BI NHF (WUPA9P)

A projekt célja az Android és iOS operációs rendszerű mobiltelefonok összehasonlítása. Kettő különböző adatforrás alapján próbál különbségeket keresni a két operációs rendszer között.

A projekt MSSQL adatbázis szervert használ az adatok tárolásáért, mint data warehouse. Az ETL folyamatokat az Integration Service (SSIS) segítségével kerültek megvalósításra. A reportok készítését PowerBI asztali alkalmazás biztosítja. Az adatelemzés Google Colab Notebook-on valósult meg Pandas és Jupyter könyvtárak segítségével.

A házi feladat során készített **videó** ezen a linken található.

A következő bekezdésekben a pontozásnak megfelelően dokumentálom a házi feladat elkészítése során elvégzett feladatokat.

ETL

Kétféle adatforrásból dolgozok.

- Első adatforrás
- Második adatforrás

Első adatforrás

Adatforrásokból való betöltése célterületre

Az első adatforrás különböző operációs rendszerű telefonok listáját tartalmazza. Az adatforrás számos metrikával rendelkezik, mindegyike nem kerül tárolásra azonban. A tárolásra került adatokat a következő kép mutatja, mely tartalmazza a tábla létrehozásához szükséges sql parancsot.

create_table_smart...2001M1\Martin (78))*

```sql
CREATE TABLE SmartPhone (
    device_id INT IDENTITY(1,1) PRIMARY KEY,
    brand NVARCHAR(1000),
    model NVARCHAR(1000),
    price INT,
    avg_rating DECIMAL(3,1),
    is_5g BIT,
    num_cores INT,
    processor_speed DECIMAL(3,2),
    battery_capacity INT,
    ram_capacity INT,
    internal_memory INT,
    refresh_rate INT,
    os NVARCHAR(50),
    extended_memory_available BIT,
    resolution_height INT,
    resolution_width INT
);
```

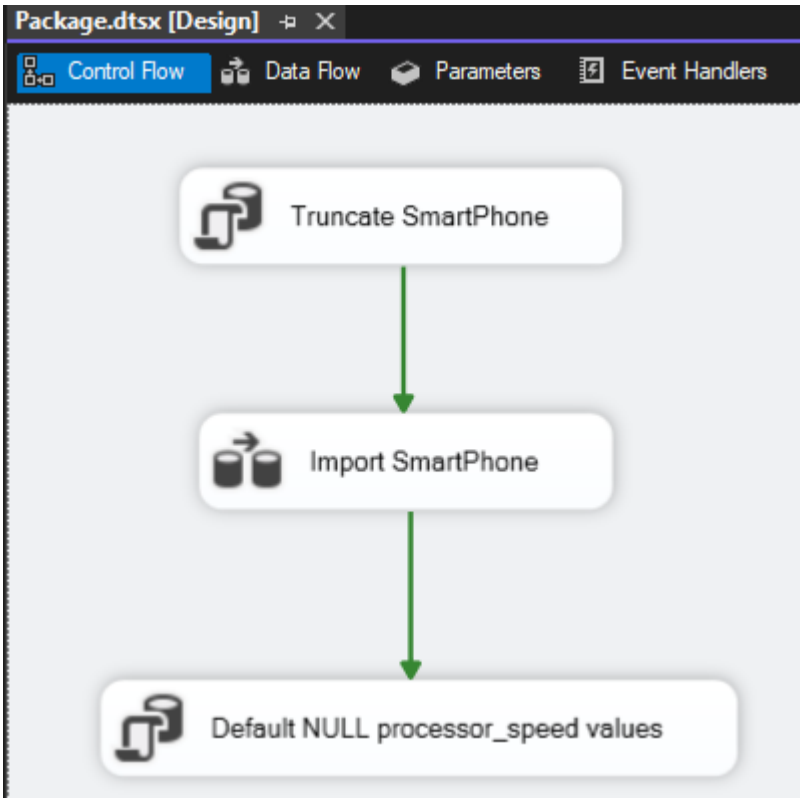
Az adatforrás a következő típusú adatokat tartalmazza általánosságban, melyek a következő típusokban kerülnek tárolásra:

- szöveges --> NVARCHAR
- egész szám --> INT
- valós szám --> DECIMAL
- logikai érték --> BIT

Elsődleges kulcsot nem tartalmaz az adatforrás, így az device_id néven, valamint adatbázis szinten jön létre minden egyes új adat beszúrásakor a táblába.

Transzformáció és ütemezés

A transzformáció három fő lépésból áll, melyet a következő kép is demonstrál.



1. SmartPhone tábla megtisztítása
2. Adatok betöltése a táblába
3. A betöltésre került bizonyos adatok módosítása

Az **első lépés** törli a korábbi adatokat a SmartPhone táblából minden egyes importálás előtt, így garantálva, hogy az adatok naprakészek legyenek. A művelet *Execute SQL Task* komponensben kerül végrehajtásra.

A **második lépés** feldolgozza, átalakítja a .csv fájlból érkező adatokat annak megfelelően, hogy azok probléma nélkül a dwh-ba tárolásra kerüljenek. Ezt egy *Data Flow Task* komponensben kerül megvalósításra. (Erről részletesen később.)

A **harmadik lépés** kicseréli bizonyos (processor_speed oszlopban található) NULL típusú adatok egy alapértelmezett értékre. Ez az alapértelmezett érték *0.00*. Fontos, hogy csak akkor kerül beállításra, ha az *avg_rating* viszont nem NULL típusú. Ez a művelet szintén *Execute SQL Task* komponensben kerül végrehajtásra.

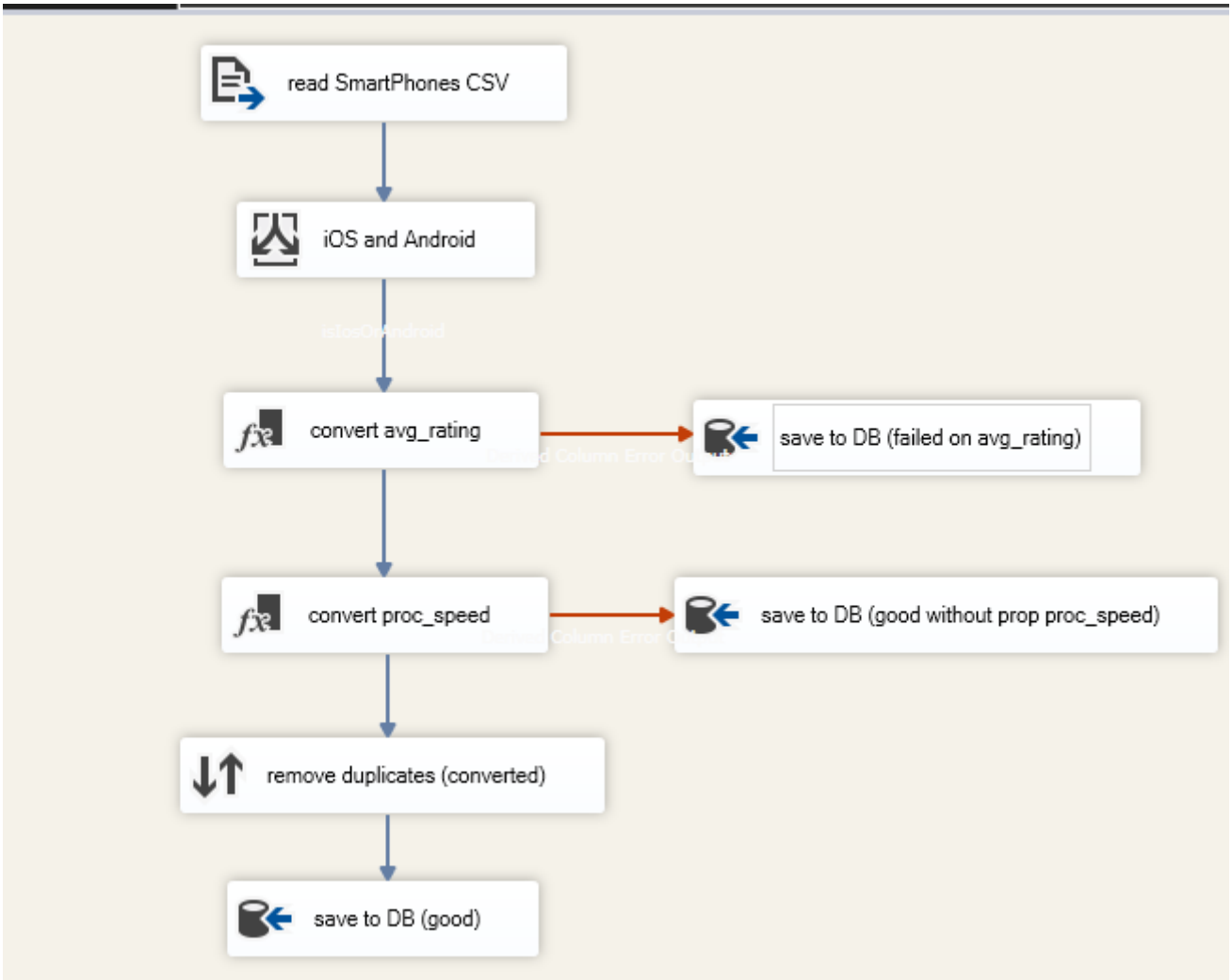
Transzformáció második lépése bővebben

Elsősorban *Conditional Split* komponens segítségével kiszűröm a kizárólag iOS vagy Android operációs rendszerrel rendelkező telefonokat. Mivel nincs szükségünk a más típusúakra, így ezek nem is kerülnek tárolásra.

