

Szakdolgozat

2023.

Tar Dániel Attila
Budapesti Gazdasági Egyetem Külkereskedelmi Kar
Nemzetközi Gazdálkodás Mesterképzés

ETHEREUM MÚLTJA, JELENE ÉS JÖVŐJE

THE PAST, PRESENT, AND FUTURE OF ETHEREUM

Belső konzulens: Dr. Budai László

Kézirat lezárása: 2023. április 30.

Revízió: 2023. augusztus 26.

Tartalomjegyzék

1	Mélyinterjú	3
2	Adataalapú kutatás	8
2.1	Adatok és statisztikák	8
2.1.1	Bázis szekció	8
2.1.2	Szekció 1 – „Bitcoin, S&P500, NASDAQ”	11
2.1.3	Szekció 2 – „Hírek, Kommentek és Bálnák”	17
2.1.4	Szekció 3 – „Technikai indikátorok”	30
2.2	Árfolyam-előrejelzés	33
3	Összegzés	45

Ábrajegyzék

1. ábra: ETH-ra vonatkozó napi adatok	9
2. ábra: Az ETH záró árfolyama (\$) és kereskedelmi volumene ETH/USDT (\$)	9
3. ábra: ETH napi záró árfolyama (\$) és ETH/USDT tranzakciók száma (db)	10
4. ábra: Az ether napi záró árfolyama vs. bitcoin napi záróárfolyama (\$)	11
5. ábra: BTC/USDT napi kereskedelmi volumene (\$), napi tranzakcióinak száma vs. ETH/USDT napi tranzakcióinak száma	12
6. ábra: Az ETH napi záró árfolyama (\$) és a FED effektív kamatlába (%)	12
7. ábra: ETH - S&P500 - NASDAQ100	13
8. ábra: ETH és S&P 500 ingatlan szektor indexe	14
9. ábra: Arany, Ezüst (uncia) napi árfolyama (\$) és ETH napi záró árfolyama (\$)	14
10. ábra: USA 5-10 éves inflációs várakozások vs. ether árfolyama (\$)	15
11. ábra: USA kincstárjegyek kamatának különbözete vs. ETH árfolyama	15
12. ábra: Korrelációs mátrix hő térképe - Szekció 1 adatai	16
13. ábra: Szórásdiagramm kapcsolatelemzés végett	17
14. ábra: Nyelvi modellek szentiment analízise példán keresztül	18
15. ábra: Hírek tábla felépítése	19
16. ábra: Ethereum hírek szentiment értéke a vizsgált időszakban	20
17. ábra: Hírek "hangulata" 2021 elején	20
18. ábra: Hírek „hangulata” 2022 elején	21
19. ábra: Top 5 subreddit oldal – kommentek száma és pontok összege	21
20. ábra: Reddit kommentek - Ethereum, ETH és ether kulcsszóval	22
21. ábra: Twitter tweeteket tartalmazó tábla felépítése	23
22. ábra: Top felhasználók és lokációk tweet szám alapján (db)	23
23. ábra: Ethereum, ETH, ether kulcsszavakat tartalmazó Twitter tweetek szentimentje	24
24. ábra: Google Keresési Trend - Ethereum témakörben	24
25. ábra: Bálnák tranzakcióinak száma (db) és értéke (ETH)	25
26. ábra: Bálnák tranzakcióinak értéke 2022-11-01 és 2022-11-20 között (ETH)	26
27. ábra: Szekció 2 adatainak korrelációs mátrixa	26
28. ábra: Ethereum Search Trend - Kereszt-korreláció 95%-os konfidencia-intervallummal	27
29. ábra: Bálnák száma (unique_whales) - Kereszt-korreláció 95%-os konfidencia-intervallummal	28

30. ábra: Bálnák tranzakcióinak száma (whales_tx_number) - Kereszt-korreláció 95%-os konfidencia-intervallummal	28
31. ábra: Hírek szentimenje (news_sentiment) - Kereszt-korreláció 95%-os konfidencia-intervallummal	29
32. ábra: MACD mutató 2023	30
33. ábra: 26 és 12 napos mozgóátlagok 2023-ban	31
34. ábra: Money Flow index	32
35. ábra: Bollinger szalagok 2023	33
36. ábra: Naiv előrejelzés példa – Validációs adathalmazra	34
37. ábra: Szezonális tesztelése	35
38. ábra: Szezonális és Trend dekompozíciója (éves periódussal)	35
39. ábra: Holt-Winters módszer – Validációs adathalmazra	36
40. ábra: Arima modell előrejelzés – Validációs adathalmazra	37
41. ábra: VAR modell előrejelzés – Validációs adathalmazra	38
42. ábra: XGBRegressor modell előrejelzése - Validációs adathalmazra	40
43. ábra: LSTM tanulási eredménye	41
44. ábra: LSTM modell előrejelzése - Validációs adathalmazra	42
45. ábra: Tanulási/Validálás adatok felbontása	42
46. ábra: Előrejelzések átlagos négyzetes hibájának gyöke (RMSE)	43
47. ábra: Előrejelzések átlagos abszolút százalékos eltérése (MAPE)	43
48. ábra: LSTM predikció az elkövetkező 14 napra	44

Előszó

A szakdolgozatom fő témája az Ethereum hálózata és ether nevezetű kriptovalutája volt. Az elméleti rész során részletesen bemutattam a kriptovaluták történetét, működési elvét, technológiáját. Megismertettem az olvasóval a kriptopénzek „ősanyját”, a Bitcoin-t. Történeti áttekintés mellett kitértem a POW rendszerére, környezetvédelmi hatásaira és illegális tevékenységekhez fűződő kapcsolataira. Mindezek után áttértem a dolgozatom fő kérdéskörére, az Ethereum-ra. Kezdve az elődökkel végig lépkedtünk az Ethereum idővonalán, megismerhettük Vitalik Buterint, a hálózat kezdetleges működését és nehézségeit. Részleteiben bemutattam egy tranzakció működését, mind a régi, mind az új POS hálózaton. Emellett létrehoztam az Ethereum teszt hálózatán egy okoszerződést, hogy bepillantást nyerhessünk az okoszerződések létrehozásába és működésébe. Az elméleti rész végén a hálózat jövőjéről és fejlesztési szakaszairól értekeztem. Jelen dokumentum nem tartalmazza az elméleti részt, csakis az izgalmasabb kutatási részt. Célom az volt, hogy adatalapon vizsgáljam az ether napi záróárfolyamának alakulását. Rengeteg adatot szereztem meg különböző forrásokon keresztül, a nyilvános GitHub (https://github.com/flamenco18/ethereum_analysis) mappám viszont nem tartalmazza az adatokat (csak a módszertanokat és megszerzés útját). A kutatásom egyik legérdekesebb szakasza az ether árfolyamára vonatkozó előrejelzések. Statisztikai, géptanulási modellek és neurális hálók segítségével készítettem a predikciókat. Fontos kiemelni azt, hogy sokat foglalkoztam géptanulási modellekkel, viszont idősoros előrejelzéssel kevesebbet. Így a kutatásom iránya és minősége még nem tökéletes, ellenben egy mélyebb kutatás alapjának vehető. Remélem a dolgozatom végére sikerül felkeltenem az olvasó érdeklődését az Ethereum-mal kapcsolatban. Vágjunk bele!

Kutatás

A kutatásom során mélyinterjút készítettem magyarországi blokklánc szakértőkkel, és adatalapú kutatást végeztem az Ethereum hálózat kriptovalutájára, az ether-re. Az elemzés során különböző hipotézisekre kerestem a választ.

Hipotéziseim:

- 1. DeFi cégek okoszerződéseken keresztül képesek felvenni a versenyt a kereskedelmi bankok által nyújtott pénzügyi szolgáltatásokkal, mint például a hitelnyújtással.**
- 2. Az Ethereum-mal kapcsolatos hírek, Google keresési trendek, Reddit kommentek és Twitter tweetek szignifikáns kapcsolatban állnak az ether napi záró árfolyamával.**
- 3. Neurális háló segítségével pontosabban előrejelezhetjük az ether jövőbeli árfolyamát, mint konvencionális statisztikai és gépi tanulási modellekkel.**

Az elméleti rész során bepillantást nyerhettünk az okoszerződések, DAO-k és DeFi-k dimenziójába. Mélyinterjú során az okoszerződések pénzügyi felhasználására, DeFi-k népszerűségére és szolgáltatásaira voltam kíváncsi. Az ether árfolyama az elmúlt évek során rendkívül volatilis volt. Kíváncsi voltam arra, hogy a bitcoin, az S&P500, NASDAQ index, az arany és ezüst árfolyama hogyan befolyásolta az évek során az ether árfolyamát. Mindezek mellett számításba vettem a pozitív-negatív hírek, cikkek, kommentek árfolyamalakító hatását, a nagybefektetők szerepét és azok "bevásárlásait". Végül előrejelzéseket készítettem az idősorok historikus értékei alapján statisztikai, gépi tanulási és neurális hálóval rendelkező modellekkel.

1 Mélyinterjú

Kutatásom kezdetekor felmerült bennem egy mélyinterjú elkészítése szakértőkkel, hogy az elméleti részben leírt technológiai vívmányok és rendszerek mennyire valósulnak meg a valós életben és Magyarországon. Így kvalitatív kutatásom eszközüül a mélyinterjút választottam, több külföldi és magyarországi szakértőt próbáltam elérni a témakörrel kapcsolatban. Három magyarországi kriptó és blokklánc szakértővel tudtam felvenni a kapcsolatot és elkészíteni a interjút. Az interjúalanyaimról röviden - Vidákovics Attila a hazai fintech szektor elismert blokklánc szakértője, aki meghatározó szerepet tölt be a decentralizált pénzügyi rendszerek hazai bevezetésében. Személyes missziójának tekinti a kriptopiacok, decentralizált pénzügyi rendszerek és blokkláncok működési elvének és logikájának megismertetését. Emellett megkérdeztem projektcsapatának tagjait - Horváth Tamást és Péntek Józsefet blokklánc szakértőket. Csapata elsőként fejlesztett működő EVM alapú blokkláncot idehaza. Mindhárom szakértő évek óta foglalkozik a témával, izgalmas és érdekes interjú született meg.

Első lépésben arra voltam kíváncsi, hogy milyen a központi bankok, mint a FED vagy ECB hozzáállása az Ethereum rendszeréhez. Megkérdezett szakértőim egyetértettek abban, hogy nagyon vegyes a hozzáállás, a kérdésben véleményt megfogalmazók általában 1-1 esetet emelnek ki, szerintük érdemesebb az egész történelmét végignézni a kriptó-anarchizmusnak és a Fednek. A jövőre vonatkozóan nehéz bármit is állítani, jelen helyzetben a Fed a decentralizált világgal szemben áll. Az Ethereum hálózat egyik nagy hátrány az, hogy a validátorok többsége az Amerikai Egyesült Államokban található meg. Sokan rebesgetik azt, hogy valamiféle törvényjavaslattal kontrolálnák a validátorokat, viszont ennek életbelépése esetén is a hálózat változatlanul és biztonságosan fennmaradna az USA-n kívüli másik 200 ezer validátor által. Az ilyen típusú korlátozásra vonatkozó hírek rebesgetése viszont a centralizált világ félelmét fejezi ki a decentralizálttal szemben. Látják az erőt, a növekedést és az iparágak viszonyulását a decentralizált megoldások felé. Véleményük szerint számukra még rémisztőbb lehet az, hogy rengeteg hagyományos, nagyvállalat indít valóban DeFi projekteket harmadik szereplő/közvetítő bank nélkül. Szerintük a központi bankok és kifejezetten a Fed nagy félelme a pénzügyi világ feletti kontroll csökkenése a kriptó világ és decentralizált technológia végett. Üzleti szempontból viszont látható, hogy a szereplők nyitnak a technológia felé még úgy is, hogy az borzasztóan gyerekcipőben jár. Az ellenérdekek ütközését a központi bankok CBDC – digitális jegybankpénz törekvései fokozhatják még a jövőben. Ezt követően a világ számítógépeként emlegetett Ethereum rendszerét és annak decentralizált funkcionalitását kérdeztem meg. Vidákovics Attila szerint fontos megemlíteni, hogy nem minden decentralizált,

ami egy decentralizált platformon fut és nincs is egyelőre olyan mérőszám, ami meghatározná ezt. Meglátásuk szerint sokan megjelenítik azt, hogy decentralizáltak, közben ideiglenesen vagy teljesen centralizáltak. Kevés hányada épül ténylegesen decentralizált alapokra, a társadalom még nincs kész arra, hogy kihasználja előnyeit. A világ még tanuló üzemmódban van, hiszen teljesen más logika és megközelítés szükséges a struktúra értelemezéséhez. Az Ethereum rendszerére visszatérve az alkotók célja a kezdetektől fogva egy decentralizált „világ számítógép” megalakítása volt. Az Ethereum rendszerére vonatkozó fejlesztési kihívások jelen vannak, viszont a folyamatos innovációknak köszönhetően néhány év múlva az okosszerződések írása, komplex programok és szkriptek írása jelentősen egyszerűsödni fog. Az Ethereum-nak habár vannak komoly kihívói, dominanciája megingathatatlan a Turing-teljes hálózatok körében. Azt, hogy a jövőben mikor éadjük el, hogy teljesen hétköznapi lesz egy okosszerződés, egy decentralizált szolgáltatás írása az Ethereum hálózatán jelenleg nagyon nehéz megmondani. Egyre több megoldás van, aminek tényleg a blokkláncon van a helye és rendkívül praktikusán használható. A továbbiakban megkérdeztem szakértőimet arról az állításról, amit az Ethereum rendszer kitalálója, Vitalik Buterin írt ki Twitterre miszerint: A jövőben már telefonunkon is képesek leszünk tranzakciót titkosítani, stakelni, a hálózat egy csomópontjaként működni. Vitalik szerint 10 éven belül érhetjük el mindezt, a szakértőimet kérdeztem arról, hogy ezt az állapotot hamarabb is elérhetjük, vagy csak később, valóság-e a várakozások. A szakértők pontosan nem tudják megmondani, hogy mikorra érhetjük el ezt a szituációt, hiszen rengeteg tényezőtől függ. Sok megoldás létezik már, viszont mélységig megérteni a technológia bonyolultságát rendkívül komplikált. Vidákovics Attila meglátása szerint 2-3 éves időtávokban óriási fejlődések vannak a blokklánctechnológiában, így nem lenne meglepődve, ha már 3 év múlva a telefonunkon keresztül érvényesítenénk blokkokat a hálózatban.

Miután a központi bankokról és a rendszer jövőjéről kérdeztem a szakértőket, áttértünk az okosszerződések, DeFi-k vonalára. Megkérdeztem őket, hogy az Ethereum 10 év múlva mennyire lesz elterjedt világszerte és képes lesz-e felvenni a versenyt a kereskedelmi bankok által nyújtott szolgáltatásokkal (pl. hitelnújtás DeFi-n keresztül). A szakértők véleménye szerint nem is kell 10 év ahhoz, hogy versenyképesé váljanak a decentralizált pénzügyi szolgáltatások és megoldások. Egyes régiókban hamarabb befog következni mindez, hiszen 1-2 milliárd ember olyan környezetben él, ahol nincs kialakított, komplex pénzügyi szolgáltatás, de ahogy a globalizációs folyamatok fokozódnak és az internet egyre jobban terjed, úgy egyre könnyebben tudnak majd munkát vállalni, pénzt szerezni. A gondolatmenetet folytatva kialakul egy tömeges igény a pénzügyi szolgáltatások irányába, amely egy rövid idő alatt kialakítható

pénzügyi környezet, mint a kriptó alapú fizetési rendszer, rendkívül kedvező. Szakértők álláspontja alapján 1-2 éven belül lesznek olyan példák, ahol jól működő kriptó alapú pénzügyi rendszereket hoznak létre az emberek. Érdekességként megjegyeztek egy kriptó gurut Joe Nakamotot, aki járja a világot és próbál kriptoval fizetni mindenhol, bankkártya és minden egyéb konvencionális pénzeszköz nélkül. A továbbiakban az okoszerződések kapcsán kérdeztem az interjúalanyaimat. Megkérdeztem azt, hogy szerintük mely szférákban tudnak domináns szereplőként fellépni az Ethereum alapú okoszerződések, előfordulhat-e akár egy éttermi fizetés lebonyolítása harmadik fél nélkül. A szakértők egyetértettek abban, hogy semmilyen akadálya nincs annak, ha egy vállalkozás egy lokálisan létrehozott okoszerződésen keresztül szeretné fogadni a rendelés után kifizetett összegeket. Szerintük a leggyorsabban abban a szférában fog elterjedni, ahol viszonylag egyszerűen és gyorsan megérthető logikával tudnak szolgáltatás kínálni és nagyon drágán működik a jelenlegi pénzrendszer, tipikusan ilyen például a hitelnyújtás. Érdekességként kiemelték, hogy Svédországban a földhivatal bloklánc alapú rendszeren fut. Ha egy okoszerződésbe tulajdonjogot tudunk bekötni, akkor a hitelfelvétel is megoldható lesz. Számítalan olyan megoldás van, hogy ha mondjuk szeretnénk hitelt felvenni, mert likviditási problémánk van, és az általunk befektetett pl. bitcoin-t vagy ether-t nem szeretnénk eladni, akkor azt fedezetként felhasználva gyorsan és elképesztően olcsón hitelt kaphatunk. Ha törlesztem az okoszerződés felé a tartozásomat, akkor felszabadulnak a fedezetként letett kriptovalutáim, ha nem, akkor az okoszerződés likvidálja a fedezetet, amit általában vagy a DeFi fejlesztőinek támogatására, vagy a következő hitelnyújtásra használnak fel. Egyre nagyobb felhasználótábora van a szolgáltatásnak, folyamatosan több tízmillió dollár vagy van olyan, ahol 100 millió dollár feletti hitelállomány fut. Az elmondottak alapján **az első hipotézisem bebizonyosodott, miszerint a DeFi cégek okoszerződéseken keresztül képesek felvenni a versenyt a kereskedelmi bankok által nyújtott pénzügyi szolgáltatásokkal, mint például a hitelnyújtással.** Vidákovics Attila kiemelte, hogy azon felhasználási formák tudnak elterjedni a legjobban, amely könnyen érthető mindenki számára. Elmondásai alapján öt év múlva már Magyarországon is elterjedtebb lesz a privátkulcsok használata és értelmezése.

Az elmúlt években számos okoszerződést, DeFi-t meghackeltek, például Ronin, Sushi Swap, Polygon hálózatát stb. Megkérdeztem a szakértőket, hogy szerintük milyen hatással lehetnek a hackertámadások az Ethereum hálózatára, okoszerződéseire és az ether árfolyamára akkor is, ha az közvetlenül nem az Ethereum rendszerében történt, hanem a kriptovaluták világában. Véleményük szerint a piac egyre immunisabb a hackertámadásokra, mondjuk, ha egy Ethereum-ot hackelnének meg, az elég komolyan érintené az egész piacot. Fontos

megjegyezni, hogy a támadások szignifikáns része az okosszerződések ellen ment végbe. Kiemelték a DAO meghackelését 2016-ban, amiről az elméleti részben is írtam. A jövőben mindig lesznek olyan okosszerződések, amikben vannak kiskapuk és kijátszható hibák. Megjegyezték a SushiSwap hackelését, amely során több mint 3 millió dollár értékű eszközt loptak el. Habár a token árfolyama a hír hallatán 5-10%-ot zuhant, hamar visszanyerte eredeti értékét. A szakértők elmondták, hogy a transzparencia miatt több hírt hallunk az okosszerződések meghackeléséről, viszont arról nagyon keveset, hogy ténylegesen hányszor hackelik meg a kereskedelmi bankokat éves szinten, hiszen ez negatívan érintené őket. Hangsúlyozták viszont azt, hogy a kereskedelmi bankok jól állnak az ellopott pénzükért, míg okosszerződések esetén elbuknánk az egészet. Nyilván mindkét oldalnak megvan az előnye és hátránya, nem biztos, hogy mindenkinek való a kripto és decentralizált világ – jegyezték meg.

