban pár paraméter nem került megjelenítésre, mivel azok konstans értékkel rendelkeztek minden futtatás esetén. Ezen értékek a következők voltak:

A „random_state” értékek minden esetben 42-re voltak állítva, így biztosítva a modellek reprodukálhatóságát. Az MLP esetén:

- solver: adam
- learning_rate: adaptive
- batch_size: 32
- max_iter: 1000
- early_stopping: True

9. táblázat: A hiperparaméter optimalizáció eredményei

	Kelenföld				Csepel				Kispest			
	Tavaszi	Nyári	Ősz	Téli	Tavaszi	Nyári	Ősz	Téli	Tavaszi	Nyári	Ősz	Téli
critierion	squared_error	squared_error	squared_error	squared_error	squared_error	squared_error	squared_error	squared_error	squared_error	squared_error	squared_error	squared_error
max_depth	40	30	20	30	20	40	20	40	30	30	30	20
max_features	sqrt	sqrt	sqrt	sqrt	sqrt	sqrt	sqrt	sqrt	sqrt	sqrt	sqrt	sqrt
min_samples_split	5	5	5	5	5	4	7	4	5	4	5	7
n_estimators	1000	250	250	250	1250	500	1250	1250	250	250	500	125
C	10	10	10	10	10	1	10	10	1	10	10	10
degree	2	3	2	2	2	2	2	2	2	3	4	2
epsilon	0.1	0.001	0.001	0.001	0.01	0.1	0.1	0.001	0.1	0.1	0.001	0.1
gamma	0.01	0.01	0.1	0.1	0.1	0.1	0.1	0.1	1	1	1	1
kernel	rbf	poly	poly	rbf	rbf	rbf	rbf	rbf	poly	poly	poly	poly
learning_rate	0.1	0.01	0.001	0.001	0.01	0.001	0.001	0.1	0.001	0.1	0.001	0.1
loss	square	exponential	exponential	square	linear	exponential	square	linear	square	linear	exponential	square
n_estimators	20	50	100	100	50	100	50	100	50	100	20	20
algorithm	ball_tree	ball_tree	ball_tree	ball_tree	ball_tree	ball_tree	ball_tree	ball_tree	ball_tree	brute	ball_tree	ball_tree
leaf_size	7	7	7	7	7	7	7	7	5	5	5	5
metric	manhattan	manhattan	manhattan	manhattan	manhattan	manhattan	manhattan	manhattan	euclidian	cosine	manhattan	manhattan
n_neighbors	11	11	11	11	11	11	11	3	3	3	3	3
weights	uniform	uniform	uniform	distance	uniform	uniform	uniform	distance	uniform	distance	distance	uniform
activation	relu	tanh	relu	relu	relu	tanh	tanh	tanh	relu	relu	tanh	relu
alpha	0.001	0.001	0.0001	0.0001	0.01	0.01	0.01	0.001	0.001	0.01	0.0001	0.0001
batch_size	32	32	32	32	32	32	32	32	32	32	32	32
early_stopping	true	true	true	true	true	true	true	true	true	true	true	true
hidden_layer_sizes	[300, 300, 300]	[500, 500, 500]	[300, 400, 400, 300]	[400, 400, 400]	[300, 400, 400, 300]	[400, 400, 400]	[500, 500, 500]	[300, 300, 300]	[500, 500, 500]	[500, 500, 500]	[300, 300, 300]	[300, 400, 400, 300]