A szöveges, egész szám és logikai érték típusú adatok betöltése során nincs külön átalakításra szükség. Sajnos a valós számokkal már más a helyzet. A valós számok a .csv fájlban tizedes vessző helyett tizedes ponttal kerültek tárolásra, vagyis konvertálásuk nehezített.

Ennek oka, hogy a *Flat File Source* komponens és az MSSQL adatbázis szerver is valós számok konvertálása esetén tizedes vesszőt vár el. Természetesen ennek fő oka a régió és nyelv jelenlegi beállítása, mely Magyarország (magyar).

Így a valós számok tizedes pontját *Derived Column* komponensben tizedes vesszőre cserélem, illetve, ha nem NULL érték, akkor konvertálásra is kerülnek. A *Data Flow*-t a következő kép mutatja.



Az *avg_rating* és *processor_speed* valós számok külön kerülnek feldolgozásra.

- Amennyiben egy telefonhoz nem tartozik értékelés, úgy a telefonhoz tartozó többi adatot is NULL-ként kezelem, és tárolom el az adatbázisban.
- Amennyiben egy telefonhoz tartozik értékelés, de nem tartozik processzor sebesség, úgy a telefonhoz tartozó többi adat még tárolásra kerül, viszont a processzor sebesség NULL értéket kap. Egy későbbi

szakaszban pedig minden ilyen helyzetben lévő telefon - értékelés van, de sebesség nincs - alapértelmezett értéket kap sebességét illetően.

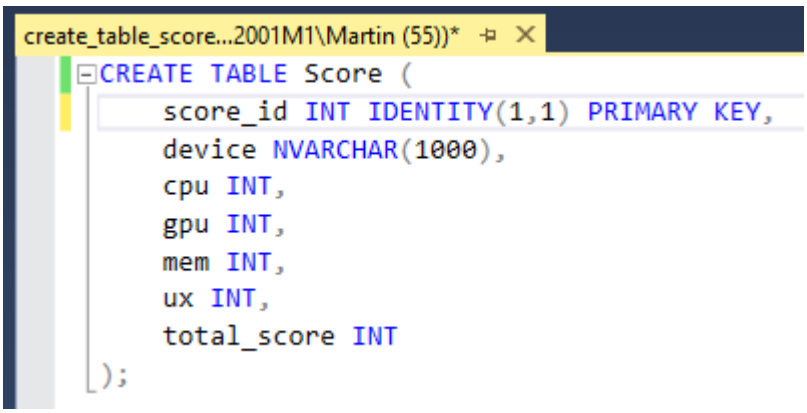
Ezt követően a telefonok *model*-je alapján eltávolítom a duplikált elemeket, illetve rendezem e szerint egy *Sort* komponensben.

Végül az adatok tárolásra kerülnek *ADO NET Destination* komponensben.

Második adatforrás

Adatforrásokból való betöltése célterületre

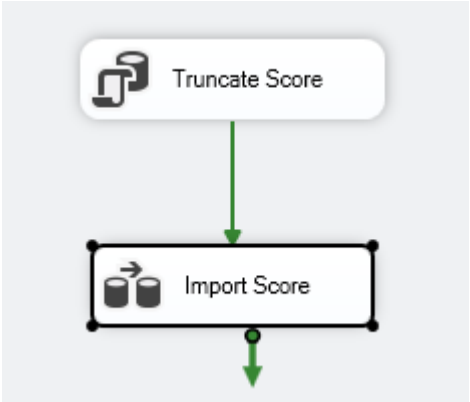
A második adatforrás kizárólag iOS és Android operációs rendszerű telefonokat tartalmaz. Mindegyik telefon négy egyedi értékelés mentén tartalmaz egy összpontszámot, illetve egy ötödik, mely az előzőeket összegzi. Az értékeléseket felhasználók adták.



Az adatforrás az előző adatforráshoz hasonlóan tárol szöveges, egész szám típusú adatokat. Szerencsére itt nincs valós szám típusú, így automatikus konverzióval mindegyik típusú adat könnyedén betölthető az mssql adatbázisba.

Transzformáció és ütemezés

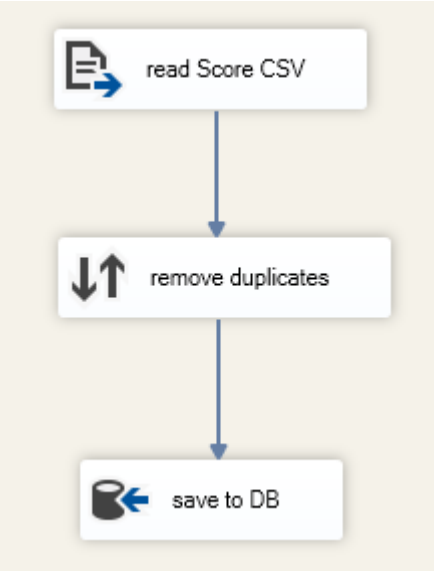
A transzformáció csak kettő fő lépésből áll, melyet a következő kép is demonstrál.



1. Score tábla megtisztítása
2. Adatok betöltése a táblába

Az **első lépés** törli a korábbi adatokat a Score táblából minden egyes importálás előtt, így garantálva, hogy az adatok naprakészek legyenek. A művelet *Execute SQL Task* komponensben kerül végrehajtásra.

A **második lépés** feldolgozza, átalakítja a .csv fájlból érkező adatokat annak megfelelően, hogy azok probléma nélkül a dwh-ba tárolásra kerüljenek. Ezt egy *Data Flow Task* komponensben kerül megvalósításra. Ennek részleteit a következő kép mutatja be.

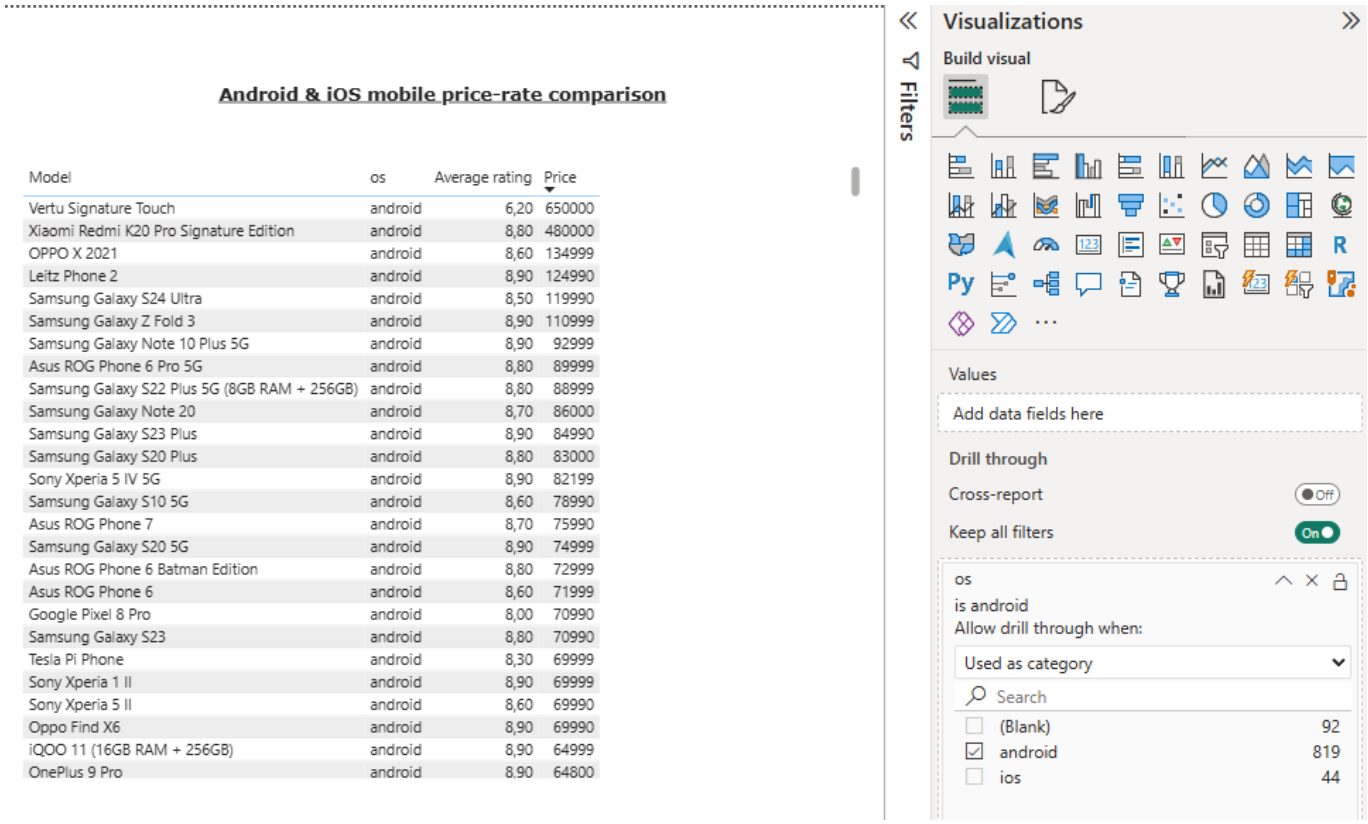


Az előző adatforráshoz hasonlóan történik az adatok importálása is a dwh-ba, azzal a különbséggel, hogy a .csv kiterjesztésű fájl minden adatot tartalmaz, nincs bennük üres/hiányzó érték. Így nincs szükség külön átalakításra egyik adatot se illetően.