A következőkben az okosszerződések magyarországi elterjedtségéről kérdeztem szakértőimet. Statisztikákat nem tudtak mondani, viszont elmondásuk alapján egyre elterjedtebbek cégek körében az okosszerződések megismerése, használata. Kiemelték magyarországi cégek közül az OTP-t, akik már saját blokkláncfejlesztő csapattal is rendelkeznek, az MVM-t, Vodafone-t és Mol-t is, akik pedig olyan megoldásokkal foglalkoznak, amelyek okosszerződéseken futhatnak. Utolsó kérdésem során az interjúalanyaim elé az okosszerződések általános jövőjéről tettem fel kérdést. Elmondásuk szerint egyik oldalról a technológia fejlődésével egyre egyszerűbb lesz különböző üzleti logikát okosszerződésekbe foglalni és ezzel együtt párhuzamosan egyre több dologra fogják használni. Eltávolodva a kripto eszközöktől és árfolyamspekulációktól kiemelték egyik fejlesztésüket, amely naponta 5-10 ezer tranzakciót beküld, az konkrétan egy dokumentum hitelesítés. Az egyik legnagyobb magyar társasházkezelő minden dokumentumot, amit bárhova is elküld, azt blokkláncon hitelesít. Ehhez egy okosszerződést írtak, amely lekezeli a folyamatot. Ilyen jellegű megoldások hétköznapiak lesznek, mert számtalan előnye van, amit ki lehet használni. Hibrid rendszerek fognak megjelenni, adatbázisok lesznek okosszerződésekkel összekötve, bizonyos műveleteket blokkláncon fognak futtatni, amelyek nem feltétlenül lesznek decentralizáltak – taglalták.

Összegezve a kvalitatív kutatásomat egy nagyon érdekes és izgalmas interjút hallhattam a szakértőktől, kezdve a központi bankok hozzáállásáról, az Ethereum és okosszerződések jövőjén át, a magyarországi felhasználásról. Kiemelték a centralizált pénzügyi világ ellentétét a decentralizáltéval, beszéltek a decentralizáció fogalmáról és annak értelmezéséről. Kihangsúlyozták, hogy jelenleg rengeteg ember él kialakított pénzügyi rendszer nélkül, ahol praktikus lenne a kripto alapú fizetési rendszer létrehozása. Az első hipotézisemet sikerült

bebizonyítani, amikor is a szakértők az okoszerződések, DeFi-kről és hitelnyújtásról beszéltek. Az interjú során kitértünk az elmúlt években történt hackelésekre is és azok hatására mind az Ethereum, mind a kriptovilág kapcsán. Magyar vonatkozásban az OTP-ről, MVM-ről, Mol-ról és Vodafone-ról hallhattam okoszerződések/blokklánc technológia iránti érdeklődést. Végül, de nem utolsó sorban az okoszerződések általános jövőjéről kérdeztem szakértőimet. Térjünk át a kutatásom második felére, amikor is adatalapú kutatást végeztem az Ethereum rendszer kriptovalutájával (ether vagy ETH) kapcsolatban!

2 Adatalapú kutatás

A kutatásom második felében egy teljeskörű adatalapú elemzést készítettem az ether árfolyamára és meghatározó tényezőire. Az elemzést Jupyter Notebookban írtam meg, Python objektumorientált programnyelven. Minden kód saját íromány, megtalálható a nyilvános GitHub Mappámban ("Repository") ([ethereum analysis](#)). Az általam írt kódokat is megjeleníttem a mondatok után hiperhivatkozásban, amelyek közvetlenül a megfelelő GitHub oldalra dobnak be. Mielőtt belevágnánk az adatokba és statisztikákba, szeretném hangsúlyozni, hogy egyes historikus adatok, mint például a FED Real Estate Indexének értéke vagy az S&P500 alakulása elérhető nyilvánosan és ingyenes a TradingView.com weboldalon. Az elemzések során GDPR által védett adatokkal is foglalkoztam (Reddit és Twitter kommentek), így ezek a GitHub mappámban nem találhatóak meg, csakis aggregált módon. Először részletesen megismerjük az összes adatot, változót, statisztikát, majd speciálisan áttérünk a szentiment analízis részre, mely során az ether árfolyamára ható kripto híreket, Reddit kommenteket, Twitter tweeteket elemezzük. Ezután az idősoros előrejelző modellekre evezünk át, ahol a legalapabb naiv előrejelzést hasonlítjuk össze gépi tanulási és neurális hálós modellekkel. A kutatásom során külön kitértem a koronavírus és az orosz-ukrán válság hatásaira is. Vágjunk bele!

2.1 Adatok és statisztikák

A kutatásom legelején **négy nagyobb szekcióra** bontottam a megszerzendő adatokat: **Bázis** adatok, **Szekció 1**, **Szekció 2** és **Szekció 3**. A fejezet során minden egyes szekciót részletesen átvesszünk, milyen adatokat sikerült megszerezniem, milyen módon, milyen a strukturáltságuk, és milyen értékeket mutatnak a vizsgált időszakban. A vizsgált időszak *2017. szeptember elsejétől egészen 2023. március hetedikéig* tart. Egyes adatok napi, más adatok havi bontásban találhatóak meg. Az első és egyben legfontosabb szekció a Bázis szekció, amely az ETH árfolyamára tér ki.

2.1.1 Bázis szekció

A Bázis szekció során az ether árfolyamát szereztem meg a világ egyik legnagyobb kriptotőzsdéje által kiadott API-n keresztül, Binance API. ([binance eth data.ipynb](#)) Az adatok megszerzése során az ETH/USDT kulcspárt választottam, hiszen az USDT az egyik legrégebb és legjobban ismert stabil coin (rögzítetten mindig egy dollárt ér az értéke). Az ether napi nyitó, záró, legalacsonyabb és legmagasabb árfolyamára vonatkozó adatok általánosak, viszont a volumen és tranzakciók száma ETH/USDT kulcspár alapú. Így például az ETH/BUSD pár

volumen és tranzakcióinak száma kiesik, viszont ez a jelen kutatás során nem releváns, hiszen az általános napi árfolyam fontosabb a kutatás számára. A bázis adatok felépítése az alábbi:

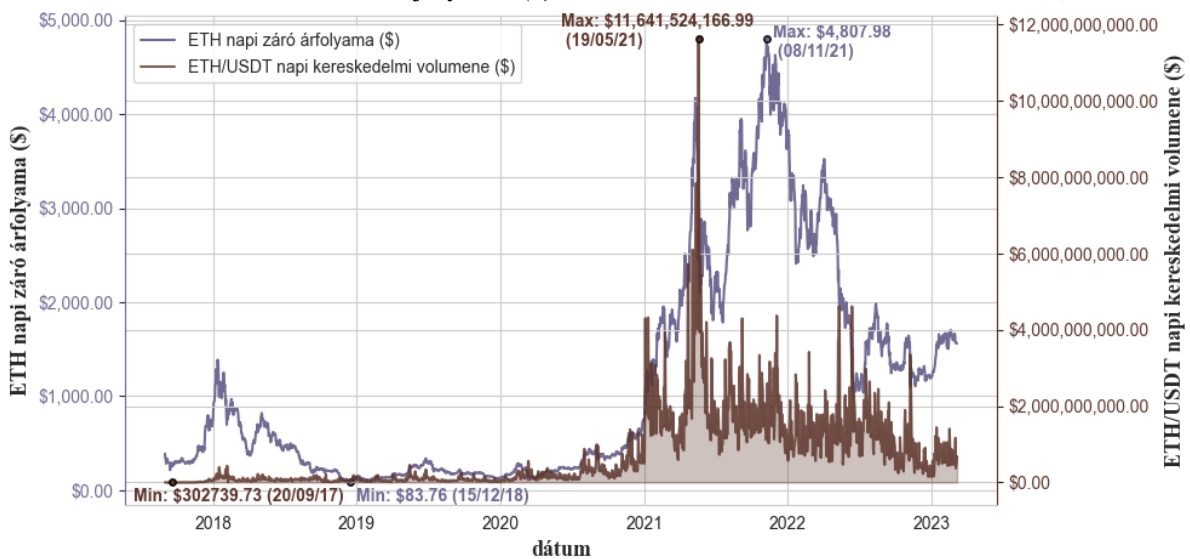
1. ábra: ETH-ra vonatkozó napi adatok

date	eth_open	eth_close	eth_low	eth_high	eth_volume	eth_num_trades	eth_qav	eth_taker_base_vol	eth_taker_quote_vol
2017-09-01	386.44	388.46	383.18	394.39	4198.68434	4124.0	1.635228e+06	1564.05531	6.096309e+05
2017-09-02	388.58	343.14	320.08	394.32	8240.32724	8217.0	2.904008e+06	2460.66776	8.684873e+05
2017-09-03	343.56	341.77	320.08	367.28	5253.62211	6153.0	1.801703e+06	2240.91753	7.679383e+05
2017-09-04	339.29	283.77	267.94	344.55	6737.81061	7403.0	2.002446e+06	2954.16805	8.770111e+05
2017-09-05	286.81	312.09	235.55	321.52	8344.41924	8999.0	2.397491e+06	4312.97282	1.251959e+06

Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Ezek közül viszont néhány adatot, mint például open, low, high és taker_base részeket nem vizsgáltam, csakis az ether napi záró árfolyamát (**eth_close**), napi tranzakciókat (**eth_num_trades**) és azok volumenét (**eth_volume** (etherben), **eth_qav** (\$)). Összesen 2014 napnyi adatot sikerült letölteni egészen a vizsgált időszak elejéig.

2. ábra: Az ETH záró árfolyama (\$) és kereskedelmi volumene ETH/USDT (\$)

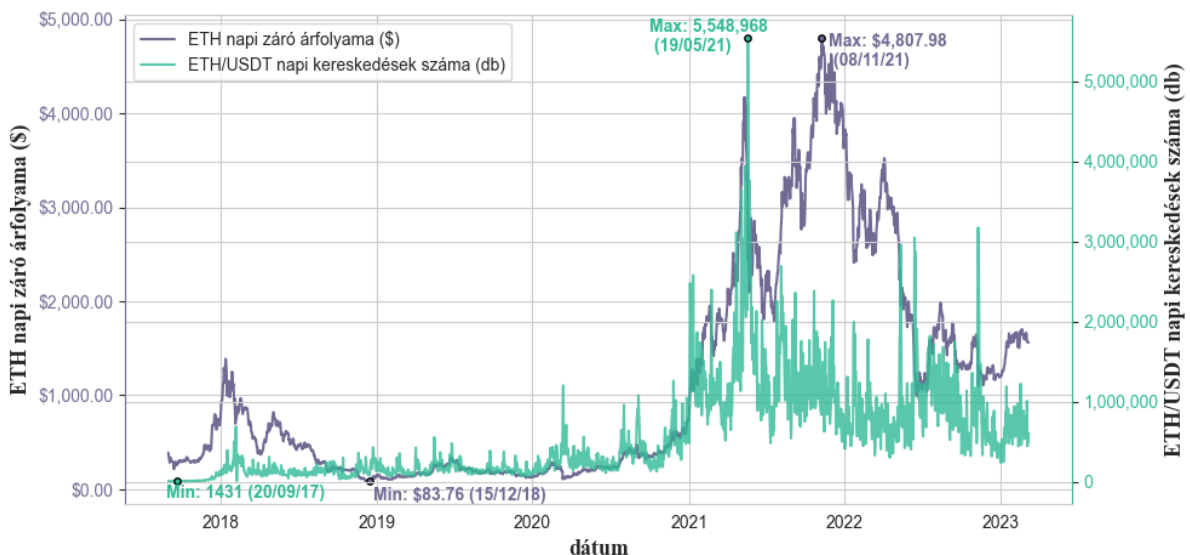


Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Ahogy az a 18. ábrán látható, az ETH árfolyama a vizsgált periódusban exponenciálisan növekedett, rendkívül volatilisen. A helyi minimum-ot 2018. december 15-én érte el (\$83.76), a helyi maximumot pedig 2021. november 8-án, amikor is elérte a \$4807 dollárt. Ez három év alatt közel 57-szeres ár növekedést jelentett, amely elképzelhetetlenül magas érdeklődést és ismertséget váltott ki az Ethereum irányába. Az ábrán látható is az exponenciális trend egészen 2020 végétől kezdődően 2022 elejéig, az orosz-ukrán válság kitöréséig. Ezt követően pedig pont az ellentettjét, egy exponenciális csökkenést láthatunk, az ether árfolyama egészen ezer

dollárig csökkent. Jelenleg az árfolyam „oldalaz”, kisebb-nagyobb növekedést tapasztalva. Az ETH/USDT pár kereskedelmi volumenét arany színnel láthatjuk a grafikonon, míg a minimum napi kereskedelmi volumen \$302 ezer dollár, addig a maximum 11 milliárd dollár feletti a vizsgált időszakban. Ebben az időpontban megannyi hír jelent meg az Ethereum potenciáljáról, teszhálózatának fejlesztéséről és jövőbeli kilátásairól. 2020 és 21 között láthatunk egy volumen növekedést, amikor is az árfolyam még alacsonyabb volt, egészen 2021 közepéig. Ezután a magasabb árfolyam hatására csökkent a volumen, valószínűsíthető, hogy sokan kivártak vagy ódzkodtak a magas ártól. 2022 közepétől viszonylag megnövekedett a tranzakciók volumene az alacsony árak folytán. Ha a kereskedések számát tekintjük meg (3.ábra), akkor bebizonyosodik az előbbieken leírt gondolatok, miszerint a tranzakciók és ezzel a kereskedelmi volumen mértéke drasztikusan megnövekedett 2021 elejére/közepére.

3. ábra: ETH napi záró árfolyama (\$) és ETH/USDT tranzakciók száma (db)



Forrás: Saját készítés ([statistics of training dataset.ipynb](https://statistics.of.training.dataset.ipynb))

Az ábra alapján a FOMO (Fear of Missing Out) effektust is megfigyelhetjük, amikor is 2021. május 21-én rengeteg tranzakció ment végbe annak reményében, hogy az árfolyam exponenciális növekedése fennmarad. Összefüggést vizsgálva korrelációs elemzést végeztem az idősorok értékeire. Fontos megjegyezni, hogy a három idősorot először is *stationary*-vé (statisztikailag állandóvá, trend és szezonáltság nélkülivé) alakítottam. Így csakis a három változó közötti értékek kapcsolatát vizsgáltam. Emellett fontos azt is megjegyezni, hogy ezek **dependens változók**, hiszen az árfolyamot a tranzakciók száma és volumene is meghatározza, így nem érdemes korrelációt vizsgálnunk. Ezzel a Bázis adatok végére értünk, térjünk át az izgalmasabb Szekció 1-re, ahol a bitcoin, S&P500 és NASDAQ indexek mellett amerikai inflációs várakozásokkal, arany és ezüst árfolyammal is foglalkozunk.

2.1.2 Szekció 1 – „Bitcoin, S&P500, NASDAQ”

Munkálataim során az ETH árfolyamára ható independens árfolyamok, indexek vizsgálata közben született meg az első nagyobb ”adatszoport”, adatszekció, amely összefoglalható, és általános képet ad a világ gazdasági helyzetéről, ciklusairól. Bentről kifelé indulva mutatom meg azon adatsorokat, amelyek az ether-re gyakorolt lényegi hatását az alfejezet végén fejtem ki. Az adatok vizualizálása elengedhetetlen azért, hogy pontosabb képet kapjunk trendjükről, mintázatukról és ether-rel az való kapcsolatukról. Az első valuta, amely eszembe jutott, mint befolyásoló tényező, az nem más, mint a bitcoin árfolyama. Az adatokat kód segítségével (hasonlóan az ether adataihoz) Binance API-n keresztül szereztem meg. ([binance btc data.ipynb](#)) Az alábbi ábrán látható egy nagyon érdekes trend:

4. ábra: Az ether napi záró árfolyama vs. bitcoin napi záróárfolyama (\$)

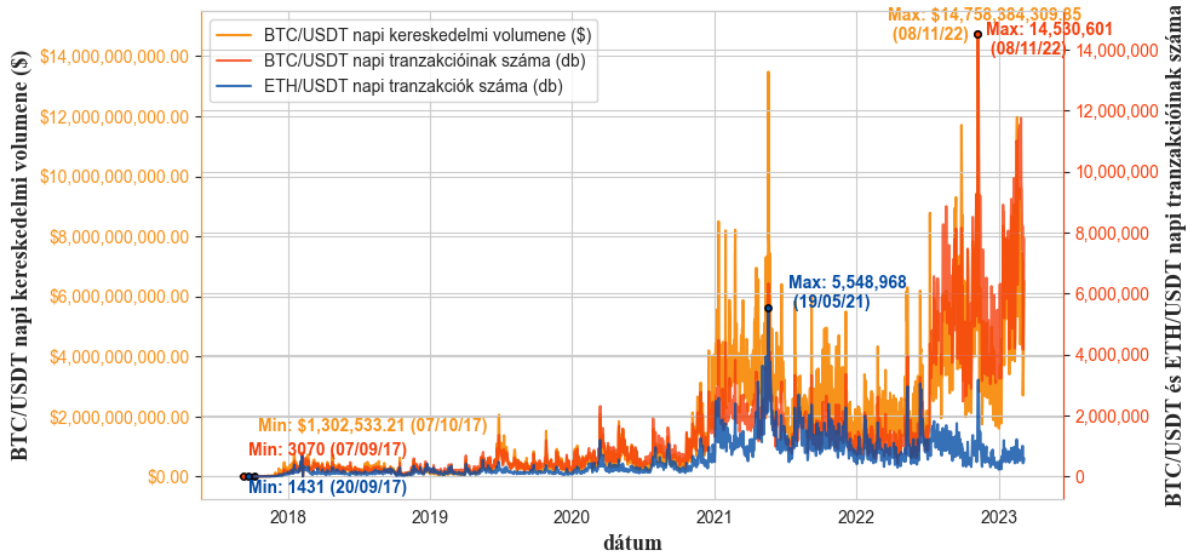


Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Közel azonos trendet és vonaldiagrammot figyelhetünk meg, a bitcoin helyi minimumának időbeli pozíciója eltér, viszont maximuma teljesen megegyezik. A bitcoin és az ether ára is 2021. november 8-án érte el a csúcát. Ekkoriban olyan hírek jelentek meg, hogy „a mutatók arra utalnak, hogy a bitcoin a 86 ezer dolláros szint felé tart”, vagy „a kriptovaluták tartják a versenyt a részvényekkel”. Emellett leolvasható például egy erősen exponenciális növekedés a bitcoin-nál rövid idő alatt (2021 elején), míg az ether esetében elnyújtva, később figyelhető meg mindez. Fontos megjegyezni, hogy az első magasabb csúcspont után (2021 elején) az ether és bitcoin árfolyama közel azonosan, átfedve mozgott a szakdolgozat írásának időpontjáig. Ezek alapján rendkívül erős és szignifikáns kapcsolat feltételezhető a bitcoin és ether árfolyama között. A korrelációs elemzésem során megvizsgáljuk számok alapján is ezt. Megvizsgáltam a BTC/USDT napi kereskedelmi volumenét, tranzakcióinak számát, az ETH/USDT napi

tranzakcióinak vonzásában. Érdekes különbség a tranzakciók számosságában és a tendenciában figyelhető meg, jelentősen eltér a csúcspont időbeli helyzete, különbözik 2022-től a tendencia.

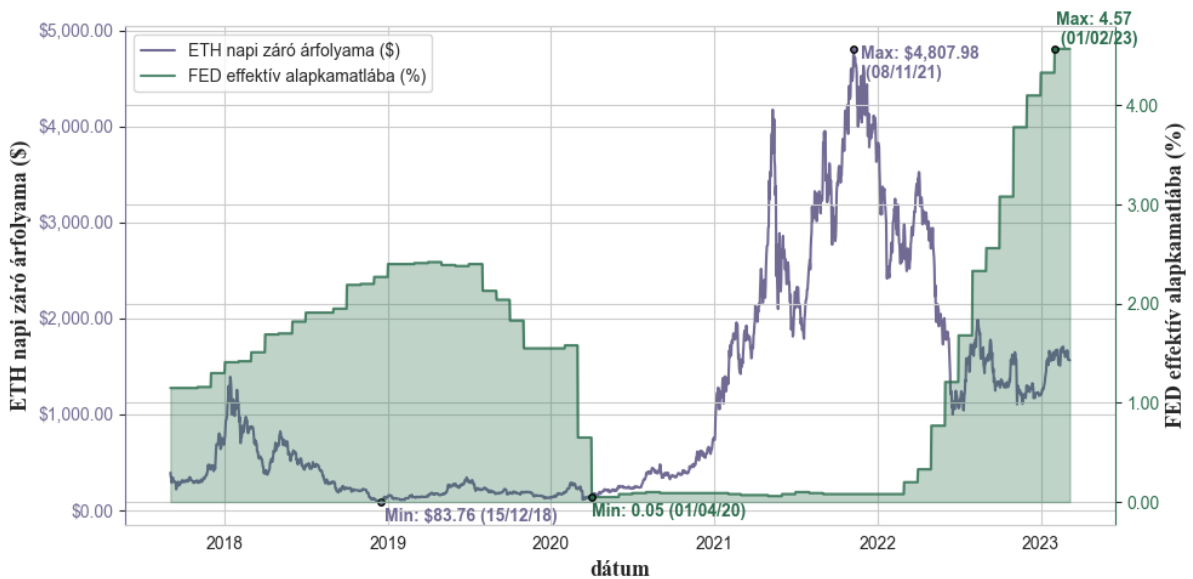
5. ábra: BTC/USDT napi kereskedelmi volumene (\$), napi tranzakcióinak száma vs. ETH/USDT napi tranzakcióinak száma



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Kifelé mozdulva a kripto világból a FED által megállapított effektív kamatláb alakulását vizsgáltam az ether árfolyamával. Ez az egyik legfontosabb kamatláb az amerikai gazdaságban, hiszen számos hitel referencia-kamatlábaként szolgál.