Report

Alább láthatók az egyes reportok kifejtve.

1. report

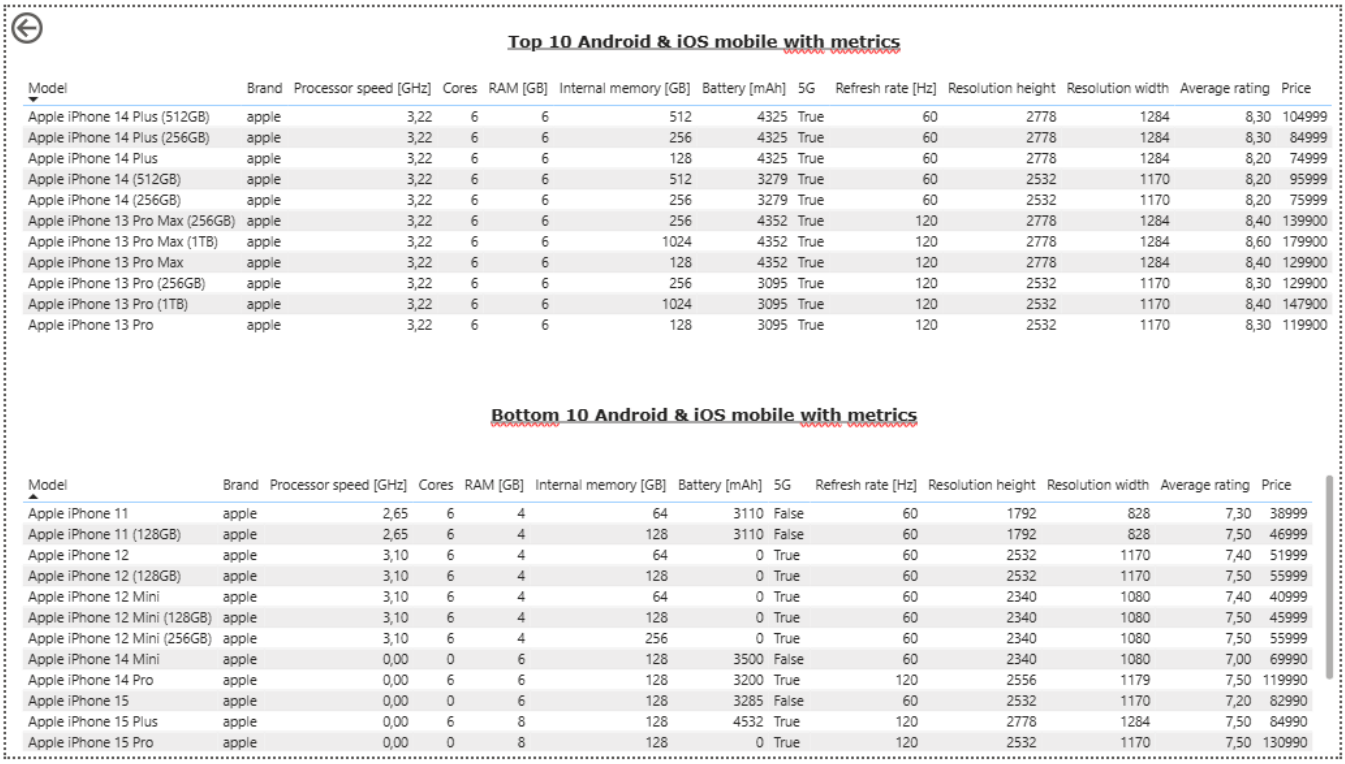


Android & iOS mobile price-rate comparison

Model	os	Average rating	Price
Vertu Signature Touch	android	6,20	650000
Xiaomi Redmi K20 Pro Signature Edition	android	8,80	480000
OPPO X 2021	android	8,60	134999
Leitz Phone 2	android	8,90	124990
Samsung Galaxy S24 Ultra	android	8,50	119990
Samsung Galaxy Z Fold 3	android	8,90	110999
Samsung Galaxy Note 10 Plus 5G	android	8,90	92999
Asus ROG Phone 6 Pro 5G	android	8,80	89999
Samsung Galaxy S22 Plus 5G (8GB RAM + 256GB)	android	8,80	88999
Samsung Galaxy Note 20	android	8,70	86000
Samsung Galaxy S23 Plus	android	8,90	84990
Samsung Galaxy S20 Plus	android	8,80	83000
Sony Xperia 5 IV 5G	android	8,90	82199
Samsung Galaxy S10 5G	android	8,60	78990
Asus ROG Phone 7	android	8,70	75990
Samsung Galaxy S20 5G	android	8,90	74999
Asus ROG Phone 6 Batman Edition	android	8,80	72999
Asus ROG Phone 6	android	8,60	71999
Google Pixel 8 Pro	android	8,00	70990
Samsung Galaxy S23	android	8,80	70990
Tesla Pi Phone	android	8,30	69999
Sony Xperia 1 II	android	8,90	69999
Sony Xperia 5 II	android	8,60	69990
Oppo Find X6	android	8,90	69990
iQOO 11 (16GB RAM + 256GB)	android	8,90	64999
OnePlus 9 Pro	android	8,90	64800

Az első report bemutatja az android és iOS telefonokat ár-érték arányban. Dinamizmusa, hogy az adatokat lehet rendezni modell, átlagos értékelés és ár szerint is növekvő vagy csökkenő sorrendben. Amennyiben csak az egyik operációs rendszerre vagyunk kíváncsiak, úgy leszúrás segítségével ez szintén beállítható.

2. report



Top 10 Android & iOS mobile with metrics

Model	Brand	Processor speed [GHz]	Cores	RAM [GB]	Internal memory [GB]	Battery [mAh]	5G	Refresh rate [Hz]	Resolution height	Resolution width	Average rating	Price
Apple iPhone 14 Plus (512GB)	apple	3,22	6	6	512	4325	True	60	2778	1284	8,30	104999
Apple iPhone 14 Plus (256GB)	apple	3,22	6	6	256	4325	True	60	2778	1284	8,30	84999
Apple iPhone 14 Plus	apple	3,22	6	6	128	4325	True	60	2778	1284	8,20	74999
Apple iPhone 14 (512GB)	apple	3,22	6	6	512	3279	True	60	2532	1170	8,20	95999
Apple iPhone 14 (256GB)	apple	3,22	6	6	256	3279	True	60	2532	1170	8,20	75999
Apple iPhone 13 Pro Max (256GB)	apple	3,22	6	6	256	4352	True	120	2778	1284	8,40	139900
Apple iPhone 13 Pro Max (1TB)	apple	3,22	6	6	1024	4352	True	120	2778	1284	8,60	179900
Apple iPhone 13 Pro Max	apple	3,22	6	6	128	4352	True	120	2778	1284	8,40	129900
Apple iPhone 13 Pro (256GB)	apple	3,22	6	6	256	3095	True	120	2532	1170	8,30	129900
Apple iPhone 13 Pro (1TB)	apple	3,22	6	6	1024	3095	True	120	2532	1170	8,40	147900
Apple iPhone 13 Pro	apple	3,22	6	6	128	3095	True	120	2532	1170	8,30	119900

Bottom 10 Android & iOS mobile with metrics

Model	Brand	Processor speed [GHz]	Cores	RAM [GB]	Internal memory [GB]	Battery [mAh]	5G	Refresh rate [Hz]	Resolution height	Resolution width	Average rating	Price
Apple iPhone 11	apple	2,65	6	4	64	3110	False	60	1792	828	7,30	38999
Apple iPhone 11 (128GB)	apple	2,65	6	4	128	3110	False	60	1792	828	7,50	46999
Apple iPhone 12	apple	3,10	6	4	64	0	True	60	2532	1170	7,40	51999
Apple iPhone 12 (128GB)	apple	3,10	6	4	128	0	True	60	2532	1170	7,50	55999
Apple iPhone 12 Mini	apple	3,10	6	4	64	0	True	60	2340	1080	7,40	40999
Apple iPhone 12 Mini (128GB)	apple	3,10	6	4	128	0	True	60	2340	1080	7,50	45999
Apple iPhone 12 Mini (256GB)	apple	3,10	6	4	256	0	True	60	2340	1080	7,50	55999
Apple iPhone 14 Mini	apple	0,00	0	6	128	3500	False	60	2340	1080	7,00	69990
Apple iPhone 14 Pro	apple	0,00	6	6	128	3200	True	120	2556	1179	7,50	119990
Apple iPhone 15	apple	0,00	0	6	128	3285	False	60	2532	1170	7,20	82990
Apple iPhone 15 Plus	apple	0,00	6	8	128	4532	True	120	2778	1284	7,50	84990
Apple iPhone 15 Pro	apple	0,00	0	8	128	0	True	120	2532	1170	7,50	130990

A második report az értékelések alapján listázza a legjobb tíz és legrosszabb tíz mobiltelefont. Dinamizumusa, hogy az adatokat szintén az egyes metrikák, tulajdonságok alapján lehet rendezni. Leszúrás segítségével pedig szabályozható, hogy éppen melyik operációs rendszerrel rendelkező telefonok adatai kerüljenek megjelenítésre.

3. report

Android mobiles with scores

Device	cpu	gpu	mem	ux	total_score
Red Magic 7 (S-8 Gen 1 18/256)	235998	445681	175442	181650	1038771
Redmi K50 Pro (M-9000 8/128)	244999	390026	169449	182367	986840
Mi 12 Pro (S-8 Gen 1 12/256)	221570	427999	164052	168607	982228
Motorola Edge 30 Pro (S-8 Gen 1 12/256)	217615	430290	159805	169686	977395
realme GT2 Pro (S-8 Gen 1 12/256)	218433	418406	162502	167575	966916
iQOO 9 Pro (S-8 Gen 1 8/256)	216549	410066	156134	171587	954336
Mi 12 (S-8 Gen 1 8/256)	215610	409770	162098	160914	948391
Galaxy S22 Ultra 5G (S-8 Gen 1 12/512)	219733	397965	167766	157386	942849
Galaxy S22 Ultra 5G (E-2200 12/512)	221164	367557	173807	151948	914476
Galaxy S22+ 5G (S-8 Gen 1 8/256)	217684	378119	157208	152510	905520

A harmadik report megjeleníti az összesített pontszám alapján a legjobb tíz android operációs rendszerrel rendelkező mobiltelefonokat. Mivel ezen adatok a Score táblából jönnek, mely nem tartalmaz jelenleg külön operációs rendszert meghatározó mezőt (oszlopot), így a *Device* oszlop szűrési feltételében megadtam, hogy csak azon telefonok kerüljenek listázásra, amelyek nem tartalmaznak iPad vagy iPhone szórészletet teljes nevükben.

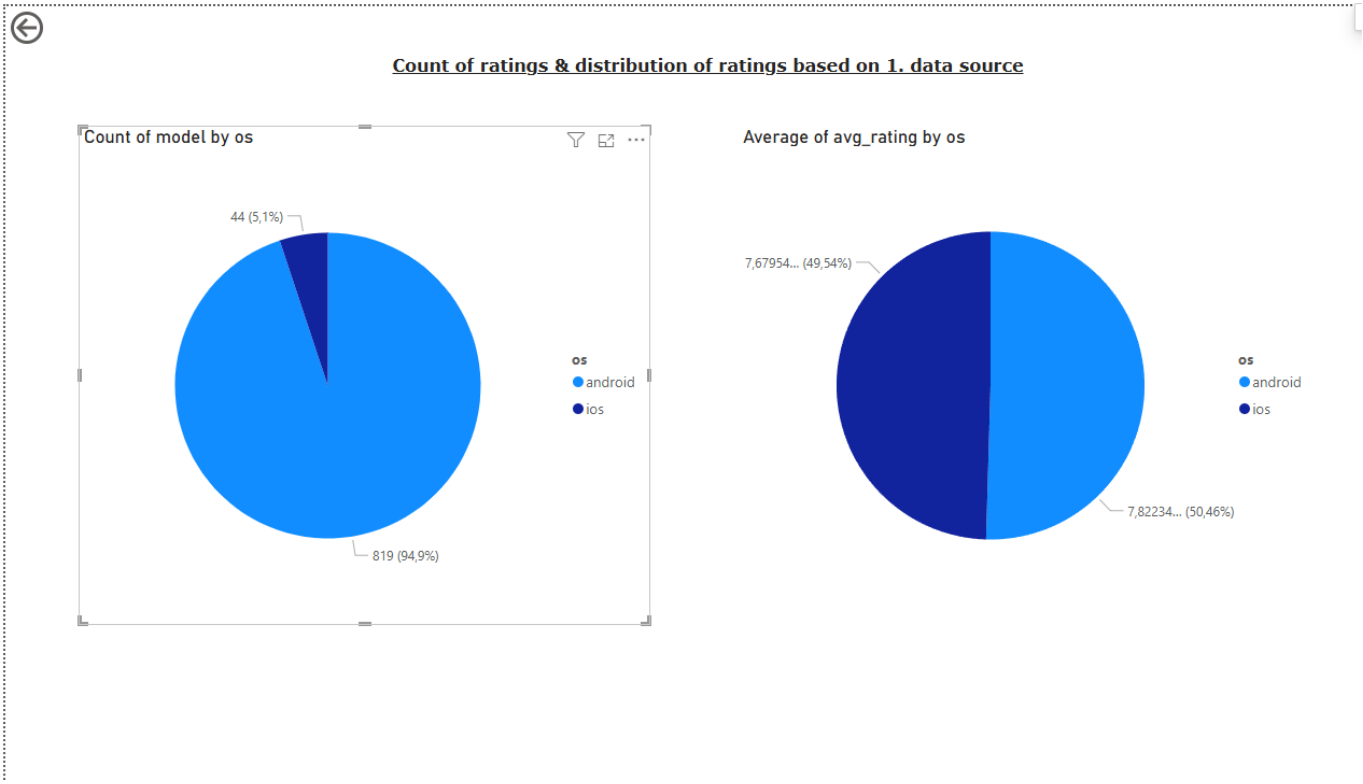
4. report

iOS mobiles with scores

Device	cpu	gpu	mem	ux	total_score
iPad Pro 5 (12.9-inch) (8+256)	351192	741383	183160	153993	1429728
iPad Pro 5 (11-inch) (8+128)	351212	722399	137145	149463	1360219
iPad Air 5 (8+64)	345275	635015	107955	140140	1228385
iPad Pro 4 (12.9-inch) (6+256)	229713	490000	125757	124585	970055
iPad Pro 3 (12.9-inch) (4+256)	229116	480648	118624	123241	951629
iPhone 13 Pro Max (6+1024)	280147	371150	163707	135420	950424
iPhone 13 Pro (6+1024)	276878	366104	163444	133505	939931
iPad Pro 3 (11-inch) (4+256)	229338	465562	118147	120285	933332
iPad Pro 4 (11-inch) (6+128)	229635	484190	97571	121386	932782
iPad mini 6 (4+64)	271405	411493	97021	132231	912150