6. ábra: Az ETH napi záró árfolyama (\$) és a FED effektív kamatlába (%)



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

A FED egészen 2020-ig 2 százalék felett tartotta az effektív kamatláb értékét, majd a koronavírus megjelenése után 2020 áprilisában drasztikusan csökkentette azt a bezárások, intézkedések, illetve a fogyasztás csökkenése végett. Ezen periódusban a pénzbőség megjelent az ETH árfolyamában, elérte maximális értékét, majd 2022 közepére közel felére csökkent. A FED drasztikus havi emelései láthatóan befolyásolta az ether árfolyamát.

Megvizsgáltam az S&P500 és NASDAQ100 indexek viszonyát az ETH historikus árfolyamával.

7. ábra: ETH - S&P500 - NASDAQ100



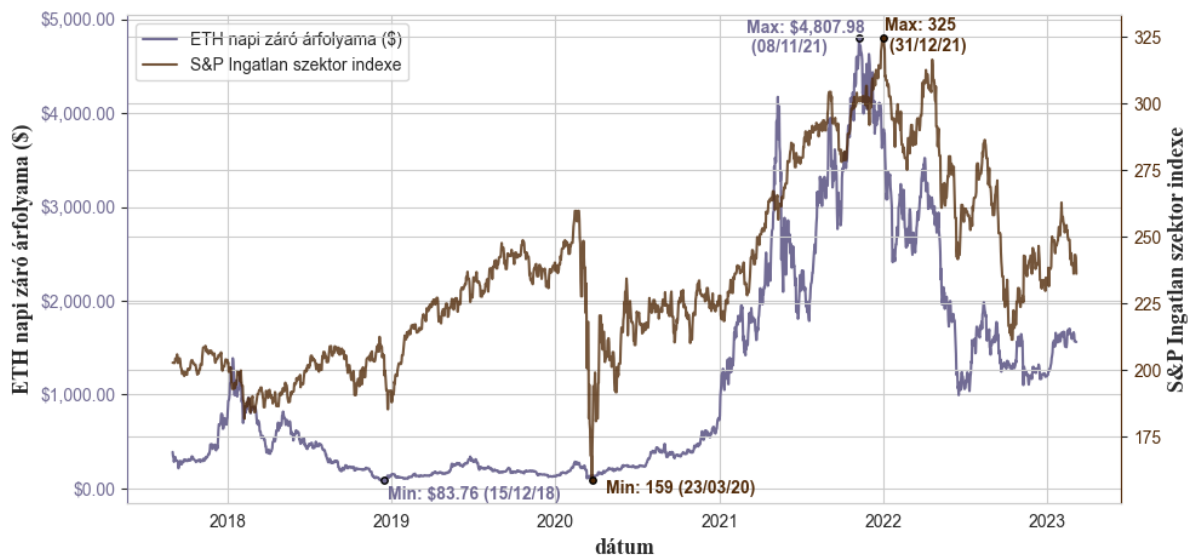
Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

A Standard & Poor's 500 indexe az amerikai gazdaság teljesítményét méri 500 fő iparági szereplők részvényein keresztül, míg a NASDAQ100 a 100 legnagyobb vállalat részvényeinek teljesítményét számos ipárból (pl. technológia, távközlés, média stb.) pénzügyi szolgáltatókon kívül. Mérsékelt lineáris növekedést tapasztalhatunk az S&P500 mutatójánál a vizsgált periódus során. A NASDAQ100 indexnél viszont egy erőteljesebb növekedést láthatunk a vizsgált időszakban és koronavírus kitörése után. Érdekeség, hogy a NASDAQ és az ether értéke 11 nap különbséggel érték el mindenkori záró csúcsukat. 2022-től viszont a tendencia mindhárom idősor esetén egyező, hol mérsékeltbben, hol drasztikusabban.

Összehasonlítottam ezen kívül a Standard & Poor's ingatlan szektorra vonatkozó indexével. Ezen mutató az 500 legnagyobb ingatlanszektorban működő cég részvényeinek alakulását aggregálja. Az alábbi ábrán egyező trendet látunk koronavírus követően valószínűleg köszönhetően a FED effektív kamatláb csökkenésének, hiszen a pénzbőség egy részét az emberek ingatlanba, részvénybe, kötvénybe fektették. Újabb érdekesség az, mintha az ingatlan

szektor index időben később, de hasonló trendet mutatna, mint az ether záró árfolyama. Korrelációs elemzésnél kitérünk részletesebben egymás hatására.

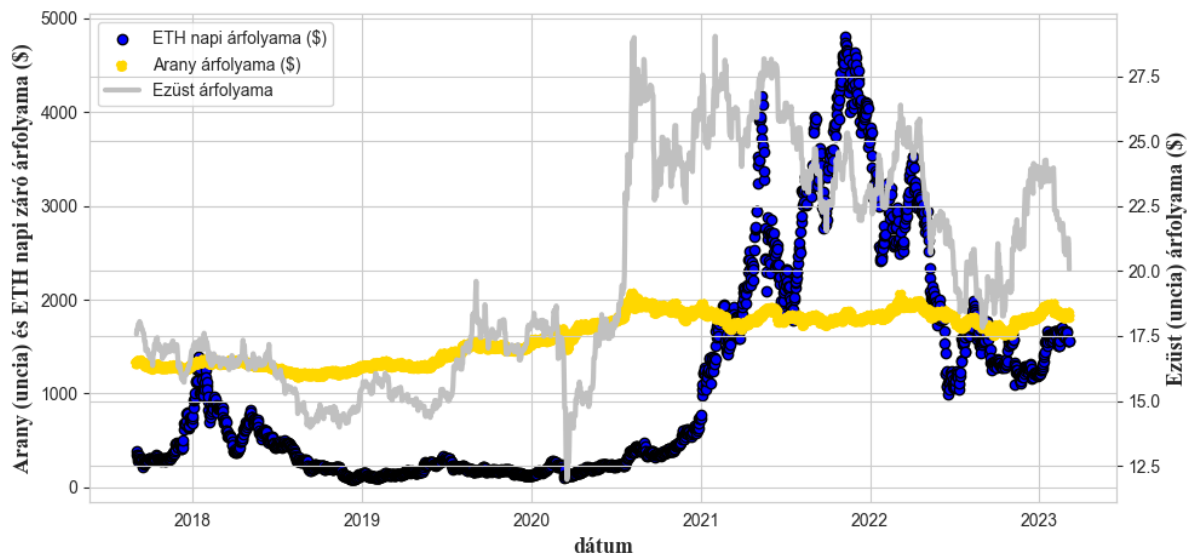
8. ábra: ETH és S&P 500 ingatlan szektor indexe



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Klasszikus árfolyamelemzésekhez hasonlóan készítettem egy vonaldiagrammot az arany, ezüst unciában mért árfolyamának alakulásáról és annak ether kapcsolatáról. Az adatokat egy weboldalról tudtam letölteni az adott Python kód segítségével. ([xau silver scraper.ipynb](#))

9. ábra: Arany, Ezüst (uncia) napi árfolyama (\$) és ETH napi záró árfolyama (\$)

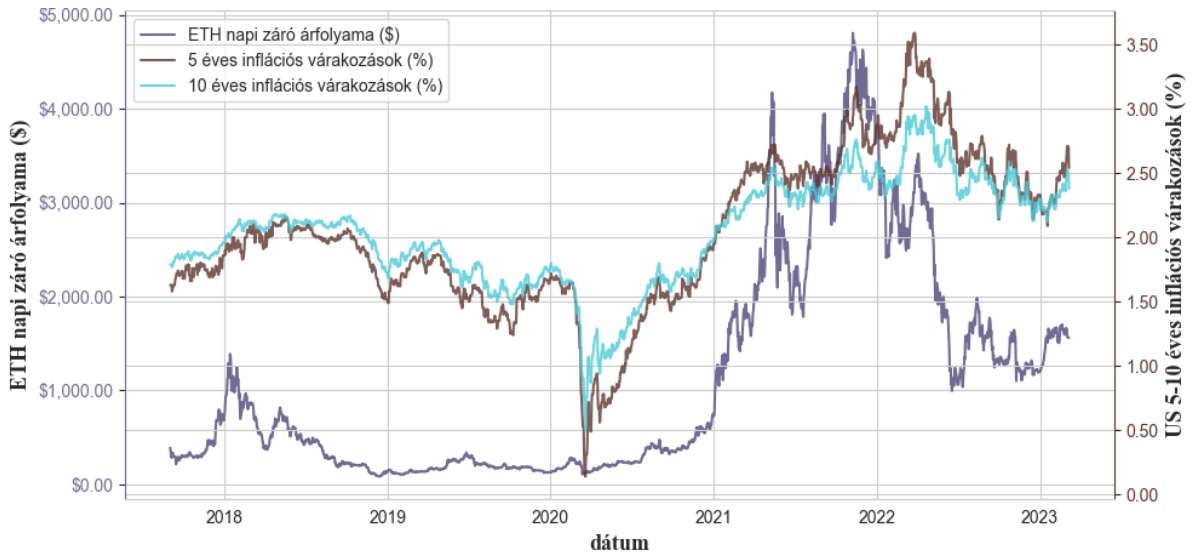


Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Legjobban az ezüst árfolyama illeszkedik az ether trendjére, az arany lineárisan és stabilan növekedett egészen a koronavírus kitörése óta. Következőkben az amerikai „breakeven” rátát vizsgáltam, amelyet a jövőbeli inflációra vonatkozó piaci várakozások mutatójaként használnak.

Általában 5 és 10 éves időtávban kalkulálják a mutatót, lényegében az inflációs várakozásokat adja vissza az inflációval védett értékpapírok és nominális értékpapírok kamatozásának különbségét alapul véve.

10. ábra: USA 5-10 éves inflációs várakozások vs. ether árfolyama (\$)



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Végül, de nem utolsó sorban az amerikai kincstárjegyek kamatainak különbségét tanulmányoztam az ether árfolyamához képest. Ha a 3 hónapos és 2 éves kincstárjegyek kamata magasabb, mint a 10 éves lejáratúaké, akkor rövidtávon belül bekövetkező recessziót prognosztizál a piac. Az alábbi diagrammon látható a lila színű idősor (10 éves lej, kincstárjegyek kamata mínusz 2 éves lej. kincstárjegy kamat) csökkenő tendenciája, amely egybe esik az ether zuhanásával.

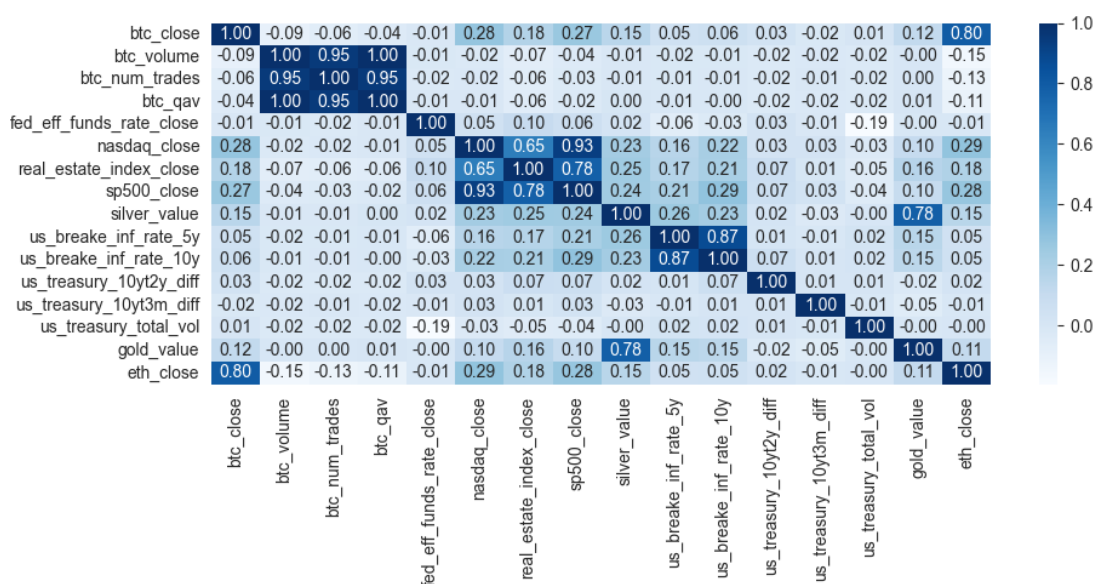
11. ábra: USA kincstárjegyek kamatának különözete vs. ETH árfolyama



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Miután minden egyes adatot bemutattam az első szekcióból, térjünk át a korreláció és szignifikancia elemzésre. Kíváncsi voltam arra, hogy számok alapján ténylegesen mely változók, mely idősorok korrelálnak az ETH záró árfolyamával. Először is minden egyes idősort „stationary”-vé (időben statisztikailag állandó) alakítottam, logaritmusos skálát vettem emellett kivontam a trendet és a szezonalitást az adatokból. Ezáltal csakis a belső, tényleges árfolyamváltozást tudjuk összehasonlítani minden egyes adatsornál. Az alábbi ábrán a szekció 1-re vonatkozó korrelációs mátrix-ot láthatjuk:

12. ábra: Korrelációs mátrix hő térképe - Szekció 1 adatai

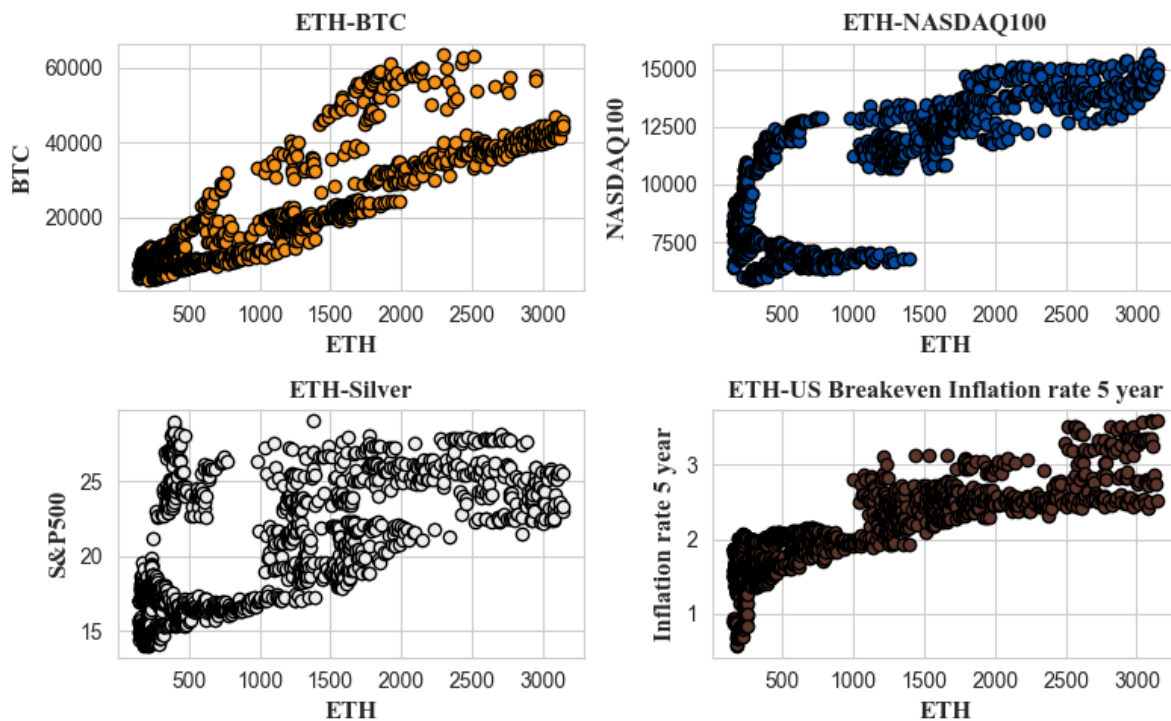


Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

A Bitcoin napi záró árfolyama és az ETH árfolyama (eth_close) között nagyon erős, pozitív irányú korreláció található. Közel közepesen erős szignifikancia figyelhető meg NASDAQ, S&P500 záró árfolyamai és az ether árfolyama között. Fontos megjegyezni, hogy például a BTC-re vonatkozó close, volume és num_trades részek egymástól függő, dependens változók (ahogyan az ábrán is észlelhető), illetve a NASDAQ100, S&P500 és S&P500 Real Estate Index is egymástól függenek, hiszen az S&P500 index egyes szereplői megjelennek a NASDAQ100, más szereplői pedig a Real Estate indexben. Hasonlóan az amerikai breakeven inflációs várakozásokra vonatkozó számok is dependesnek. Így a legfontosabb, egymástól független változók, amelyek hatással vannak és korrelációban állnak az ETH árfolyamával a következők: BTC napi záró árfolyama (erős, pozitív irányú korreláció), a NASDAQ100 napi záróárfolyama (közel közepesen erős, pozitív irányú korreláció), ezüst napi záró árfolyama (gyenge, pozitív

korreláció) és az amerikai 5 éves inflációs várakozások rátája (gyenge, pozitív korreláció). Következőekben figyeljük meg ezen változók szórásdiagrammját.

13. ábra: Szórásdiagramm kapcsolatvizsgálás végett



Forrás: Saját készítés ([statistics of training dataset.ipynb](https://statistics.of.training.dataset.ipynb))

Az adatok megtisztítottam a kiugró értékektől („outlier”), hiszen a kriptovaluták esetében a napi volatilitás akár extrém értéket is felvehet. Az ether és a bitcoin árfolyama között láthatunk egy lineáris kapcsolatot a szórásdiagrammon, az ETH-NASDAQ100 és ETH-Silver esetén viszont kevésbé nyilvánvaló a kapcsolat linearitása. Az ETH-US Breakeven Inflation rate 5 year (inflációs várakozások) esetében viszont egy kissé elnyújtott lineáris kapcsolatot tapasztalhatunk. A szekció 1 adataira az előrejelző modelleknél térünk vissza, ahol egyes idősorokat (mint NASDAQ100 vagy Ezüst árfolyam) magyarázóváltozóként fogok felhasználni.