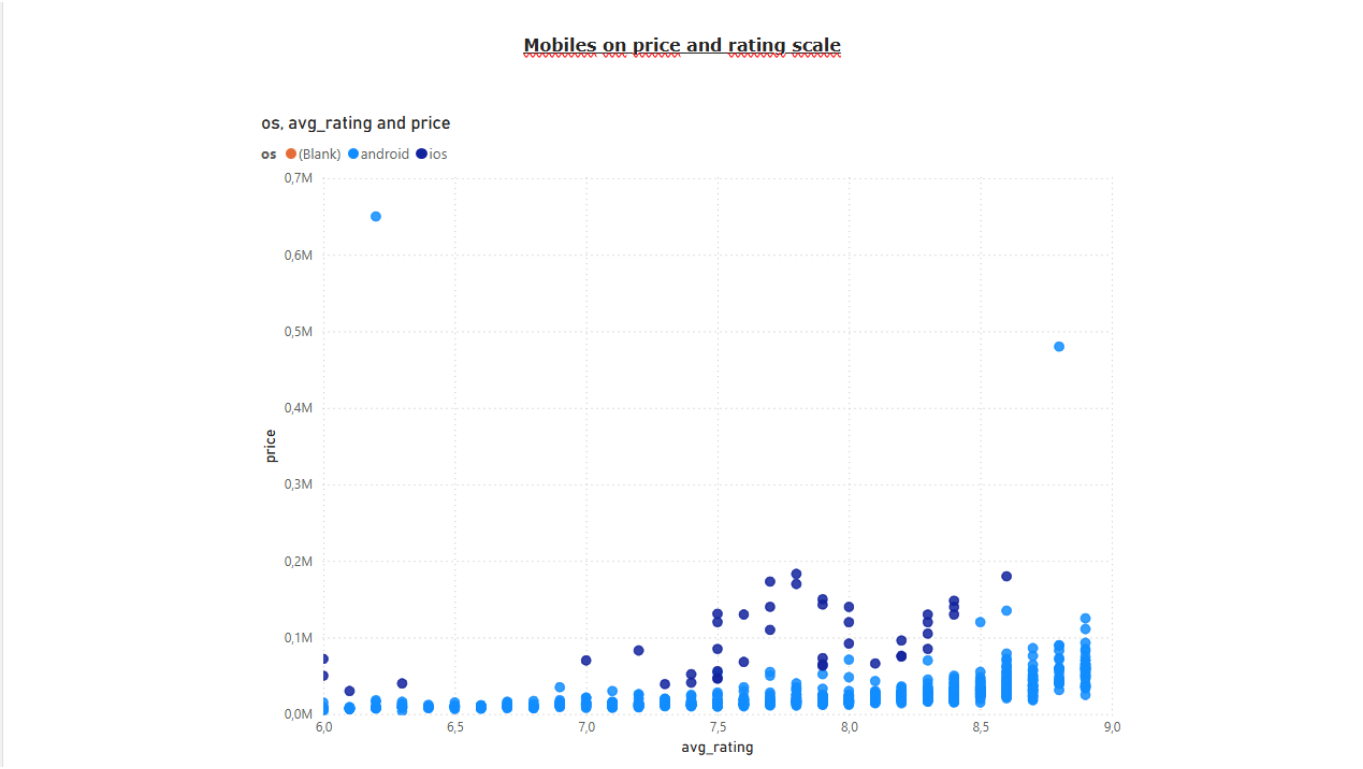
A negyedik report az előzőhez hasonlóan jeleníti meg a legjobb tíz iOS mobiltelefont.

5. report



Az ötödik report bemutatja az első adatforrás alapján a leadott értékelések és operációs rendszerek közötti eloszlást. Az látszik, hogy míg Android operációs rendszerre jóval több szavazatot adtak le, addig a telefonok értékelése során kizárólag egy picivel tud felül kerekedni az iOS típusú telefonokon. Az Android telefonok egy átlagos 7,8-as, míg az iOS telefonok egy átlagos 7,7-es értékelést könyvelhetnek el maguknak (kerekítés után).

6. report



Végül a hatodik report megjeleníti a két operációs rendszerrel rendelkező eszközök eloszlását egy ár és érték koordináta rendszerben az első adatforrás alapján. Megfigyelhető, hogy 1-2 kivétellel, de az Android

telefonok általánosságban olcsóbbak, mint az iOS eszközök legalább ugyanolyan jó, sőt jobb értékelést is el tudnak érni.

Adatelemzés

Az adatelemzés a specifikációtól eltér, hiszen az első adatforrás helyett egy olyan harmadik adatforrást használ, amely tartalmazza az egyes eszközök megjelenésének **pontos dátumát**. Továbbá megpróbálja az adatok alapján előre jelezni a következő kiadott telefonok metrikáit. Mivel az adatforrás kevés iOS eszközt tartalmaz korrekt adatokkal, így külön iOS eszközök predikciót nem tudtam elkészíteni.

- Harmadik adatforrás

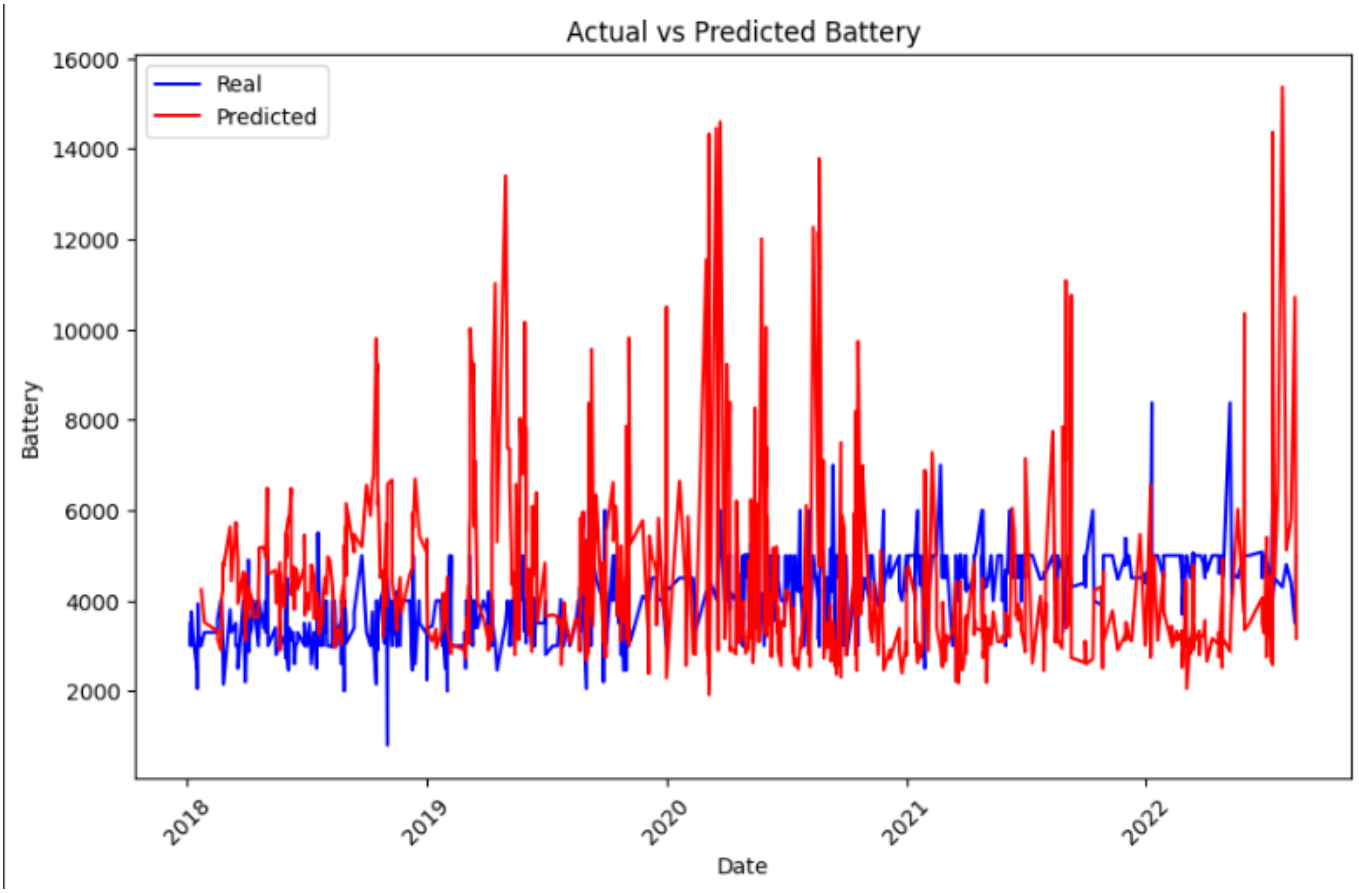
Az adatelemzéshez tartozó kód ezen a linken található meg.

Általánosságban a következő beállítások alapján készültek az adatelemzések:

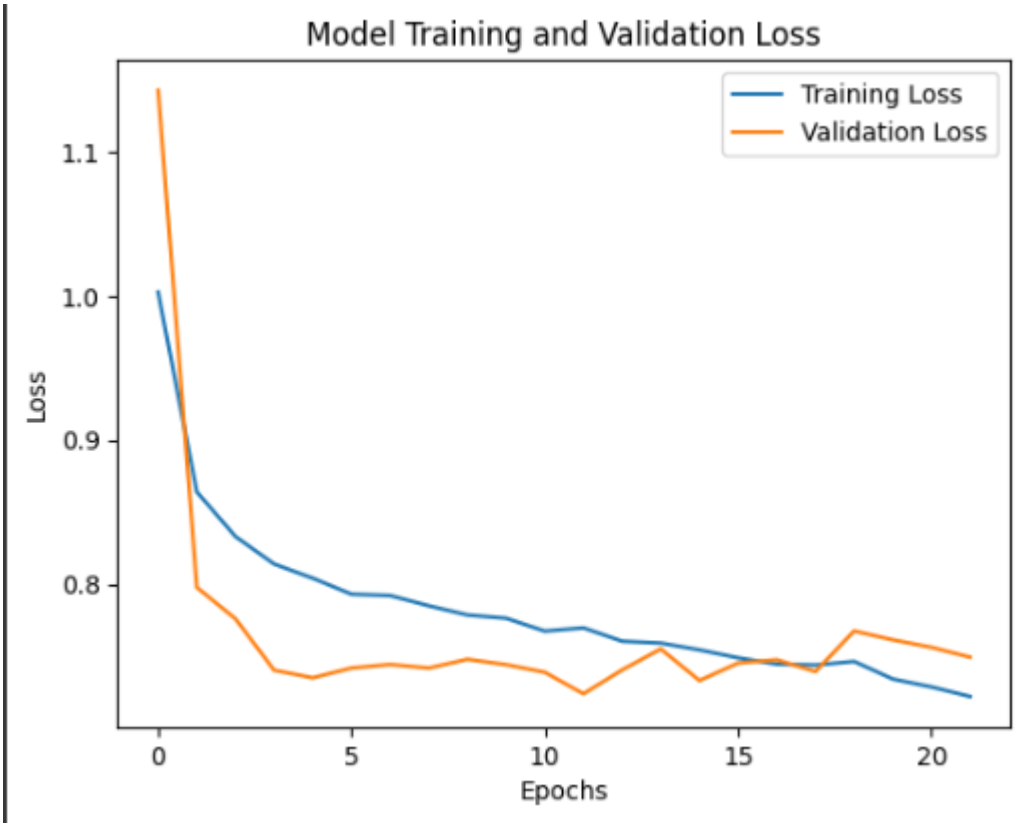
- 1 éves távlat
  - training set: 2018
  - validation set: 2021
  - test set: 2022
- 4 éves távlat
  - training set: 2011
  - validation set: 2017
  - test set: 2018

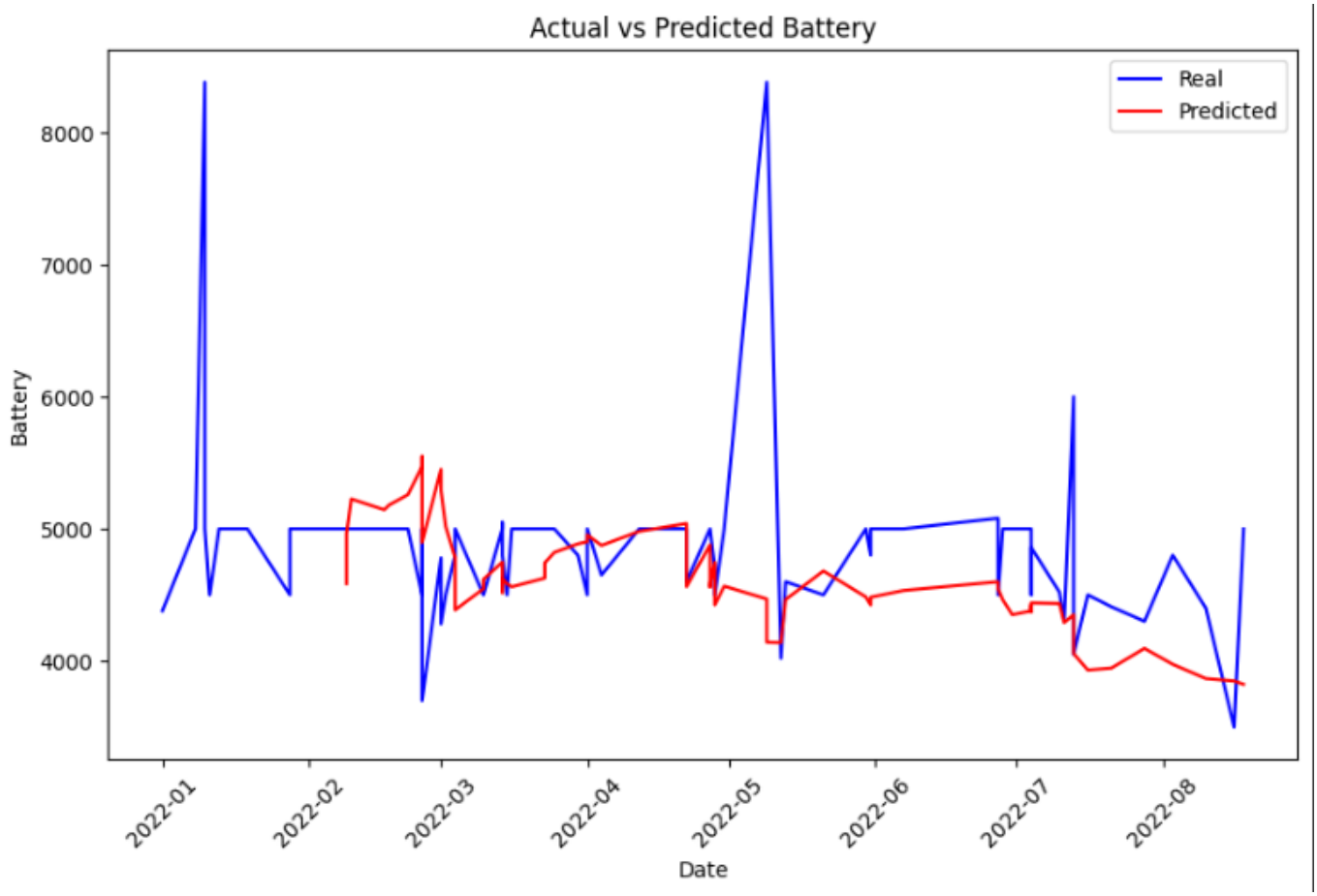
Akkumulátor

Készítettem kizárólag az android eszközök akkumulátor élettartamáról egy idősoros előrejelzést. Itt sajnos bizonyos kiugró értékek eléggé megnehezítették egy konzisztensen növekvő fejlődés képét. A valós adatokon jól látszódik, hogy egy négy éves távlatban folyamatosan fejlődik valamennyit átlagosan, ugyanakkor nem annyira, mint azt a predikció mutatja.



Ugyanakkor, ha a megjósolandó évek számát csökkentjük, és sokkal közelibb adatokkal dolgozunk, akkor valamennyivel jobb predikciót kaphatunk. Ekkor a veszteségek is sokkal kevesebbek.





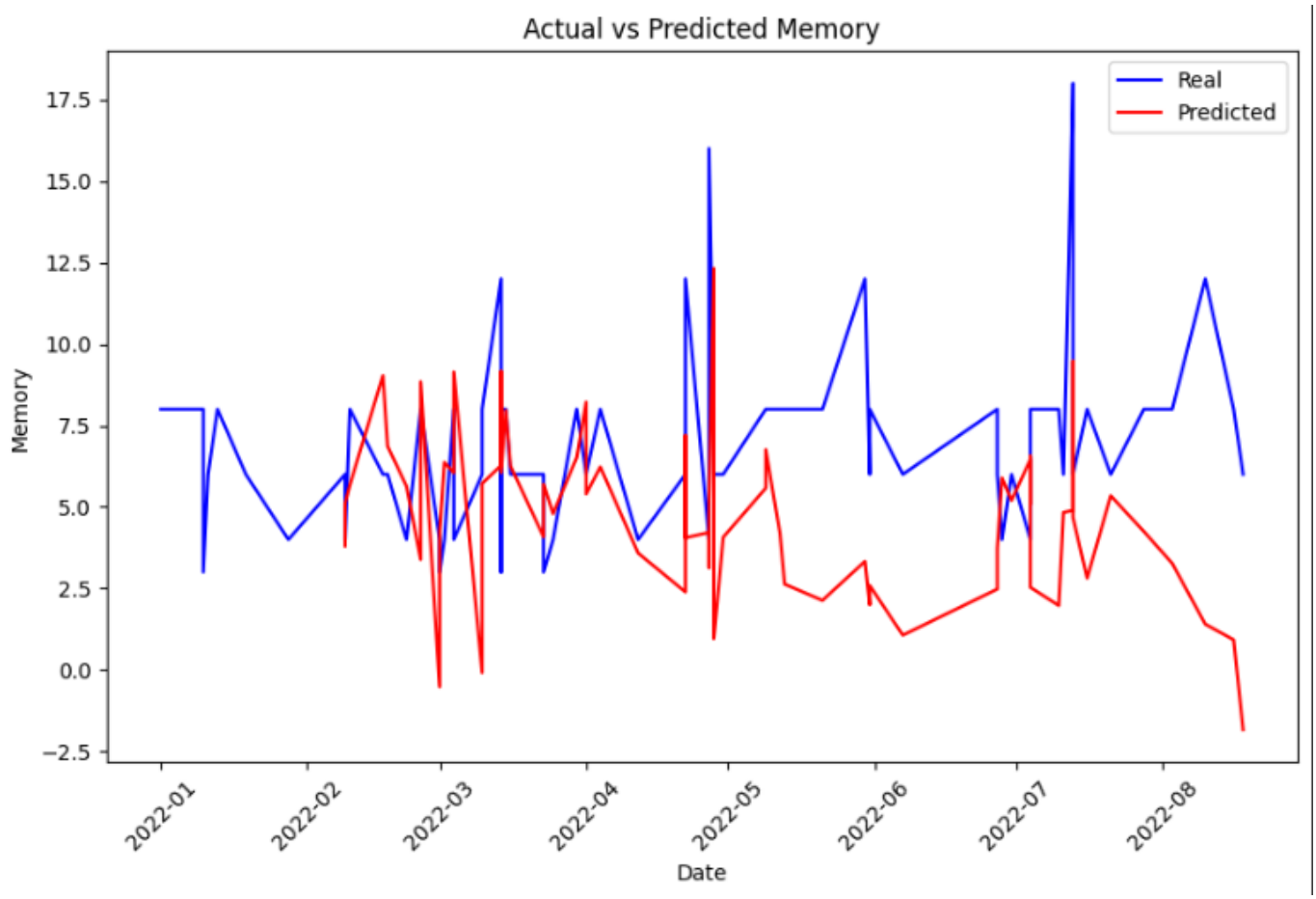
Itt a jól látható kiugró értékeken kívül általánosságban elmondható, hogy a predikció jó, sőt a 2022-05-ös időszakban szinte betrafálja.