2.1.3 Szekció 2 – „Hírek, Kommentek és Bálnák”

A „Szekció 2” elnevezésű adatok során az Ethereum rendszeréről, ETH árfolyamáról szóló híreket, kommenteket, keresési trendeket elemeztem, illetve a „bálnák” mozgását, tranzakciót követtem le. Az alfejezet elején a módszertanokat, adatokat vizualizáljuk, megismerjük, majd a végén térek ki az összefüggések, kereszt-korrelációk elemzésére. Kifejezetten angol nyelvű híreket, kommenteket és posztokat szereztem meg, hiszen a legjobb előre betanított nyelvi modellek, mint a VADER szentiment modell ([GitHub](#)) vagy Flair modell ([GitHub](#)), kifejezetten

angol szövegadatbázison tanultak be. De hogy mi is az a szentiment analízis? A **szentiment analízis** nem más, mint érzelemelemzés, más néven „véleménybányászat”. Egy adott szövegben vagy beszédben kifejezett érzelmek vagy vélemények elemzésének és azonosításának folyamata. Természetes nyelvfeldolgozási (NLP) technikák és algoritmusok felhasználásával a nyelvi modellek képesek a szöveg mögöttes hangulatát pozitív, negatív vagy semleges kategóriába sorolni. Ezáltal jól behatárolható a közvélemény hangulata és hozzáállása az Ethereum kapcsán a vizsgált időszakban. Három nyelvi modellt teszteltem, az előbbieken elhangzott Flair, Vader, és TextBlob modellt. A Flair egy mélytanulás (deep learning) és kontextuális alapú nyelvi modell, amely szóbeágyazásokat használ az érzelmek azonosítására egy adott szövegben. Képes a hangulatot finomabb szinten azonosítani, ami azt jelenti, hogy nagyobb szövegek esetén képes egyes aspektusainak hangulatát azonosítani. A VADER (Valence Aware Dictionary and Sentiment Reasoner) ezzel ellentétben egy olyan lexikon és szabályrendszerek által meghatározott hangulatelemző modell, amelyet konkrétan közösségi médiák kommentjein, posztjain és hangulatain tanítottak be. A TextBlob egy egyszerű, szabályalapú hangulatelemző modell, amely egy előre definiált „hangulatlexikon” segítségével osztályozza egy adott szöveg hangulatát. Mindháromban közös, hogy egy-egy szöveg, komment értékelése során hangulatintenzitási pontszámot adnak vissza -1 és 1 között, értelemszerűen a -1 közeli pontok negatív, a nulla közeli semleges, az 1 közeli pedig pozitív hangulatot tükröznek. ([sentiment analysis.ipynb](#))

14. ábra: Nyelvi modellek szentiment analízise példán keresztül

```
#Example
flair_sentiment("The price of ether has fallen 30% in the last week")
-0.9999

#Example
textblob_sentiment("The price of ether has fallen 30% in the last week")
0.0

#Example
vader_sentiment("The price of ether has fallen 30% in the last week")
-0.3612
```

Forrás: Saját készítés ([sentiment analysis.ipynb](#))

Ahogy a 14. ábrán is látható, különböző értékeléseket adnak egyes modellek a kifejezetten negatívnak mondható szalagcímre. Tesztelések és próbálgatások után a Vader szentiment modelljét választottam, hiszen rengeteg személyes véleményt (Twitter tweet, Reddit komment)

elemeztem, illetve sok esetben a hírek szalagcíme meglepően hasonlít a felbukkanó hatásvadász kommentekre. Így minden egyes részhez a Vader modelljét használtam fel, természetesen némi adattisztítás (felesleges karakterek, új sor jelzők, tab jelek kitörlése) után.

Hírek. Megérzésem szerint manapság a pozitív és negatív hírek szerepe megnövekedett, az egyik legfontosabb befolyásoló tényezőként írható le. Ezen szekció adatait az Ethereum-ra vonatkozó hírekkel kezdtem meg. A világ legnagyobb keresőoldalán (Google) hírrovatában kezdtem el a kulcsszavas keresésemet és scrapelésemet Python segítségével. ([google_news_scraper.py](#)) Fontos megjegyzés, hogy sajnos limitációk miatt csak 200 releváns hírt tudtam leszűrni a vizsgált periódusban. Kiegészítésképpen az internet két legnagyobb kripto hírportáljáról (CoinTelegraph, Coindesk) töltöttem le Ethereum-ra releváns híreket és fejléceket (headline). ([scrape_coindesk.py](#) és [scrape_cointelegraph.py](#)) Összesen 6190 hírt tudtam megszerezni CoinTelegraphról és 1511 hírt Coindeskről. Így több, mint 7500+ hírt, szalagcímet, rövid hírösszefoglalót, linket gyűjtöttem össze a kijelölt időszakra napi bontásban.

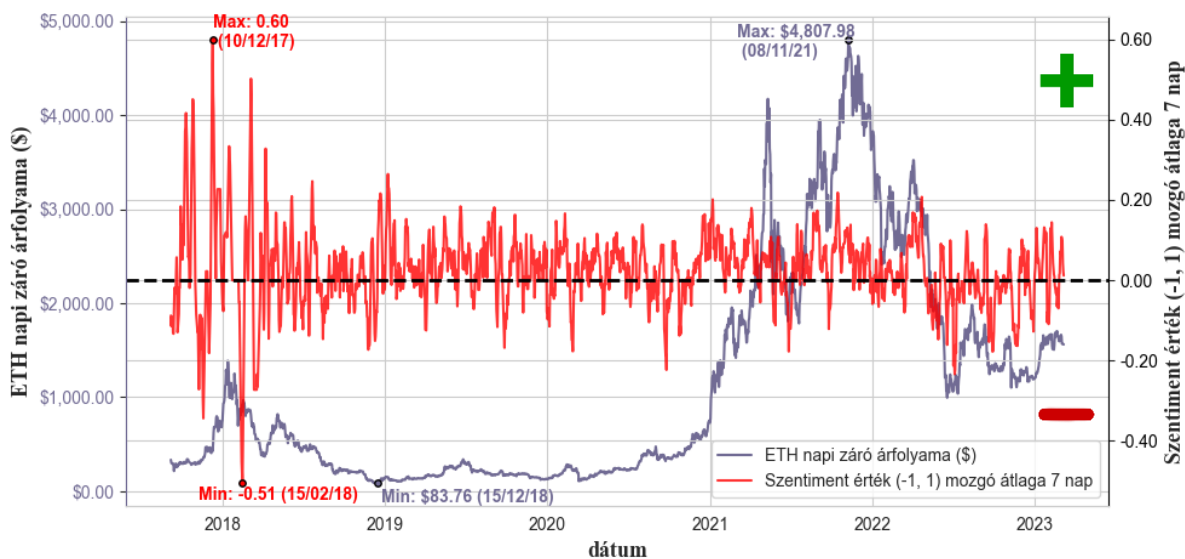
15. ábra: Hírek tábla felépítése

```
Data columns (total 7 columns):
#  Column          Non-Null Count  Dtype
---  -
0  date              1511 non-null   object
1  headline          1511 non-null   object
2  short_text        1511 non-null   object
3  link              1511 non-null   object
4  long_text         1511 non-null   object
5  sentiment         1511 non-null   float64
6  final_sentiment   1511 non-null   float64
```

Forrás: Saját készítés ([daily_training_dataset.ipynb](#))

Ezt követően az előzőekben megismert Vader modelljével szentiment analízist végeztem a hírek szalagcímén. Az alábbi diagrammon a hírek szentiment értékét láthatjuk historikusan. A skála -1 és +1 között mozog, -1 közelében negatív hírek születtek az adott napon, +1 közelében pedig inkább pozitívabb hírek.

16. ábra: Ethereum hírek szentiment értéke a vizsgált időszakban



Forrás: Saját készítés ([statistics_of_training_dataset.ipynb](#))

A szentiment pontok volatilitása napi szinten rendkívül magas, így vizualizáció szempontjából az értékek elmúlt 7 napi mozgó átlagát vettem. A diagramm alapján kivehető a legjobb és legrosszabb hírek időbeli elhelyezkedése, megfigyelhető a 2021. év eleji pozitív hír áradat és árfolyamemelkedés, illetve a 2022. év eleji drasztikus csökkenés. Vizsgáljuk meg ezen időszakokat fókuszáltan:

17. ábra: Hírek "hangulata" 2021 elején

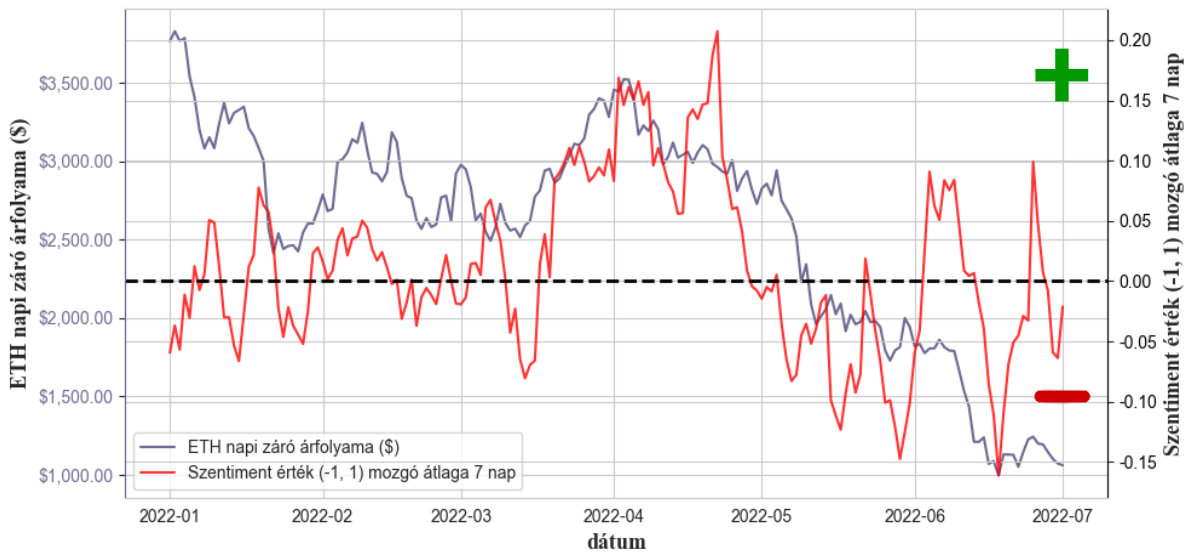


Forrás: Saját készítés ([statistics_of_training_dataset.ipynb](#))

A 17.ábrán jól kivehető a pozitív hírek áradata 2021 elején, olyan szalagcímek jelentek meg, hogy: „Ethereum will become the main asset for investors in 2021” vagy „Ether is flying to the moon? Here’s what’s fueling the ETH rocket”. Mindezek mellett a kripto sajtó az ether

történelmi csúcsaitól és adoptációs lehetőségeitől volt hangos. Az árfolyam masszív növekedésbe kezdett 2021 közepéig. Érdekes lesz megvizsgálni azt, hogy vajon az árfolyam hat a hírekre vagy fordítva. Később mindez kiderül. Térjünk át a 2022 eleji időszakra, amikor egy grandiózus esést tapasztalhatunk mind az árfolyamban, mind a hírek polaritásában köszönhetően a háború kitörésének. Olyan hírek jelentek meg, mint: „Az ETH árfolyama mélyponton van”, „két hét alatt 30%-ot esett” vagy „nehezen tart ki az ether a nyomás alatt”.

18. ábra: Hírek „hangulata” 2022 elején



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Sejtésem és a fenti ábrák szerint valamelyest hat a pozitív-negatív hírek aránya az árfolyamra, második hipotézisemet az alfejezet végén vizsgálom. Ezt követően a világ egyik legnagyobb közösségi oldalán, a **Reddit**en kerestem az Ethereum, ETH, ether kulcsszóval rendelkező kommenteket. Összesen több, mint 4600 kommentet tudtam leszedni 2017 szeptemberétől kezdődően, a legtöbbet a „CryptoCurrency” és „ethfinance” subredditokból (aloldalak). ([reddit_scraper.ipynb](#))

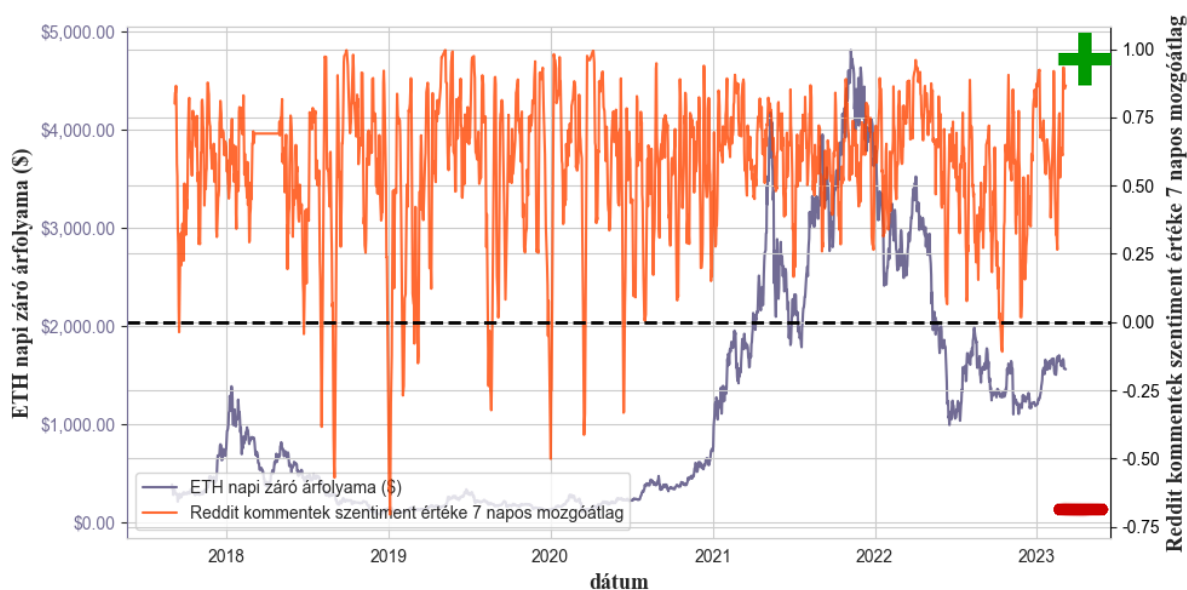
19. ábra: Top 5 subreddit oldal – kommentek száma és pontok összege

subreddit	comments	total_score
CryptoCurrency	1100	3149
ethfinance	767	13022
ethtrader	611	4295
ethereum	494	3652
ledgerwallet	94	169

Forrás: Saját készítés ([sentiment_analysis.ipynb](#))

Redditre a kommenteket a felhasználók le vagy fel pontozhatják, így érhető el az, hogy egyes kommenteken 25 score (felpontozás) van, vagy -25. (lepontozás) Az elemzés előtt kiszűrtem a 0 pont alatti kommenteket, amelyek általában gyanúsán spamek vagy bot által generált üzenetek. A kommenteket előzetesen adattisztítottam, és lefuttattam a jól ismert Vader modellt. A kapott szentiment értékeket súlyoztam az adott napi totál score (fel-leszavazások) értékével, illetve a napi bontásban nehezen értelmezhető szentiment értékeket a hírekhez hasonlóan 7 napi mozgó átlagra alakítottam és elkészítettem egy diagrammot:

20. ábra: Reddit kommentek - Ethereum, ETH és ether kulcsszóval



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Érdeemes előzetesen megjegyezni, hogy az adott Reddit subreddit aloldalak az Ethereum-ot támogató felhasználók kedvenc lokációja, így ahogyan az ábrán is látszik, a pozitív hangulat jócskán kitartott a vizsgált időszakban. Néhány időpontban láthatunk negatívabb eseményeket, ilyen a 2020 eleji időszak, amikor is a koronavírus megjelenése közben a kommentelők kritizálták az Ethereum Merge-re vonatkozó terveket. A számosság kevés ahhoz, hogy konkrétan kimondható legyen egy pozitív hatás, viszont szemügyre vehető a pozitivitás intenzitása és kiugró értékei a 2021-es nagy felütés előtt.

A Reddit kommenteket követően a Twitter tweetek-re tértem át, a Snsrape Python modul segítségével sikerült közel 25 millió tweetet letöltenem. ([twitter scraper.ipynb](#)) A tábla felépítése a következő:

21. ábra: Twitter tweeteket tartalmazó tábla felépítése

#	Column	Dtype
0	datetime	object
1	tweet_id	int64
2	text	object
3	likes	int64
4	username	object
5	user_followers	int64
6	user_location	object
7	user_verified	bool
8	final_sentiment_test	float64

Forrás: Saját készítés ([daily_training_dataset.ipynb](#))

A Twitter üzenet írók felhasználónevét, követőszámát, publikus lokációját, és igazolt státuszát nyertem ki az kódokból. Több, mint 2,5 millió egyedi felhasználó Ethereum-mal kapcsolatos tweetjét sikerült megfognom, lokáció szerint a legtöbbet az USA-ból és Angliából.

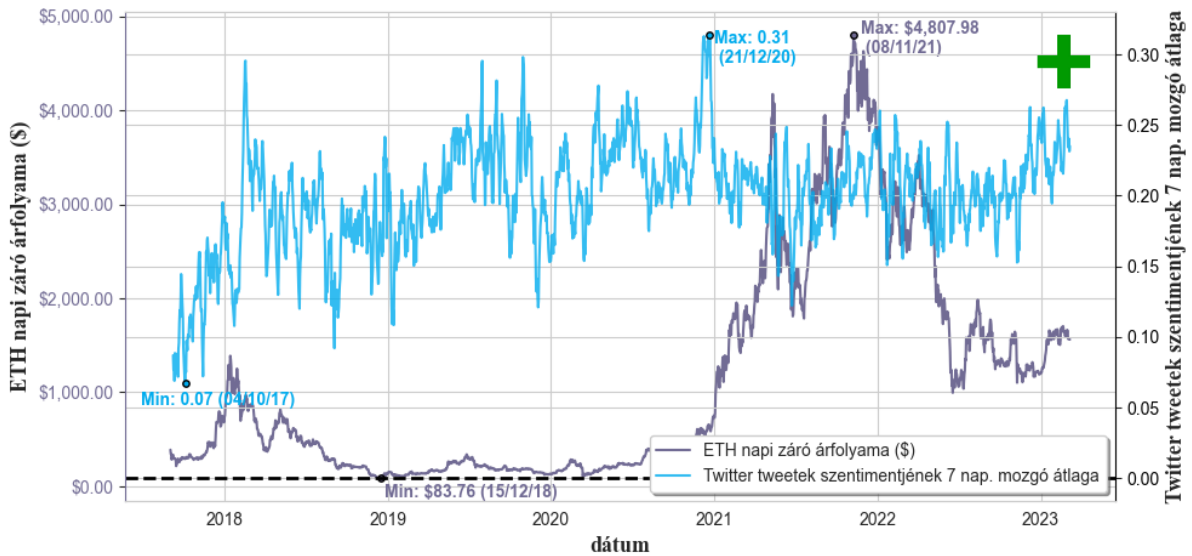
22. ábra: Top felhasználók és lokációk tweet szám alapján (db)

username		user_location	
coinspectator	128071	United States	354541
theautomatski	125271	Metaverse	288767
CyberFM_Rock	104657	United Kingdom	229345
ClankApp	88386	California, USA	194014
robostopia	87109	Worldwide	188952

Forrás: Saját készítés ([daily_training_dataset.ipynb](#))

A „coinspectator”, „automatski” felhasználók például botok, amelyek Twitter posztok formájában kripto híreket közölnek az oldalukon. Ezután a tweetek textjét megtisztítottam, majd Vader modell alá vetettem. Miután mind a 25 millió Twitter kommentet kiértékelte a modell, kiszűrtem a 0 követővel rendelkező felhasználókat, hiszen rengeteg spam célú robot található meg az oldalon. Mindezek után a „hangulatpontokat” súlyoztam a felhasználó követőszámával napi szinten. A lenti diagrammon az értékek 7 napos mozgó átlagát láthatjuk a könnyebb vizualizáció, trendkövetés végett. Fokozatos és lineáris növekedést tapasztalunk a szentimentek maximum értékéig (2020 vége), majd egy kisebb csökkenés melletti stabilizációs folyamatot. Sejtésem szerint közepes erősségű szignifikancia figyelhető meg a napi árfolyam és napi hangulatpontok között, az alfejezet végén mindenre fény derül.