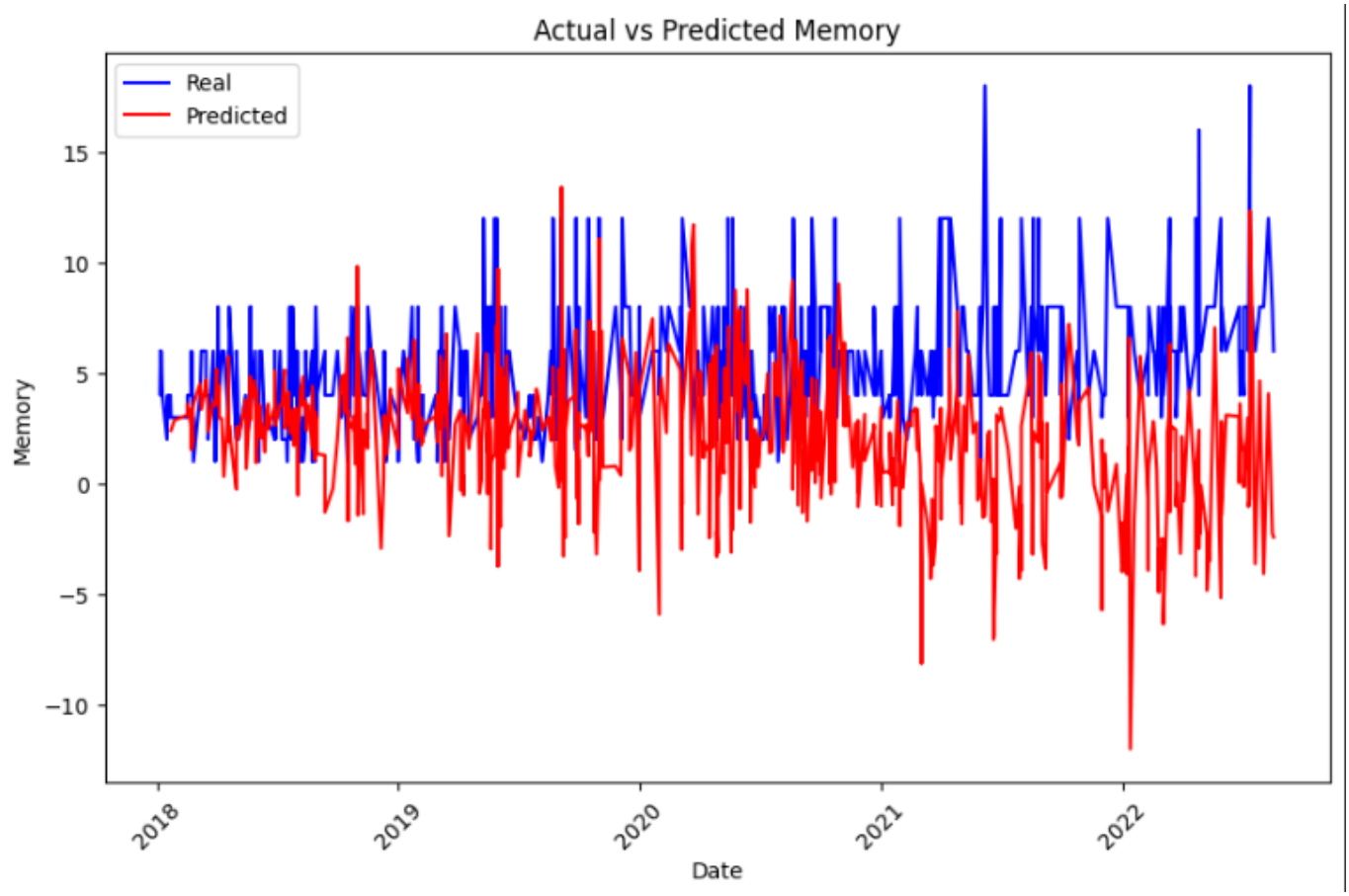
RAM

Nézzük, hogyan alakul android, majd bármely operációs rendszerrel rendelkező telefon esetén a RAM predikciója.

Android

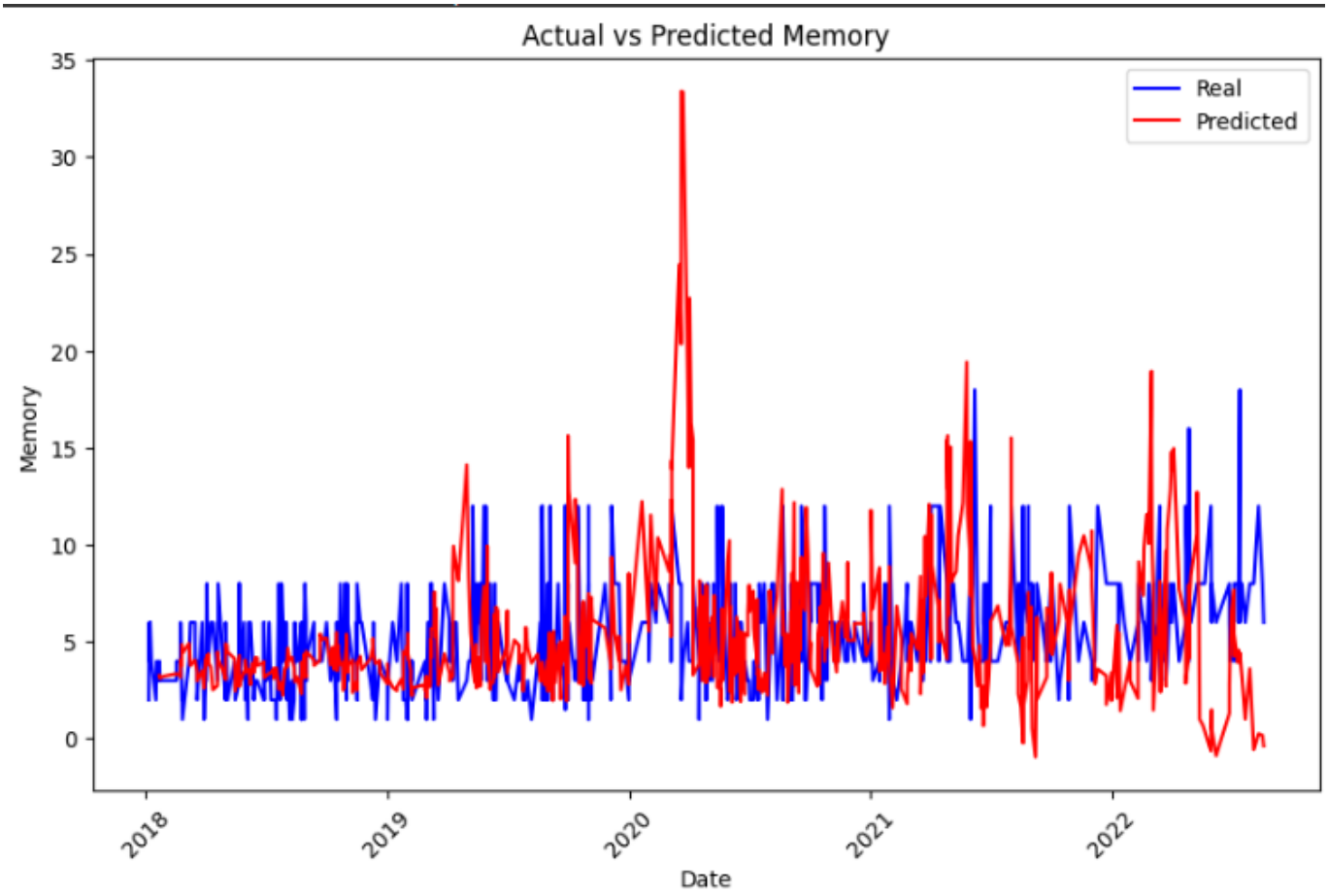
Az első képen egy év predikcióját, míg a másodikon négy év predikcióját vizsgáltam. Az első képen a kiugró értékek szintén bezavarhatnak, ugyanakkor az év elején és közepén sikerül pontosan becsülnie. A második képen viszont 2019-től a valódi mennyisége egy RAM-nak egy telefon legalább kétszeresére nő, azonban ezt a predikció 2021-től kevesebbnek jósolja, ami hibás predikcióhoz vezet. Itt a training set 2011-től, a validation set 2017-től indult.





All OS

Érdekes módon, ha a training és validation set az előzőekhez hasonlóan kerül beállításra és négy év távlatát vizsgáljuk, akkor sokkal szebb és pontosabb predikciót kapunk, egy nagyon nagy kiugró becslést leszámítva.

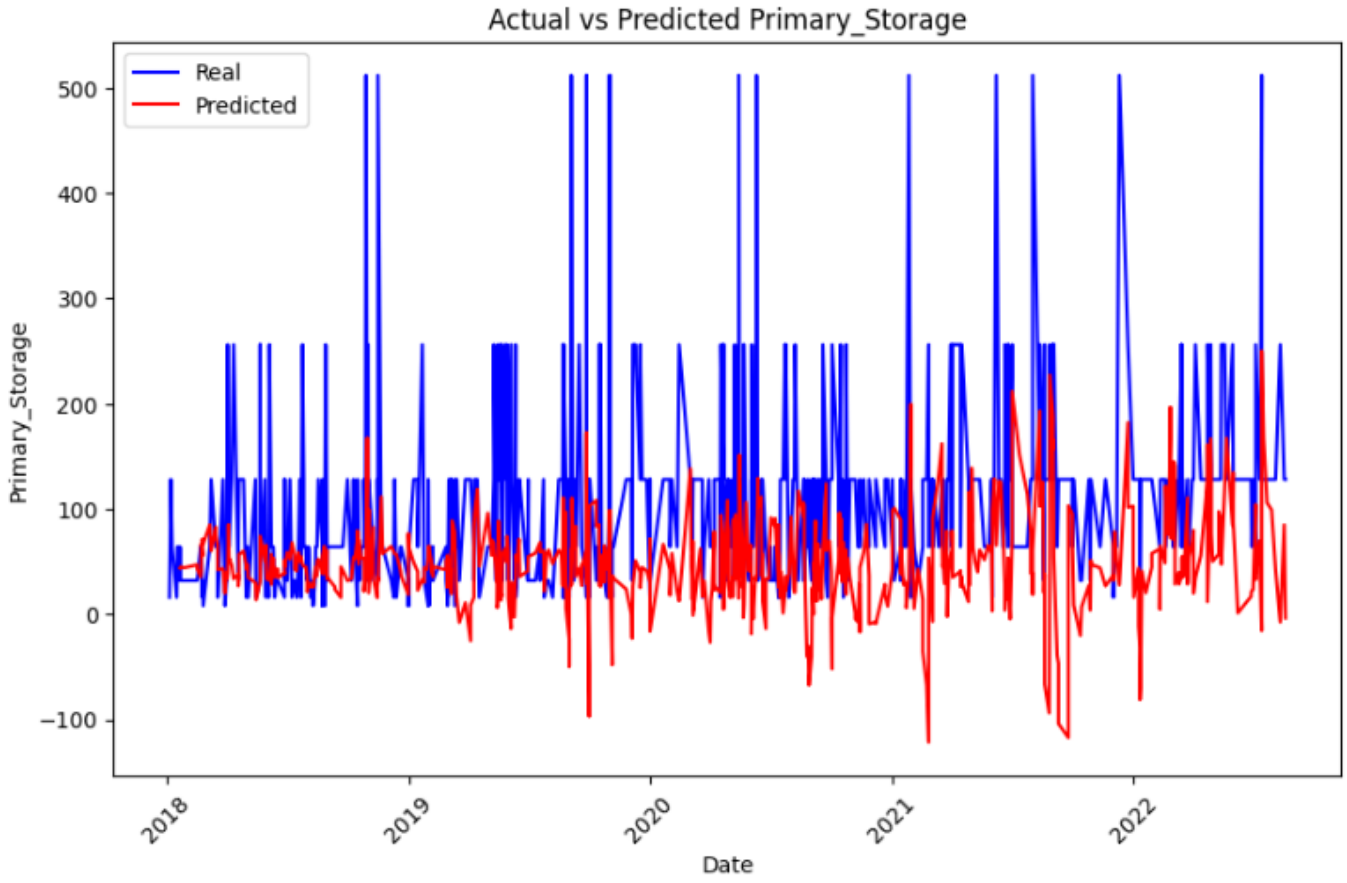


Belső tárhely

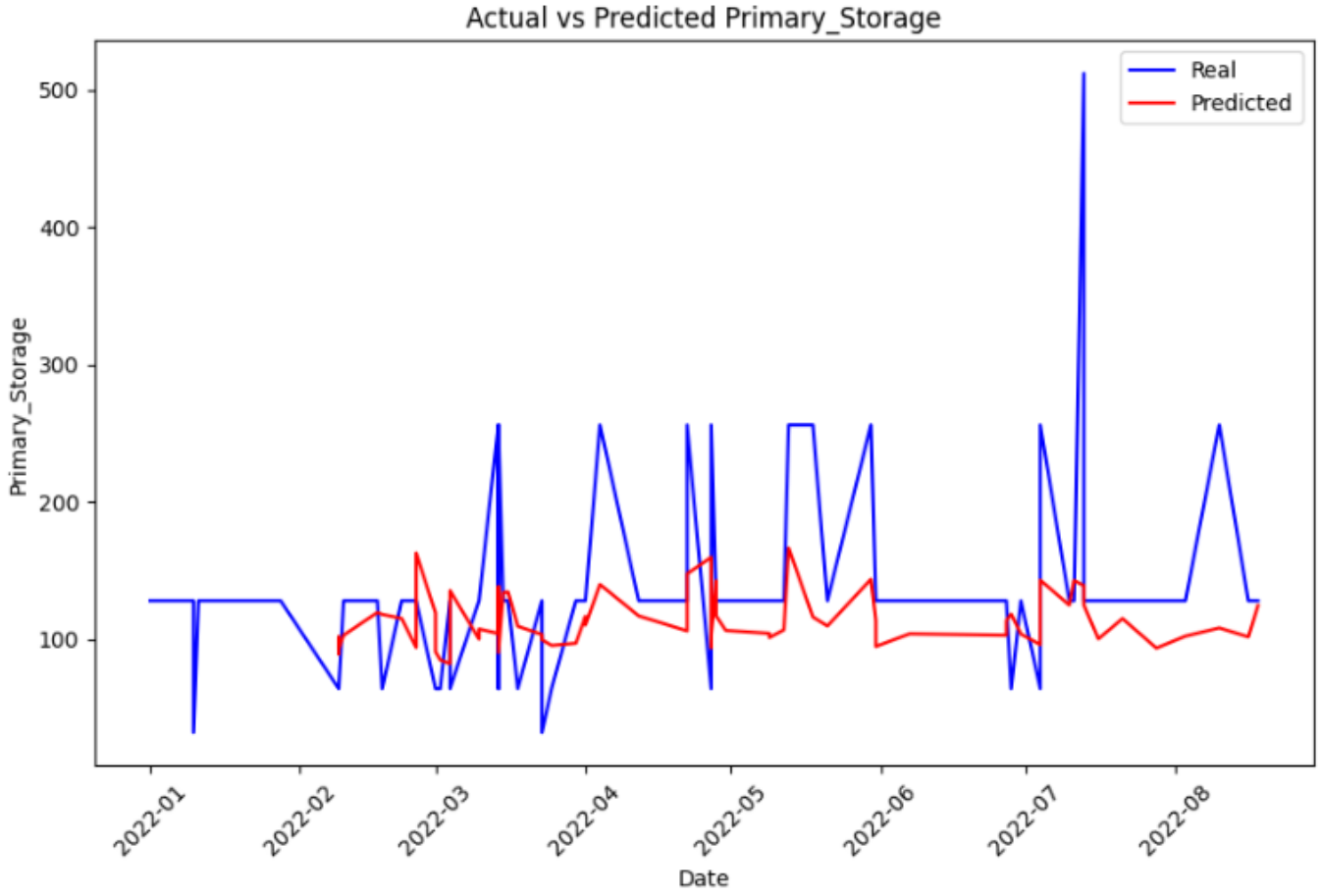
Nézzük, hogyan alakul bármely operációs rendszerrel rendelkező telefon, majd android esetén a belső tárhely predikciója.

All OS

Ha minden beállítást megismétlünk a RAM-hoz hasonlóan, akkor éppen ellenkezőleg egy rosszabb becslési képet kapunk. A belső tárhely ugyanis sokkal gyorsabb fejlődésen ment keresztül, mint az akkumulátor vagy a RAM.



Vagyis, ha kisebb távlatban vizsgáljuk a fejlődést, akkor sokkal szebb, de még mindig nem a legpontosabb képet fogjuk sajnos kapni.



Android

Csak android típusú telefonok esetén így szintén egy éves távlatra jósolva kapunk valamennyivel szebb képet, mintha 4 éves távlatra vizsgálnák több régebbi adat alapján.