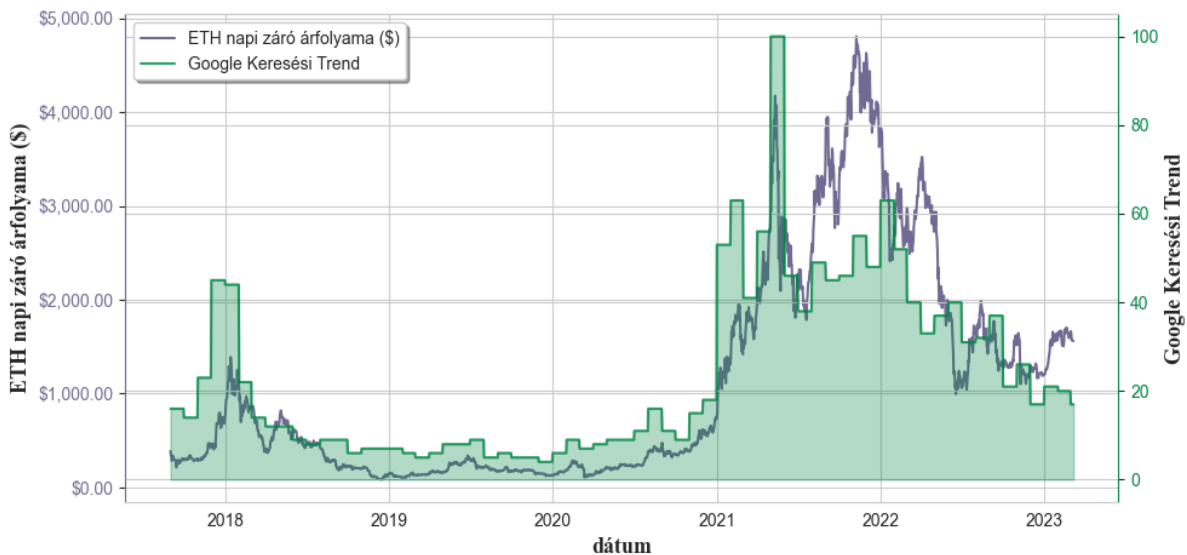
23. ábra: Ethereum, ETH, ether kulcsszavakat tartalmazó Twitter tweetek szentimentje



Forrás: Saját készítés ([statistics_of_training_dataset.ipynb](#))

A Twitter tweet-ek után a világ legnagyobb keresőoldalának adatait vettem össze az árfolyam alakulásával. A **Google Trends** ([link](#)) weboldalon **havi bontásban** historikusan tudunk rákeresni Ethereum témakörben a keresési trend világszintű alakulására. A nullától százig terjedő skála a keresési népszerűséget méri, azaz nem a számosságot hasonlítja össze az idő múlásával, hanem a többi keresési kulcsszó/témakör népszerűségével, keresettségével.

24. ábra: Google Keresési Trend - Ethereum témakörben



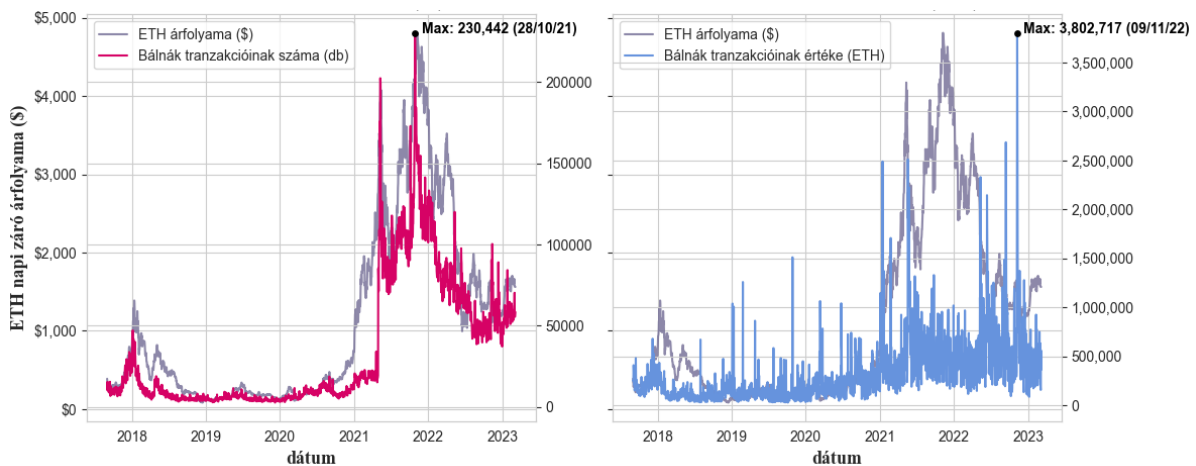
Forrás: Saját készítés ([statistics_of_training_dataset.ipynb](#))

A keresési trend és az emberek érdeklődése az első nagy felfutás végén rendkívül megnövekedett, és egészen populáris maradt a 2022 közepi zuhanásig. Az ábra alapján egy

minimális megérzésem támadt a keresettség és az ether korreláció meglétére, az alfejezet végén kiderül.

Az egyik legizgalmasabb része a kutatásnak a **bálnák** szerepe és aktivitása volt, amelyhez a Google Bigquery publikus adattárházát hívtam le. ([whales.sql](#)) Az összes Ethereum címet tartalmazó táblából a top 10 ezer legtöbb ether-rel rendelkező címet gyűjtöttem ki, majd ezen „bálnák” tranzakcióinak számát és értékét (ETH-ban) kértem le historikusan. A lenti diagrammokon szemügyre vehető, hogy a bálnák tranzakcióinak száma az árfolyam trendjéhez hasonló, viszont értékének tendenciája kissé különböző. A jobb oldali ábrán látható, hogy 2022. november 9-én több, mint 3,5 millió ether tranzakcióját bonyolították le a bálnák. Vizsgáljuk meg ezt az időszakot tüzetesebben!

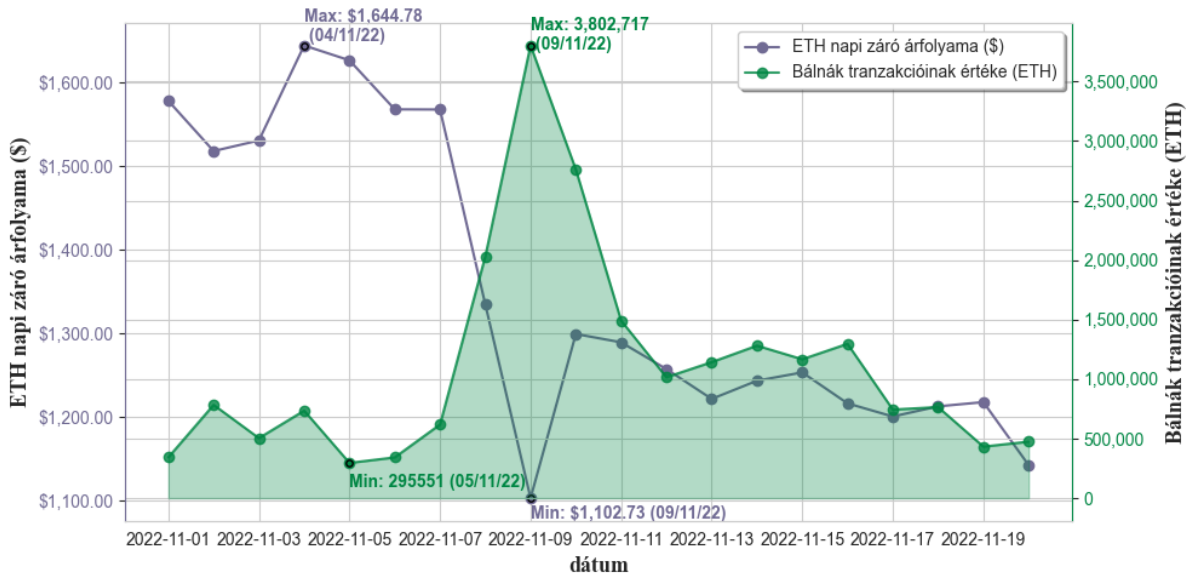
25. ábra: Bálnák tranzakcióinak száma (db) és értéke (ETH)



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

A 26. ábrán egyértelműen a bálnák jelenlétét és árfolyamot befolyásoló tevékenységüket figyelhetjük meg. A vizsgált periódusban (2022. november 11. és november 20. között) az ether árfolyama kevesebb, mint 6 nap alatt közel 70%-ot esett. A bálnák hetedikén és nyolcadikán több millió ETH tranzakcióját folytatták le, majd kilencedikén elérték mind a vizsgált periódus, mind az egész időszak csúcsát, amikor is közel 4 millió darab ETH-val kereskedtek. Fontos megjegyezni, hogy ebben az időszakban ment csődbe az FTX kriptotőzsde, illetve a Binance kihátrált annak megmentése mögül, a bálnák valószínűleg a pénzüket mentették a centralizált kriptotőzsdékből. Az árfolyam vélhetően ennek is köszönhetően majdnem megfeleződött, 1100 dollárra esett vissza. Ezután viszont szépen diagnosztizálható a bálnák bevásárlási folyamata tizedikén (ld. 25. ábra). Eddigi tapasztalataim és elemzéseim szerint az egyik legerősebb befolyásoló tényező a szekció 2 adatai közül a bálnák mozgása lehet.

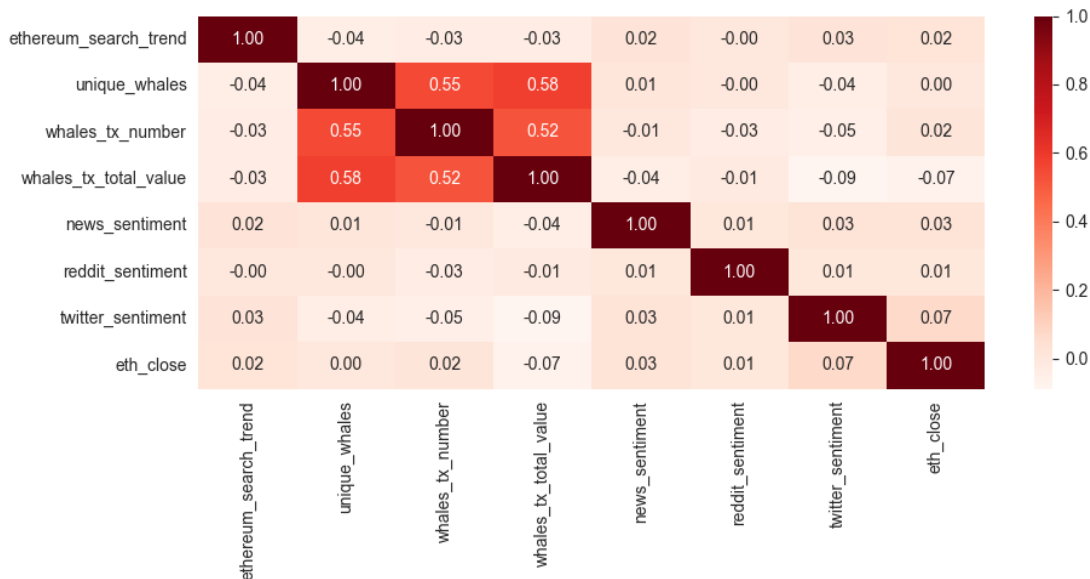
26. ábra: Bálnák tranzakcióinak értéke 2022-11-01 és 2022-11-20 között (ETH)



Forrás: Saját készítés ([statistics_of_training_dataset.ipynb](#))

Miután megismertünk minden adatot Szekció 2-ből, térjünk át a kapcsolatok elemzésére, korrelációs mátrixra, kereszt-korrelációs együtthatókra és a második hipotézisem kifejtésére. Előzetesen minden adatot stationary-vé alakítottam (időben statisztikailag állandóvá) alakítottam, először logaritmikus skálán vettem, majd trend hatásának eltávolítása végett napi változásokat/megtérüléseket vettem alapul. Így csakis az idősorok belső változását hasonlítottam össze egymással.

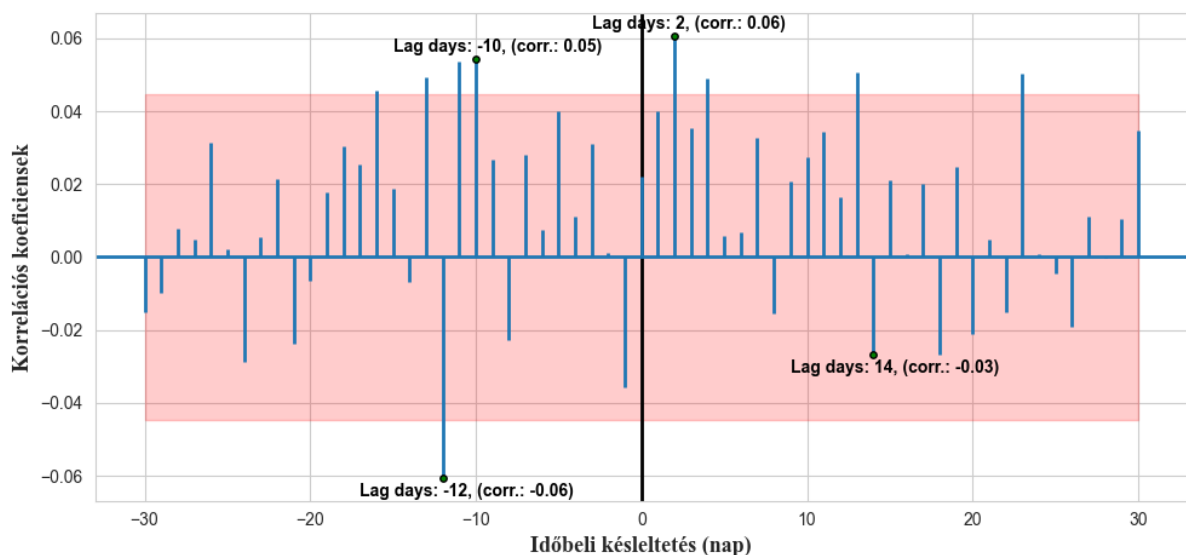
27. ábra: Szekció 2 adatainak korrelációs mátrixa



Forrás: Saját készítés ([statistics_of_training_dataset.ipynb](#))

Fontos megjegyezni azt, hogy a fenti korrelációs mátrix két idősor napi logaritmikus változásának korrelációját méri, így nem ad teljes képet azokról az effektusokról, amelyek hamarabb vagy később fejtik ki hatásukat az ether árfolyamára. Ezeket kereszt-korrelációval vizsgálom meg a későbbiekben. Az ábra alapján kivehető, hogy az **eth_close - twitter_sentiment**, és az **eth_close - whales_tx_total_value** idősorok napi változásában gyenge korrelációt (-0.07 és 0.07) tapasztalhatunk. Meglepetésre míg így is ezek voltak a legerősebbek a többi idősorhoz képest, hiszen a news_sentiment 0.03, reddit_sentiment pedig 0.01 erősségű korrelációt mutat az ether napi záró árfolyamváltozásával szemben, így elhanyagolható kapcsolatuk. Mi történik akkor, ha nem az adott napi változások közötti korrelációs kapcsolatot keressük, hanem az egyik idősort időben valamelyik irányba eltoljuk. Fontos: Minden esetben az ether napi záróárfolyamához képest csúsztattam el a megnevezett idősorokat! Vizsgáljuk meg a kereszt-korreláció erősségét a következőben:

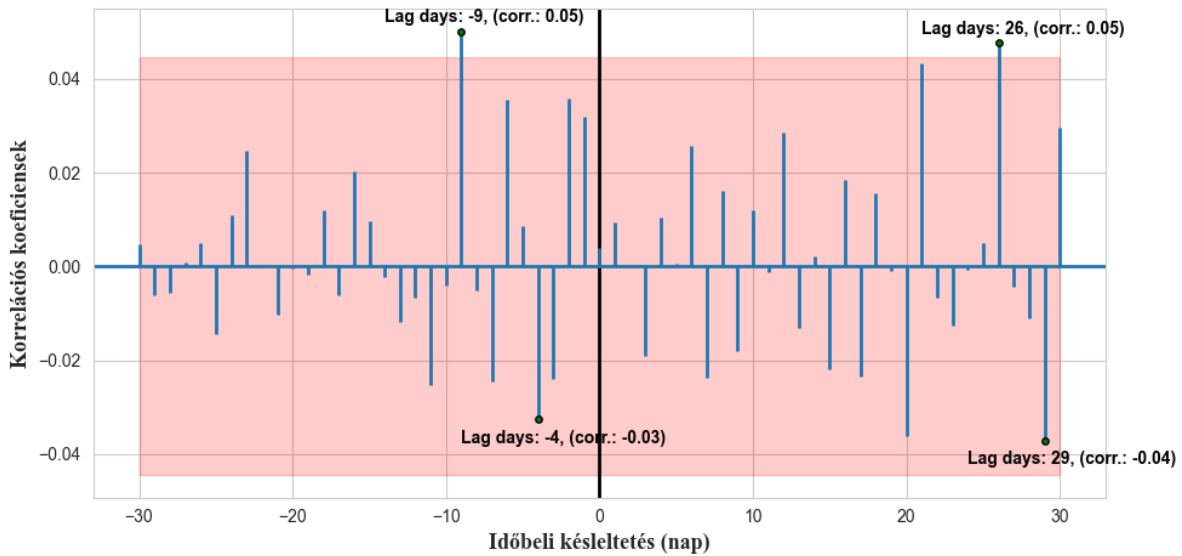
28. ábra: Ethereum Search Trend - Kereszt-korreláció 95%-os konfidencia-intervallummal



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Az Ethereum keresési trendjénél tudtam elérni javulást, viszont sajnos az adatok havi bontása miatt csalókás lehet a 10 napos idősor csúsztatás. Az ábra alapján az Ethereum keresési trendjének 10 nappal korábbi változása és az árfolyam jelenlegi változása között minimálisan erősebb 0.05-ös korrelációt tapasztalhatunk. A bálnák adott napi számának változása és az árfolyam között 0-s korrelációt tapasztaltunk, 9 napos eltolás esetén viszont 0.05-öset. Kereszt-korrelációs elemzés során előfordulhat véletlenszerű szignifikancia és csúsztatás („lag”), így érdemes racionálisan tekinteni az eredményekre.

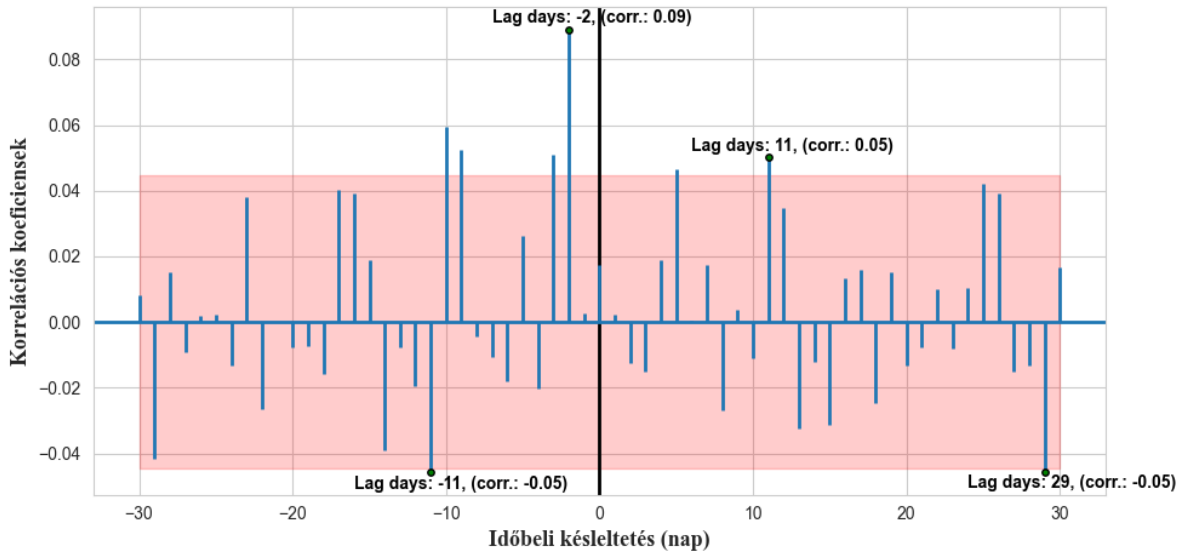
29. ábra: Bálnák száma (unique_whales) - Kereszt-korreláció 95%-os konfidencia-intervallummal



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

A bálnák tranzakcióinak számában viszont hatalmas előrelépést jelent a két napos eltolás, hiszen a korreláció mértéke 0.02-ről 0.09-re javult. Racionálisan nézve a bálnák megnövekedett tranzakcióinak száma később fejti ki hatását az árfolyamra.

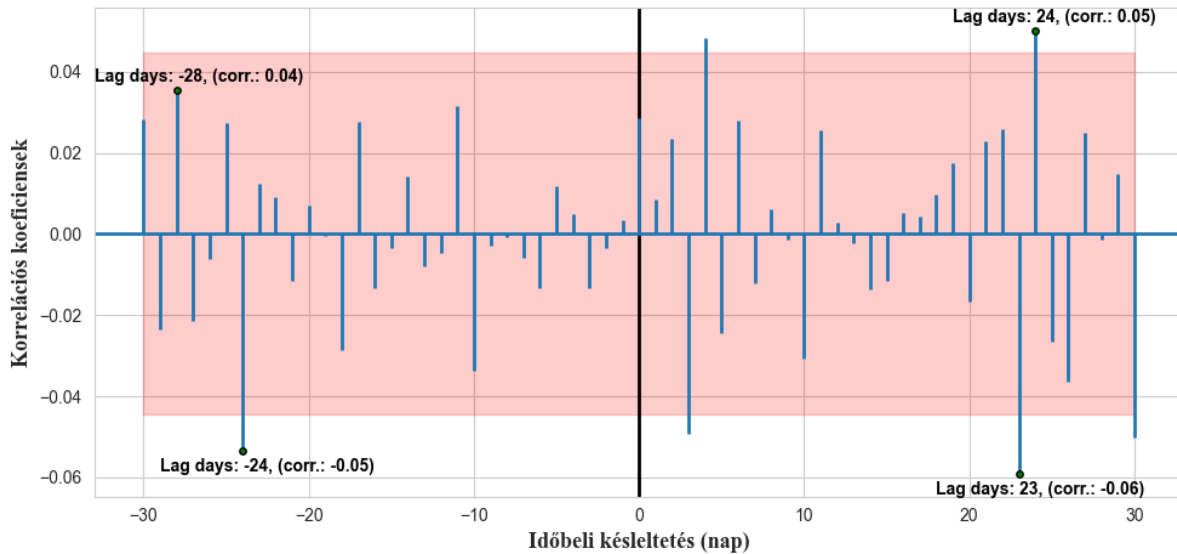
30. ábra: Bálnák tranzakcióinak száma (whales_tx_number) - Kereszt-korreláció 95%-os konfidencia-intervallummal



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

A hírek szentimentváltozása során viszont láthatjuk, hogy elcsúsztatás esetén sem javult lényegesen a korreláció, 24 napos eltolás esetén egy -0.05-ös korrelációt mutat az ábra, ám racionálisan nézve ez csak egy anomália lehet.

31. ábra: Hírek szentimenje (news_sentiment) - Kereszt-korreláció 95%-os konfidencia-intervallummal



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Végül, de nem utolsó sorban Granger Casualty teszt alá vettem az idősorokat. A Granger-Casualty statisztikai teszt segítségével megállapítható egy idősorról, hogy az alkalmazható-e egy másik idősor prediktálására. A Granger okozati összefüggés tesztje két regressziós modell illesztését foglalja magában: az egyikben csak az Y múltbeli értékei szerepelnek előrejelzőként, a másikban pedig az Y és az X múltbeli értékei is. Fontos megjegyezni azt, hogy szó szoros értelemben az eredmény nem jelent oksági kapcsolatot két idősor között. Minden egyes párra lefuttattam a tesztet, viszont csakis az Ethereum keresési trendre találtam szignifikáns értéket. ([statistics of training dataset.ipynb](#))

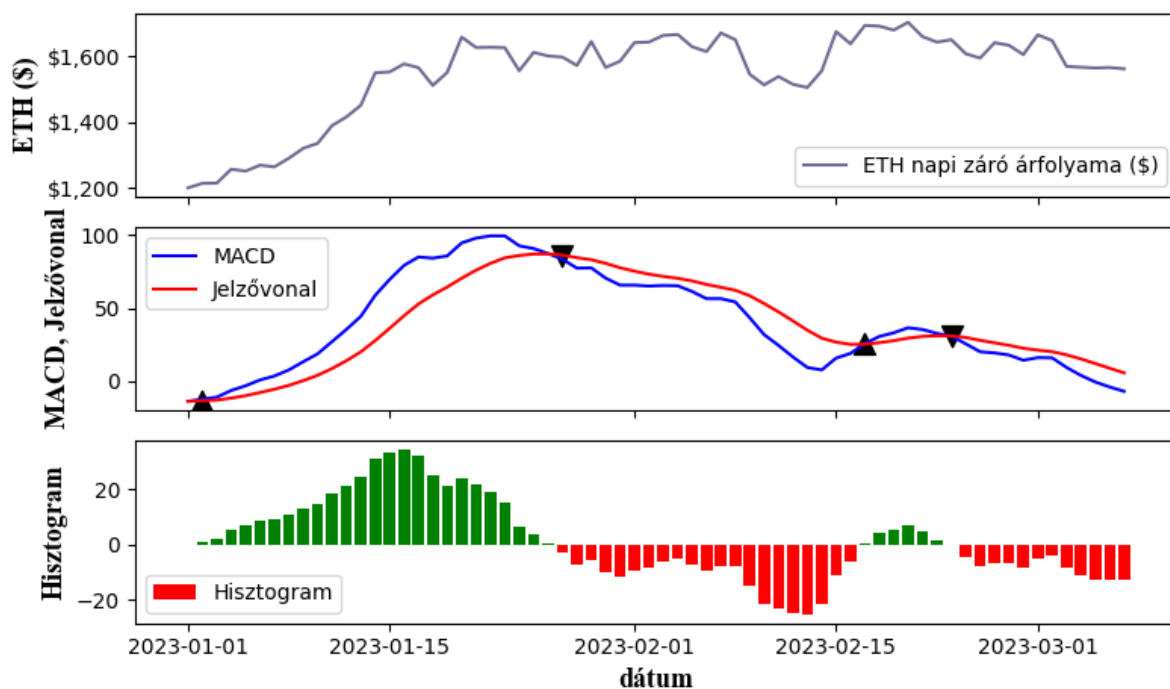
A második hipotézisem, amely így szól: **Az Ethereum-mal kapcsolatos hírek, Google keresési trendek, Reddit kommentek és Twitter tweetek szignifikáns kapcsolatban állnak az ether árfolyamával** megdőlt, hiszen a korrelációs elemzések és Granger tesztek során csakis gyenge korrelációt tapasztaltam. Így a **második hipotézisemet megcáfolom**, a kapcsolat erőssége gyenge.

Összegezve rengeteg adatot sikerült összegyűjteni, adattisztítani, elemezni. Keresési trendek és hírek kapcsán kevesebbet, Reddit kommentek és Twitter tweetek esetén viszont lényegesen több információt sikerült gyűjteni. Az idősorok hullámán a bálnákat is megfigyelhettük, milyen tranzakciókat folytattak le, milyen értékben az elmúlt években és ez milyen hatással volt az árfolyamra. Az utolsó szekció adatai jönnek, következzenek a technikai indikátorok!

2.1.4 Szekció 3 – „Technikai indikátorok”

Végül, de nem utolsó sorban technikai indikátorokkal egészítettem ki az adathalmazt. Momentum, trend, volumen és volatilitási indikátorokat hoztam létre a bázis szekció adataiból. Minden egyes indikátortípusból szeretnék egyet bemutatni, amelyek a legrelevánsabbak tudnak lenni az előrejelzés során. Elengedhetetlen az, hogy kapcsolatvizsgálatot és korrelációt ebben a szekcióban nem vizsgálok, hiszen az indikátorok dependens magyarázóváltozók az árfolyammal. A **momentummutatók** olyan indexek és mutatók, amelyek az értékpapír, kriptovaluta árának sebességváltozásának értelmezésére alkalmazhatók. A momentum indikátorokat önmagukban nem elegendők teljeskörű elemzéshez, mert nem a mozgás irányát, hanem csak az árváltozást magába ölelő időkeretet. Emellett a momentumindikátorok segítségével elemezhetjük az egyes részvények esésének vagy emelkedésének sebességét. Az egyik legfontosabb az MACD (Moving Average Convergence Divergence), amely a két súlyozott (26 napos és 12 napos) mozgóátlag különbségét hasonlítja össze a 9 napos súlyozott mozgóátlaggal. A különbséget ábrázoló vonalat MACD vonalnak, a 9 napos mozgóátlagot pedig jelzővonalnak hívják. A vételi jelzést az MACD akkor generálja, amikor alulról lépi át a jelzővonalat, értelemszerűen az eladási jelzést pedig akkor, amikor felülről. Vizsgáljuk meg 2023-ra a mutatót:

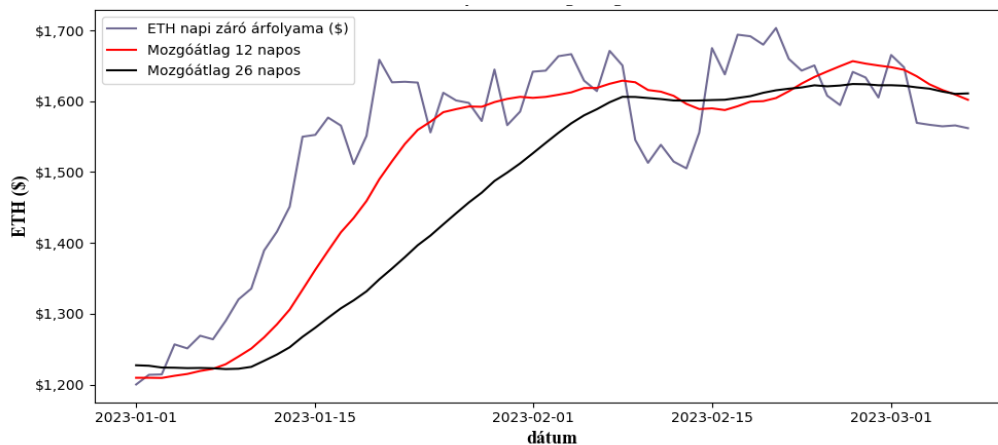
32. ábra: MACD mutató 2023



Forrás: Saját készítés (statistics_of_training_dataset.ipynb)

A diagrammon láthatjuk a vételi és eladási jeleket fekete háromszöggel. Január elején például egy vételi jelet láthatunk, az MACD metszi a Jelzővonalat, az árfolyam 20%-nál is többet emelkedett a következő jelig. Január végén ezzel ellentétben eladási jelet mutatott az indikátor, az árfolyam és a hisztogram is negatív növekedési pályára lépett. A **trendmutatók** a trend felismerésében, lekövetésében nyújtanak hatalmas segítséget. A legismertebb indikátor a mozgóátlag, amely x periódusra visszamenően átlagolja a napi záró árakat, harmonizálja és kisimítja a görbét. A hosszú távú trendekkel foglalkozó kereskedők számára a 200 napos, 100 napos és 50 napos egyszerű mozgóátlagok népszerűek.

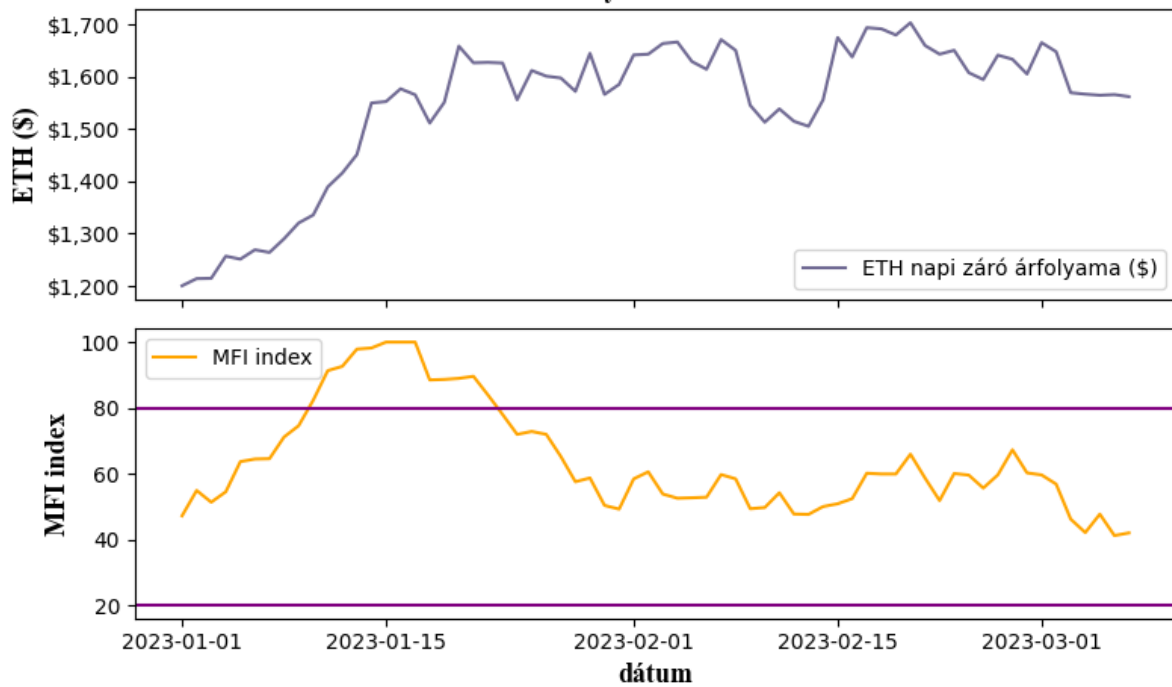
33. ábra: 26 és 12 napos mozgóátlagok 2023-ban



Forrás: Saját készítés ([statistics of training dataset.ipynb](https://statistics.of.training.dataset.ipynb))

A **menyiségi (volumen) mutatók** bemutatják azt, hogy más kereskedők hogyan érzékelik a piacot, elemzésük segít a trendek és a minták további értelmezésében. A volumen indikátorok egyik fő előnye a korai előrejelzés várható árfolyammozgásokra. A Money Flow Index (MFI) egy olyan mutató, amely képes a kereskedelmi nyomást mind eladási, mind vételi oldalról úgy, hogy közben figyelembe veszi az időt és az árat. Szakértők volumennel súlyozott relatív erősségi indexnek is nevezik ezt a mutatót, hiszen nem csak az árat, hanem a volument is számításba veszi. A Money Flow Index (MFI) az RSI-hez hasonlóan értelmezhető, kereskedési jeleket generál a pénzügyi eszköz helyzetéről - túlvásárolt vagy túladott zónában helyezkedik-e el.

34. ábra: Money Flow index



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

2023 elején az indikátor átlépte a túlvásárolt zónát, csökkenése előre jelezte az árfolyam esését 2023 februárjára. Az utolsó típus a volatilitási indikátorok, hiszen tőzsdei kereskedés során nem csak a piac trendjét, konszolidációját érdemes megfigyelnünk, hanem a volatilitással is foglalkoznunk kell. A volatilis időszakok a kriptokereskedelemben hatalmas kilengéseket okoznak megnehezítve a kereskedést.

A legismertebb mutató a Bollinger szalagok (Bollinger Bands), amely 3 részre tagolható: a felső, alsó és középső sávok. A középső sáv a 20 napos vagy sávok mozgóátlaga, a felső sávnál az előzőidőszak szórása (+2) adjuk hozzá a mozgóátlaghoz, az alsó sáv pedig előzőidőszak szórása (-2) adunk hozzá a középső sávhoz.

Amikor a piaci volatilitás nő, a sávok kitágulnak, amikor a volatilitás csökken, akkor pedig összeszűkülnek. A kereskedők Bollinger-sávokat akkor veszik használatba, amikor egy nyugodtabb, konszolidáltabb fázis után az árfolyam áttöri a felső, vagy az alsó sáv határát.

35. ábra: Bollinger szalagok 2023



Forrás: Saját készítés ([statistics of training dataset.ipynb](#))

Összegezve a leírtakat rengeteg adatot sikerült gyűjtenem az elmúlt közel 6 évből napi szinten, kezdve az ether-re vonatkozó adatokkal, hírekkel, Twitter, Reddit kommentekkel, bitcoin árfolyammal, NASDAQ100, S&P500 árfolyammal stb. Kiegészítettem egyéb technikai indikátorokkal is, amelyek segíthetnek az előrejelzés során a modellnek. Minden adatok végül a [daily training datasetben](#) mergettem össze és készítettem elő predikcióra!

2.2 Árfolyam-előrejelzés

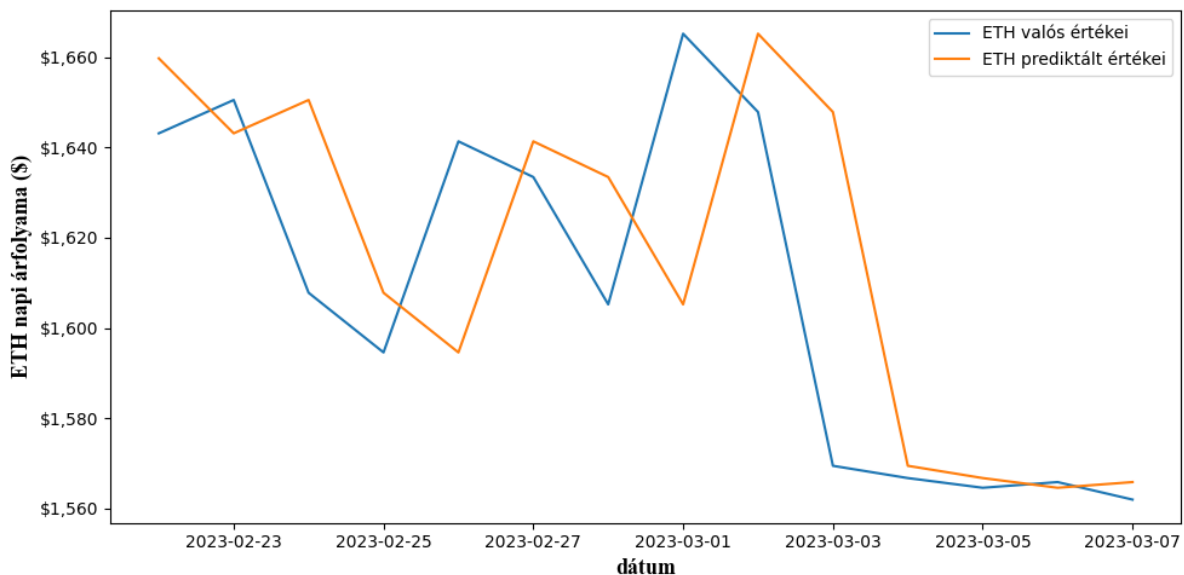
A kutatásom legizgalmasabb része következik, az ether árfolyamának előrejelzése. Rengeteget olvastam idősoros előrejelzési technikákról, statisztikai és gépi tanulási modellekről. Előzetesen bemutatom a felhasznált módszertanokat a legegyszerűbbtől kezdve a legkomplexebb neurális hálóig. Egyes modellek figyelembe veszik a trend és szezonális tényezőket, más modelleknél szükséges az adatok statisztikailag állandóságát (stationary) kialakítani. A legtöbb modell az árfolyam változását prediktálja, nem pedig az árfolyamot. A kiértékelés során végső soron a prediktált árfolyamot fogom összehasonlítani a valódi árfolyammal. Az adatokat **tanulási**, **validációs** és **valós (tesztelési)** adathalmazokra osztottam fel, a tanulási adatszetten tanulnak be a modellek, a validációson tesztelem és finomhangolom őket, majd a valós adatokra előrejelzek.

Naiv előrejelzés

A naiv előrejelzés az egyik legkönnyebben megérthető módszer, amely azon a feltételezésen alapul, hogy egy változó jövőbeli értéke megegyezik a jelenlegi értékével. Más szóval, a következő időszakra vonatkozó előrejelzés nem más, mint a jelenlegi időszakban megfigyelt érték *csúsztatott* értéke.

Ezt a módszert gyakran használják viszonyítási alapként (base line model) a kifinomultabb előrejelzési módszerek összehasonlításához. Bár a naiv előrejelzés könnyen megvalósítható, nem mindig pontos, különösen akkor, ha az adatokban vannak olyan mögöttes trendek, nagyfokú volatilitások vagy minták, amelyek nehezen vehetők figyelembe.

36. ábra: Naiv előrejelzés példa – Validációs adathalmazra



Forrás: Saját készítés ([naive forecast.ipynb](https://naive-forecast.ipynb))

Az 36. ábrán láthatjuk a jobbra elcsúsztatott predikciót, érdekes megjegyezni azt, hogy ez a módszertan sok esetben jobban teljesített, mint a következő versenytársai.

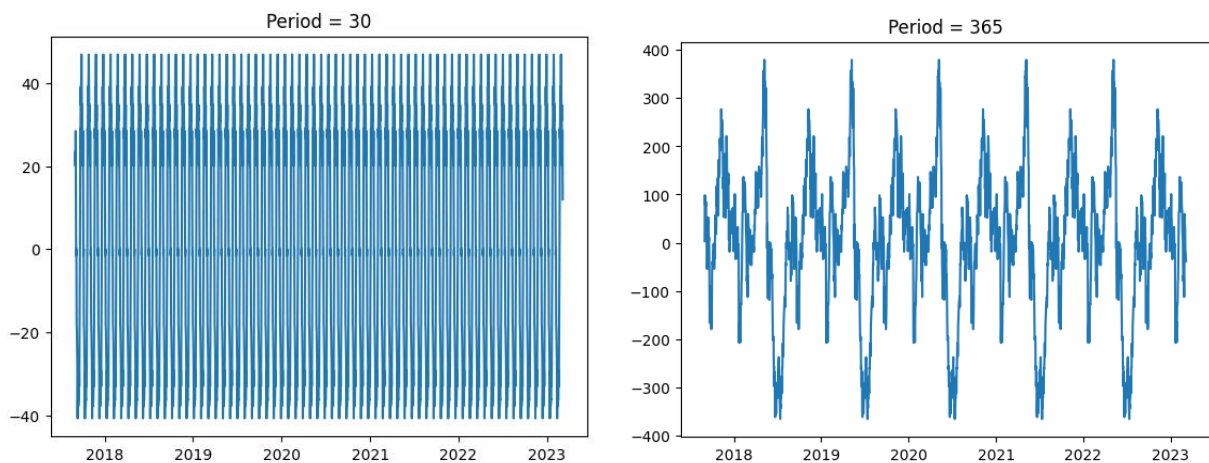
Holt-Winters módszer

A Holt-Winters-módszer, más néven hármas exponenciális simítás egy statisztikai előrejelzési technika, amely során az idősor szezonálisát és trendjellegét megragadva prediktálunk a jövőre. A módszertan az egyik legnépszerűbb megközelítése, különösen a szezonális (pl. repülő utasok száma), vagy lineáris trendet (pl. egy ország villamosenergia-termelése az elmúlt 60 évben) mutató adathalmazok esetén.

A módszer az idősor három összetevőjét veszi figyelembe: az adatok szintjét (átlagérték), trendjét (irányváltást) és szezonálisát (ismétlődő mintázatokat). Az első egyenlet, amelyet

a szintkomponens kiszámításához használnak, a múltbeli megfigyelések súlyozott átlagán alapul. A második egyenlet a trendkomponens kiszámításáért felelős, az aktuális szint és az előző szint közötti különbségen alapul. A harmadik egyenlet, amelyet a szezonális komponens kiszámításához használnak, az aktuális megfigyelés és az előző évek megfelelő szezonjának átlagértéke közötti különbségen alapul. Mind a trend, mint a szezonális hatást fejthet ki az árfolyamra, így elemzésem során mindkét megközelítésre teszteltem. Először a szezonális periódusát/ciklusát próbáltam meghatározni.

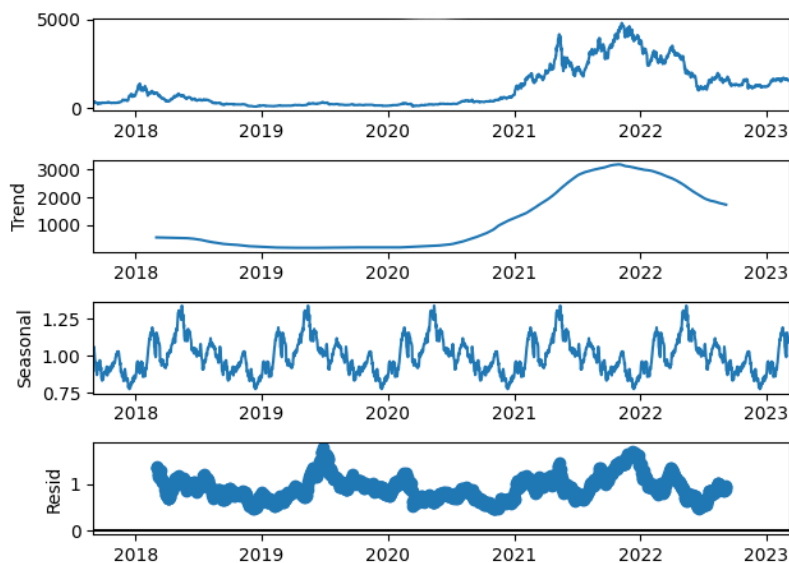
37. ábra: Szezonális tesztelése



Forrás: Saját készítés ([holt_winters_model.ipynb](#))

Habár az előzetes szezonális tesztek során megfigyeltem 14 napos, 30 napos és 365 napos ciklusokat, az előzetes alfejezetek során láthattuk, hogy az ether árfolyamában nem jellemző a szezonális.

38. ábra: Szezonális és Trend dekompozíciója (éves periódussal)

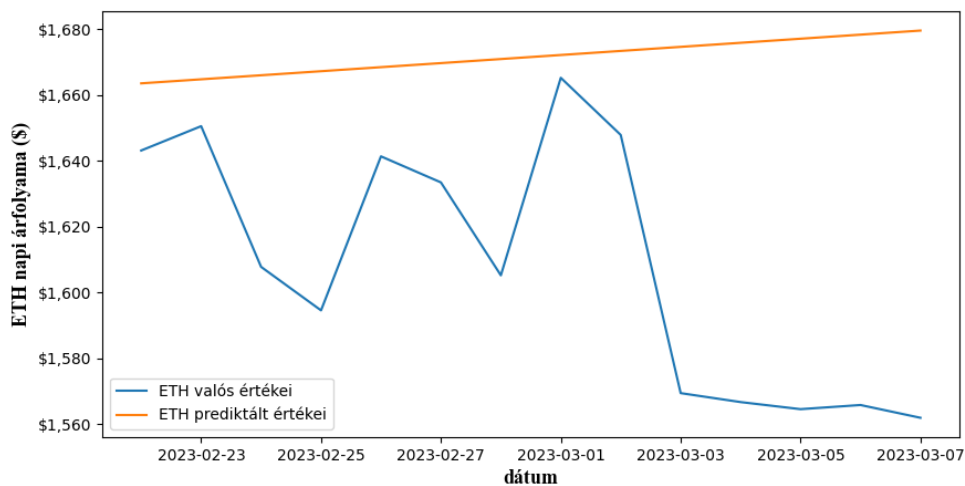


Forrás: Saját készítés ([holt_winters_model.ipynb](#))

Ezután az árfolyam dekompozícióját végeztem el 365 napos periódusra. Legfelül a tényleges árfolyamot, alatta az exponenciális trendet, szezonális ciklikusságát, majd a reziduálisokat (fehér zaj) láthatjuk. A módszertan lényege az exponenciálisan "kisimító" egyenletek (szint, trend és szezonális) kombinálásával előrejelzés készítése a következő időszakra.

A Holt-Winters-módszer különösen hasznos olyan idősoros adatok esetében, amelyek trend- és szezonálisitást is mutatnak. Feltételezi azonban, hogy az idősor három összetevője időben stabil, így nem biztos, hogy jól működik olyan adatok esetében, amelyek hirtelen változásoknak vagy szabálytalan mintáknak vannak kitéve. Így ahogyan az alábbi ábrán látható, gyors és erős volatilitásokra, trendváltásokra a metódus nem képes reagálni.

39. ábra: Holt-Winters módszer – Validációs adathalmazra



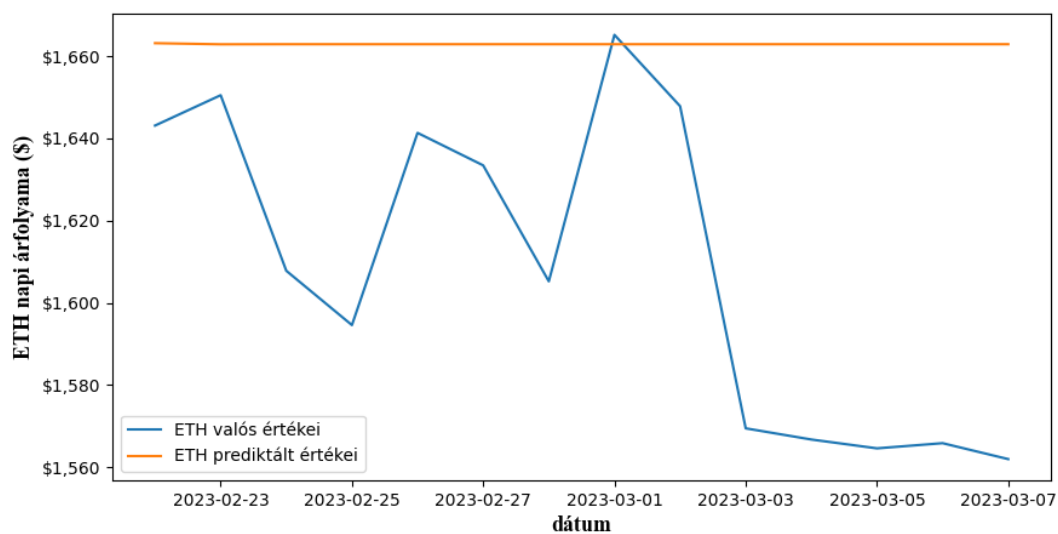
Forrás: Saját készítés ([holt_winters_model.ipynb](#))

ARIMA modell

Az ARIMA (Autoregresszív integrált mozgóátlag) egy olyan statisztikai modell, amely az **autoregresszió**, a **differenciálás** és a **mozgóátlag** fogalmát ötvözi. Az **autoregresszió** jelentése ebben az esetben az, hogy a vizsgált idősor jövőbeli értékét a múltbeli értékeinek segítségével jelezzük előre. Az ARIMA modellek kapcsán ez úgy történik, hogy a sorozatnak egy vagy több késleltetett értékét is bevonjuk a modellbe. A **differenciálás** célja a trend eltávolítása az egymást követő megfigyelések közötti különbség kiszámításával, amely többször is elvégezhető, amíg egy statisztikailag állandó (stationary) sorozatot nem kapunk. A **mozgóátlag** az idősor gördülő átlagának kiszámítására utal egy fixált méretű "időablakon" keresztül, ami segít a volatilitások kisimításában. Az ARIMA-modellt így három paraméter határozza meg: p , amely az autoregressziós szám, d , amely az adatsor differenciálását jelöli és

q, amely a mozgóátlaghoz használt időablak mérete. Az ARIMA-modell az idősor múltbeli értékeinek segítségével képes jövőbeli értékek előrejelzésére, különösen hasznos, ha az idősor szezonalitást vagy trendeket mutat. Arra is használható, hogy modellezze és eltávolítsa a szezonális vagy trendkomponensek hatásait egy idősorból, így hasznos lehet az adatelemzésben és a döntéshozatali folyamatokban. Az ARIMA modellekhez készült egy beépített `auto_arima` funkció, amely segítségével meghatározhatjuk a p, d és q optimális értékét iterációkon keresztül. Tanulási és validálási adatszeteknél normális és logaritmikus értékeket használtam az árfolyam prediktálására.

40. ábra: Arima modell előrejelzés – Validációs adathalmazra



Forrás: Saját készítés ([arima_model.ipynb](#))

Fontos kiemelni, hogy az eddigi módszereknél csakis az árfolyam historikus árát tudtam felhasználni tanulási adathalmazként. Az ARIMA modellek kiegészíthetők exogén idősorokkal, viszont egy árfolyamelőrejelzés kapcsán ezek nem használhatók (nem tudjuk a jövőbeli híreket, véleményeket, más árfolyamokat, indexeket). Lineáris és szezonális modelleknél az egyik legpopulárisabb, részvény és kripto előrejelzéseknél kevésbé használható.

VAR modell

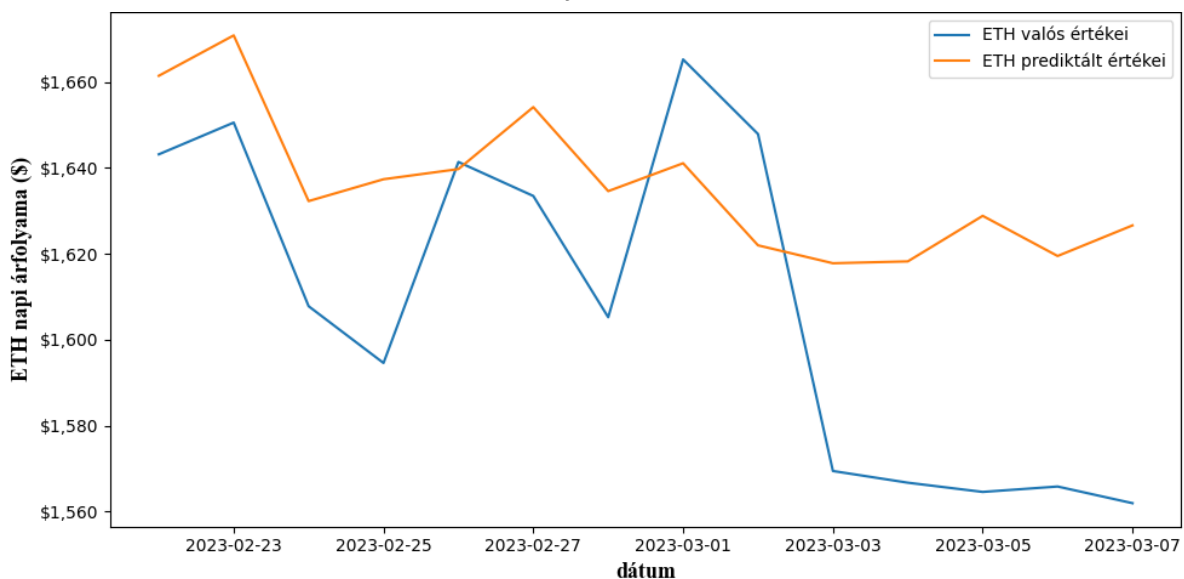
A vektor autoregressziós (VAR) modell egy olyan statisztikai modell, amely képes több idősor közötti összefüggések és kapcsolatok elemzésére, egyfajta sztochasztikus folyamatmodell. Lineáris függvényeként modellezi a célzott változó késleltetett (múltbeli) értékeit, a modellben szereplő többi változó késleltetett értékeit és a hibák („fehér zajok”) vektorát. Fontos megjegyezni, hogy érdemes előzetes vizsgálatokat (Granger okozati teszt) folytatni arról, hogy mely idősorokról feltételezhető az, hogy idővel hatással vannak egymásra.

A modellparaméterek becslése lineáris egyenletrendszer megoldását jelenti. A p (késleltetési érték) kiválasztása, amely a jövőbeli értékek előrejelzéséhez használt múltbeli megfigyelések számát határozza meg, általában statisztikai tesztek vagy egyszerű próbálgatások útján választják ki. Kódolás során a „maxlags” argumentum segítségével a tudjuk meghatározni késleltetések maximális számát (p).

Miután a VAR-modell betanult az adatokon, előrejelzések készíthetők az endogén változók múltbeli értékeire történő rekurzív alkalmazásával. Az előrejelzési horizont a becsléshez használt adatok hosszától és a prediktált periódusok számától függ.

A VAR-modellnek van néhány korlátja, például a linearitás feltételezése az adathalmazok között és az időben állandó idősorok kialakítása. Mindazonáltal egy nagyon népszerű és széles körben használt modell előrejelzésekhez, különösen a több egymással összefüggő változót tartalmazó rendszerek esetében. Gyakran használják a közgazdaságtanban és a természettudományokban.

41. ábra: VAR modell előrejelzés – Validációs adathalmazra



Forrás: Saját készítés ([var_model.ipynb](#))

XGB Regressziós modell

A VAR modellt követően az XGBoost Regressor gépi tanulási modelljét próbáltam ki a validációs adathalmazon. De milyen is az XGBoost modell, hogyan működik, milyen előnyei vannak?

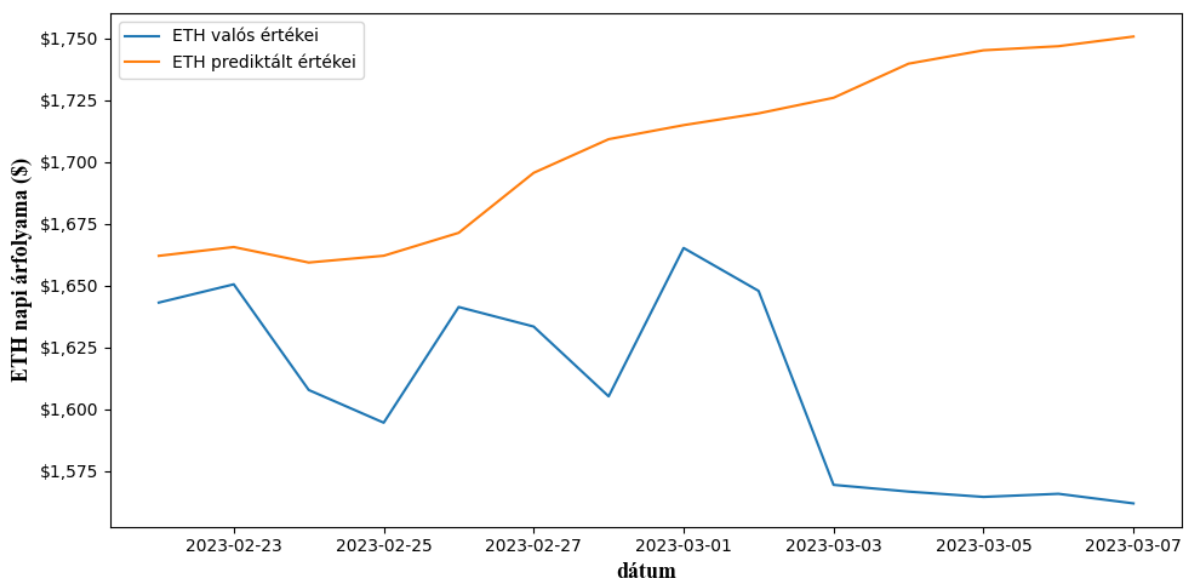
Általánosan elmondható, hogy egy gépi tanulási modell célja az, hogy a tanulási folyamat során általánosságban megtanulja az összefüggéseket és mintázatokat. Lényegében szeretnénk

egy generalizálást végrehajtani és elkerülni azt, hogy a modell túltanulja vagy ne tanulja meg a tanulási adatbázist.

Az XGBoost (Extrém gradiens erősítés) egy rendkívül erős és ismert gépi tanulási modell, előszeretettel használják klasszifikációra és regresszióra különböző versenyeken és megmérettetésekben. A gradiens erősítés egy olyan technika, amely több „gyengén” tanuló, például döntési fák kombinálásával egy erősebb modellt hoz létre. Az XGBoost a gradiensnövelő algoritmus egy speciális típusát használja, amely a veszteségfüggvényt egy sor iteráción keresztül optimalizálja, és minden egyes iteráció egy új döntési fát ad a modellhez. Hihetetlenül jól skálázható, hiszen akár több millió soros és oszlopos (dimenziós) adatokat is képes megtanulni rövid idő alatt. Az adatok „túltanulásának” megelőzésére rengeteg regularizációs paramétert tudunk beállítani, ilyen például a gamma, reg_lambda, reg_alpha kulcsargumentumok, vagy a tanulási ráta. Mindezek mellett párhuzamos futtatásokat állíthatunk be, így időt takaríthatunk meg a tanulás folyamán. Amennyire személyre szabható a modell és paramétere, annyira nehéz megtalálni a legoptimálisabb kombinációt.

A folyamat során előzetesen tanulási adathalmazt alakítottam ki úgy, hogy minden egyes idősor késleltetett időben állandó adatait kapcsolom a jelenlegi árfolyamértékhez. Így a modell képes megtanulni a múltbeli adatok hatását és befolyását. Ezután egységes skálára hoztam őket, hiszen néhány százalék alapú (pl. Twitter sentiment), néhány darabszám, nagy része pedig amerikai dollár alapú. A magyarázóváltozók erősségét megvizsgálva az ether-hez kapcsolódó nyitás, zárás, mozgóátlagok, technikai indikátorok mellett a NASDAQ, S&P500 és bitcoin értékeit használtam előrejelzéshez. Inkrementális stratégiát használtam előrejelzés során, magyarul minden előrejelzéshez az előző előrejelzés értékeit használtam.

42. ábra: XGBRegressor modell előrejelzése - Validációs adathalmazra



Forrás: Saját készítés ([xgb_model.ipynb](#))

Az 42. ábrán látható, hogy a modell erőteljesen felülprediktálja a validációs adathalmazon belüli időszakot. Érdeemes lesz számokban megvizsgálni a különbséget a fejezet végen az „Értékelés” szakaszban.

LSTM neurális háló

Végül, de nem utolsó sorban egy neurális hálót LSTM (Long-Short Term Memory) építettem fel az árfolyamra. Teszteltem a Tensorflow modulból származó rétegeket mind inkrementális, mind többnapos előrejelzési módszerekre, viszont a legjobb eredményt a Pytorch moduljával tudtam elérni.

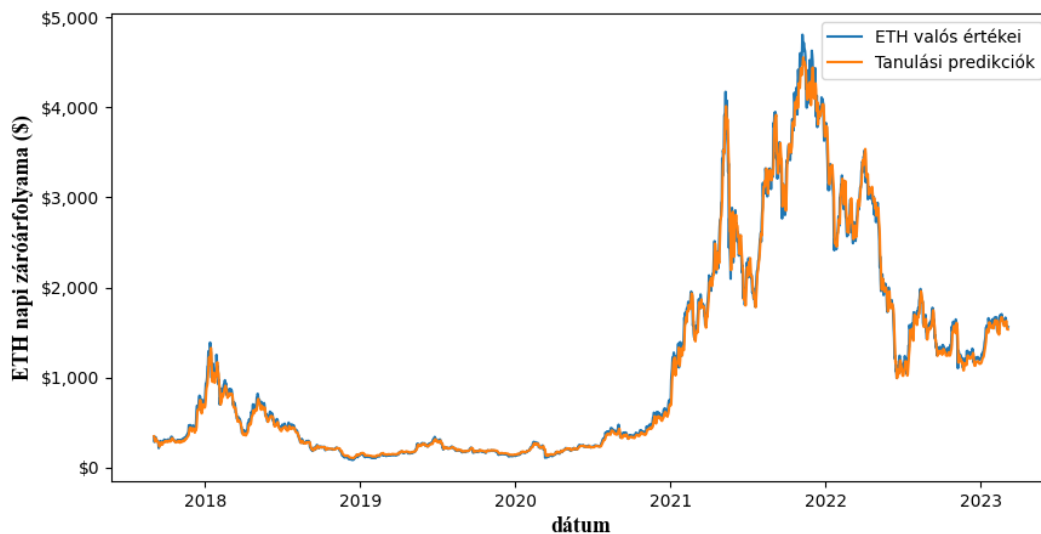
Az LSTM egy olyan típusú rekurrens neurális hálózat, amely memóriacellákat és ”kapukat” használ arra, hogy hosszú távon is szelektíven emlékezzen a legfontosabb információkra és adatokra. Az információáramlást három kapu irányítja a hálózatban, az egyik legfontosabb a felejtési kapu, amely feladata az irreleváns adatok elfelejtése és eltüntetése (szigmoid funkcióval és szorzással képes). A bemeneti kapu dönti el azt, hogy milyen új információkat tárol el a háló a folyamat során. Az aktuális bemenetet és az előzetes rejtett cellákból származó (el nem felejtett) adatokat veti össze (szigmoid és tanh funkciókkal) és kreál kimeneteket a neuronok között. A harmadik kapu a kimeneti kapu, amely az előző bementi kapuból származó értéket veti össze az aktuális bemenettel és gyárt újabb inputokat a rejtett cellákba.

Ezek a kapuk segítenek az LSTM modellnek szelektíven tanulni és emlékezni vagy elfelejteni a bemenetek hosszú sorozata alatt. A sejtek állapota a múltbeli bemenetek

"memóriájaként" működik, a kapuk határozzák meg, hogy mennyi múltbeli információt tartsanak meg, és mennyi új információt adjanak hozzá. Ez teszi az LSTM-et népszerű választássá a szekvenciális adatokat tartalmazó feladatokhoz, például előrejelzések mellett természetes nyelvi feldolgozáshoz és a beszéd felismeréshez.

A folyamat során az ETH múltbeli árfolyamával tudtam elérni a legjobb eredményeket, az alábbi ábrán a tanulási minőségét/eredményét láthatjuk:

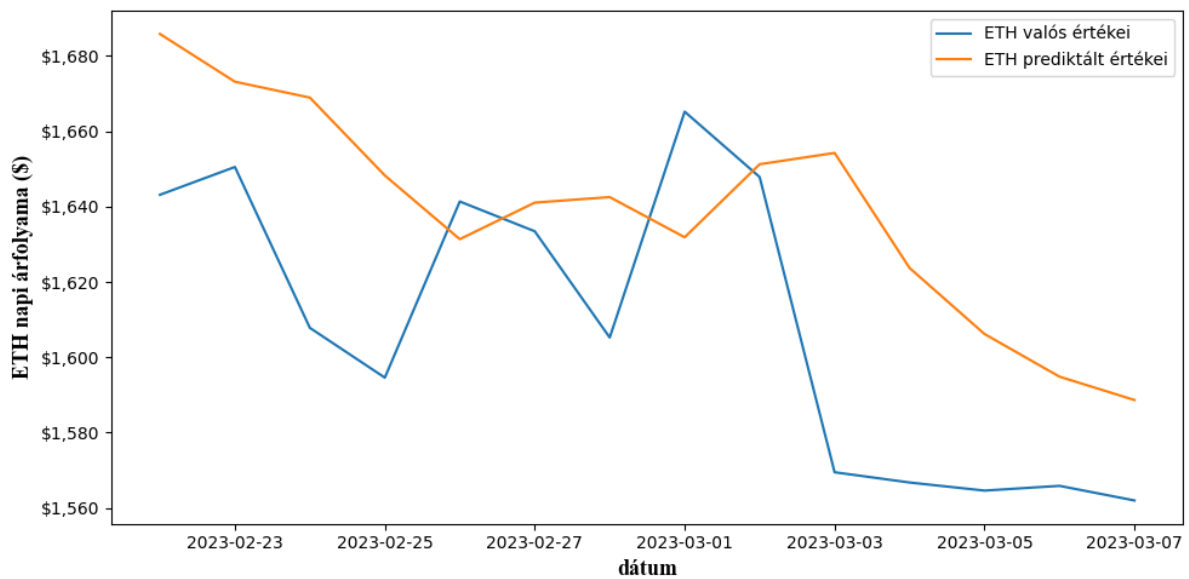
43. ábra: LSTM tanulási eredménye



Forrás: Saját készítés ([lstm_model.ipynb](#))

Elmondhatatlanul jól megtanulta az ether erősen volatilis historikus árfolyamát a modell, a validációs adatokra következett a tesztelés. A 44. ábrán a 14 napos előrejelzés láthatjuk, a neurális háló jól leköveti a trendet, érzékelteti az árfolyam esését (picit késleltetve) február végén és március elején. Kissé felülprediktál érték szempontjából, viszont jól megfogja a következő időszak trendjét.

44. ábra: LSTM modell előrejelzése - Validációs adathalmazra

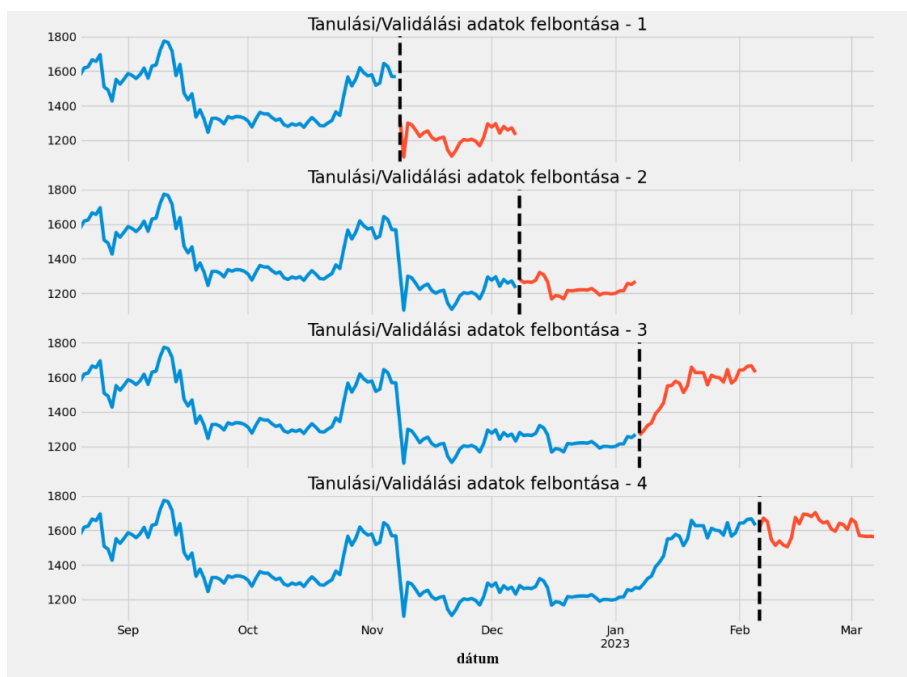


Forrás: Saját készítés ([lstm_model.ipynb](#))

Értékelés

A következőkben értékeljük minden egyes modell eredményét számokban. Az adatok tanulási-validálási adatokra bontásánál **10 periódust** vizsgáltam, így 1 napos előrejelzés esetén 10-szer 1 napra vizsgáltam a pontosságot, 7 napos előrejelzésnél pedig 10-szer 7 napos periódusokra, ahogyan a 45. ábrán is látható.

45. ábra: Tanulási/Validálás adatok felbontása



Forrás: Saját készítés ([xgb_model.ipynb](#))

Értékeléshez az átlagos négyzetes hibát, átlagos abszolút hibát és átlagos abszolút százalékos eltérést számoltam ki. A hangsúlyt az átlagos négyzetes hibára és annak négyzetgyökére helyeztem, hiszen az árfolyam gyakori magas volatilitására érzékenyebb a mérőszám, így jobban mérhető a predikció pontossága.

46. ábra: Előrejelzések átlagos négyzetes hibájának gyöke (RMSE)

előrejelzési hossz	Naiv	Holt-Winters	Arima	Var	XGB	LSTM
1 nap	24,8681	24,4040	24,5118	8.3536	28,5660	27,4513
3 nap	38,0148	50,2007	49,6641	11.5543	53.3714	51,6113
7 nap	38,8613	77,9358	78,0130	44,6529	64,3082	44,6416
14 nap	46,2068	121,8498	119,8629	39,5418	118,8046	58,0182

Forrás: Saját készítés

Az átlagos négyzetes hibák négyzetgyöke alapján rövidtávon a **VAR modell** teljesített a legjobban, míg hosszútávon a **Naiv előrejelzési** módszer. Érdeemes megjegyezni azt, hogy az elmúlt időszakban kevésbé volt volatilis az ETH árfolyama, így ennek is köszönhető a naiv előrejelzés sikeressége.

47. ábra: Előrejelzések átlagos abszolút százalékos eltérése (MAPE)

előrejelzési hossz	Naiv	Holt-Winters	Arima	Var	XGB	LSTM
1 nap	1,5384%	1,5495%	1,5192%	0,5348%	1,7789%	1,7217%
3 nap	1,9817%	2,9862%	2,9371%	0,7309%	3,1182%	2,8667%
7 nap	1,9302%	4,6489%	4,6761%	2,7217%	3,8625%	2,3935%
14 nap	6.6330%	7,3146%	7,3284%	2,1999%	7,5983%	3,1437%

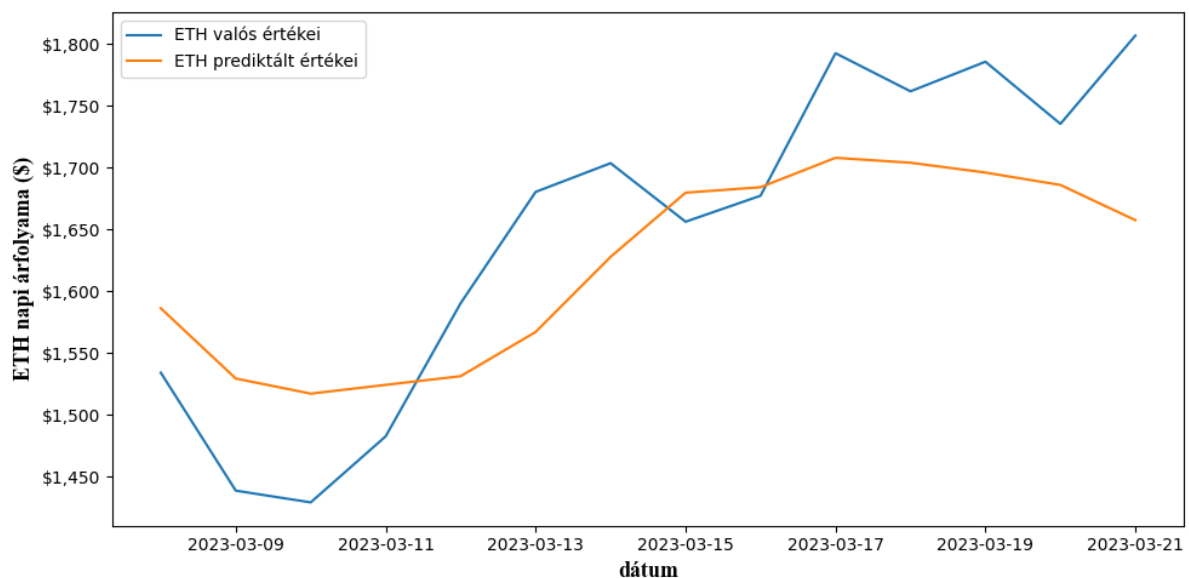
Forrás: Saját készítés

Az átlagos abszolút százalékos eltérések alapján a **VAR modell** kimagaslóan a legjobb volt, csak 7 napos predikciók esetén tudta leverni a Naiv előrejelzés a modellt. Az eredmények alapján **megcáfolom** a harmadik hipotézisemet, amely így szól: **Neurális háló segítségével pontosabban előrejelezhetjük az ether jövőbeli árfolyamát, mint konvencionális statisztikai és gépi tanulási modellekkel.** A megcáfoláshoz viszont egy **olyan kitévelt** tennék, hogy a neurális hálók komplexitása hihetetlenül bonyolult, elmondhatatlanul sok időt ölel fel a megfelelő rétegek, kombinációk megtalálása. Mindezek mellett az ábrák alapján a neurális háló

tudta a legjobban lekövetni a jövőbeli trendet, extrém volatilitásnál a naiv előrejelzés nem tudna olyan jól érvényesülni, mint a tesztek alapján. Megfelelő magyarázóváltozók és a legjobb opciók használatával a neurális hálóval történő előrejelzések **jelentősen javíthatók!** A jövőben szeretnék egy teljeskörű neurális háló elemzést is végezni az árfolyamra további idősorokkal kiegészítve.

Végül, de nem utolsó sorban vessünk egy pillantást a **valósan jövőbeli adatok** predikciójára. Mi történt volna, ha a kutatás pillanatában előrejeleztem volna a még nem ismert adatokra? A tanuló adathalmazom 2023. március 7-ig terjed ki, így a március 8-tól kezdődő adatokat még nem látta/láthatta a modell. Az alábbi ábrán szemlélhető az LSTM modell előrejelzése a következő 14 napra.

48. ábra: LSTM predikció az elkövetkező 14 napra



Forrás: Saját készítés ([lstm_model.ipynb](#))

A modell előrejelzése helyesen lekövette a 14 nap mozgását és tendenciáját, viszont nem annyira jól az ether tényleges jövőbeli értékét. Viszont, ha a trendet követve befektettem volna, akkor plusz 15%-os növekedést tudtam volna elérni az előrejelzés segítségével.

3 Összegzés

Kutatásom során három hipotézisemre kerestem a választ, az első hipotézisemre mélyinterjút készítettem magyarországi blokklánc szakértőkkel. Kíváncsi voltam arra, hogy a decentralizált pénzügyi cégek képesek-e felvenni a versenyt a kereskedelmi bankok által nyújtott pénzügyi szolgáltatásokkal. A szakértők elmondása alapján rendkívül populáris a hitelnyújtás DeFi cégeknél, jelenleg is több tíz milliárd dollár értékű hitelállománnyal rendelkeznek, így az első hipotézisem bebizonyosodott.

A kutatásom második felében adatalapú elemzéseket, vizsgálatokat folytattam le az Ethereum rendszer kriptovalutájára, az ether-re (ETH). Előzetesen rengeteg Python kódot írtam adatgyűjtéshez, több mint 50 idősort sikerült megszereznem a vizsgált periódusra. Az adatokat szekciókra bontottam, a Bázis szekció tartalmazza az ether-re vonatkozó alapadatokat, mint nyitó, záró árfolyam, napi ingadozás mértéke, volumene stb. A Szekció 1-ben a világgazdaságot meghatározó árfolyamokat, indexeket töltöttem le. Ilyen volt a bitcoin értéke, az S&P500, NASDAQ, és amerikai kincstárjegyek historikus értékei. Ezen kívül az amerikai inflációs várakozásokat, kincstárjegyek kamatainak különbözetét, az S&P500 ingatlan szektor indexét, arany és ezüst árfolyamát sikerült összegyűjteni. Korrelációs és trendvizsgálat során a bitcoin árfolyama erős szignifikanciát mutat az ether-el, a NASDAQ, S&P500 mellett az arany és ezüst árfolyama közepesen erős vagy gyenge szignifikanciát mutatott. A Szekció 2 adatai között az Ethereum rendszerével kapcsolatos híreket, kommenteket és bejegyzéseket analizáltam, és vizsgáltam meg árfolyambefolyásoló hatásukat. Több, mint 25 millió ether vagy ETH vagy Ethereum kulcsszót tartalmazó Twitter bejegyzést értékeltem ki és vizualizáltam hangulatindexüket az árfolyammal. Emellett a keresési trendek és bálnák bevásárlásait vettem össze az árfolyam alakulásával. Korrelációs, kereszt-korrelációs és Granger hatás elemzést végeztem a második hipotézisem végett. Az eredmények alapján gyenge korrelációt tapasztaltam az árfolyam és a hírek, kommentek változása között, Granger hatás szerint pedig lineáris előrejelzéshez se alkalmasak a hangulatelemzések, így a második hipotézisemet megcáfoltam. Szekció 3 során az árfolyamra vonatkozó technikai indikátorokat, mozgóátlagokat elemeztem és vizualizáltam.

A harmadik hipotézisemre előrejelzési modelleket készítettem Python segítségével. Alap statisztikai modellektől kezdve a neurális hálógig, mindent több periódusra teszteltem, több hibamérő módszer és eljárás alapján. A legjobbnak a Vector AutoRegression (VAR) modell bizonyult a Naiv előrejelzés mellett. Fontos kiemelni, hogy az árfolyam utóbbi időszakban oldalazott, így érdekesebb lehet a következő kutatás során nagyobb időtávra kiterjeszteni a

tesztelést. A neurális háló segítségével még valós előrejelzést is készítettem a vizsgált időszakot követő periódusra, a tendenciát sikeresen lekövetve a modell több mint 15%-os növekedést tudott volna kölcsönözni befektetésemnek.

Kutatásom és témaválasztásom felnyitotta a szememet, revolúció folyamatát tapasztalhatjuk a világgazdaságban. Rendkívül felgyorsultak a fejlesztések a modern kriptográfiában és kriptovaluták világában. A jövőben többet szeretnék foglalkozni gépi tanulási modellekkel, neurális hálókkal és természetesen kriptovalutákkal.

Mellékletek

Mélyinterjú kérdések

1. Az Ethereum hálózata az egyik legnépszerűbb mind a befektetők, mind a laikusok számára. Ön mit gondol a közösség, központi bankok (pl. FED, ECB) hozzáállásáról az Ethereum kapcsán?
2. Sokan a web3 korai úttörőjének tartják a világ számítógépéként emlegetett Ethereum rendszerét, Ön szerint mennyire állja meg és fogja megállni ez az állítás a helyét?
3. A jövőben már akár telefonunkon is képesek leszünk tranzakciót titkosítani, validálni, akár stakelni, és a hálózat egy csomópontjaként működni. Ön szerint legközelebb mikorra érhetjük el ezt az állapotot?
4. Ön szerint 10 év múlva az Ethereum mennyire lesz elterjedt világszerte? Képes lesz-e felvenni a versenyt a kereskedelmi bankok által nyújtott szolgáltatásokkal?
5. Ön szerint mely szférákban tudnak domináns szereplőként fellépni az Ethereum alapú okosszerződések? Előfordulhat akár egy éttermi rendelés (adatalapú reakciókkal mindkét oldalról) megoldása okosszerződések keresztül?
6. Számos okosszerződést, DeFi-t hackeltek meg az elmúlt években (Poly Network, Ronin), milyen hatással lehet mindez (akkor is, ha nem az Ethereum-on történt) az Ethereum hálózatára és az ether árfolyamára?
7. Magyarországon mennyire terjedtek el az okosszerződések? Esetleg tud mondani néhány példát?
8. Milyen jövőt lát az okosszerződések előtt általánosságban?