Tartalom jegyzék terv:

Magyarországi elektromos tőzsde leírása 5 oldal

Gyűjtött adatok jellemzése, egymáshoz való viszonyuk, adat tisztítás 5 oldal

Tradicionális statisztikai módszerek leírása 5 oldal

1-2 kipróbálása, milyen eredményeik vannak az adatainkon 8 oldal

Neuron hálók statisztikai felhasználása cikkek, eredmények 2 oldal

Neuron hálók leírása 5 oldal

Saját neuron háló fel paraméterezése a feladathoz 15 oldal

Summa 45 oldal

[Feladat 3](#_Toc495439892)

[Magyar áram tőzsde 4](#_Toc495439893)

[Tradicionális statisztikai módszerek 9](#_Toc495439894)

[Regresszió 9](#_Toc495439895)

[Lineáris regresszió 9](#_Toc495439896)

[Kétváltozós eset 9](#_Toc495439897)

[Becslési módszerek 9](#_Toc495439898)

[Legkisebb négyzetek módszere 9](#_Toc495439899)

[Elaszticitás 10](#_Toc495439900)

[Mozgó átlagolás 10](#_Toc495439901)

[Exponenciális simítás, 11](#_Toc495439902)

[Vagy hivatkoznak még rá az irodalomban, mint Exponenciális mozgóátlag angolul Moving Average-Exponential Formula. 11](#_Toc495439903)

[Egyszeres exponenciális simítás 11](#_Toc495439904)

[Brown féle kettős exponenciális simítás 12](#_Toc495439905)

[Legközelebbi szomszéd 12](#_Toc495439906)

[ARIMA modell 13](#_Toc495439907)

[Modellek összehasonlítása: 13](#_Toc495439908)

[Neurális hálók 15](#_Toc495439909)

[Irodalom jegyzék: 16](#_Toc495439910)

Feladat

Idősor-előrejelzésre alkalmas módszerek áttekintése, implementálása, összehasonlítása, használhatóságuk feltérképezése. A módszerek igen széles spektrumát használták, javasolták már erre a célra (, legközelebbi szomszéd alapú modellek, ARMA, ARIMA modellek, mesterséges neurális hálók, regressziós modellek stb.), ezek áttekintése és egy csoport kiválasztása, majd megvalósítása és tesztelése a feladat. Részben szimulált, részben valós, mért adatokon célszerű tesztelni a megvalósított módszereket.

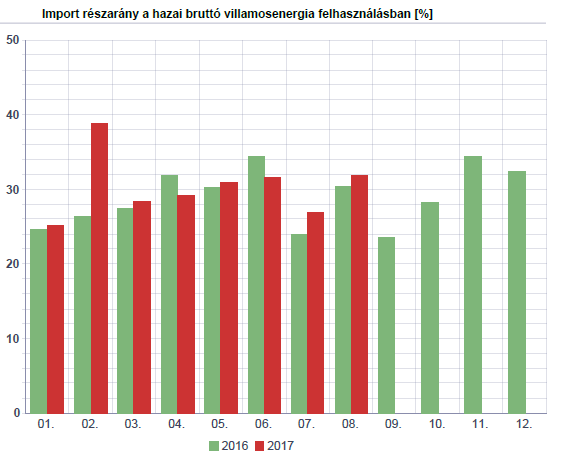
A magyar áramtőzsde (HUPX) 2009-es indulása óta már több évnyi klíringeredmény áll rendelkezésre mélyrehatóbb elemző vizsgálatok elvégzéséhez. Az elmúlt 6 évben ráadásul három féle környezetben is működött a magyar tőzsde: önállóan, a cseh-szlovák-magyar összekapcsolásban, valamint a 4M piac-összekapcsolásban.

## Magyar áram tőzsde

A pénzügyi idősorok előrejelzését nagyban nehezíti, hogy ezek általában zajosak, nem stacionáriusak, nemlineárisak és kaotikusak, továbbá gyakran fordul elő bennük strukturális törés is Ezen okok miatt a pénzügyi/tőzsdei idősorok előrejelzése az egyik legnagyobb kihívás a piaci szereplők számára

A mavir honlapról beszerzett dokumentumok és diagramok tanulmányozása után kiderült, hogy a hazai termelés nem elegendő a hazai kereslet kielégítésre. Az igényt átlagban 30% import révén tudják biztosítani. Az alábbi diagramomon (1. ábra) jól látszik az import aránya a hazai villamos energia felhasználásba.

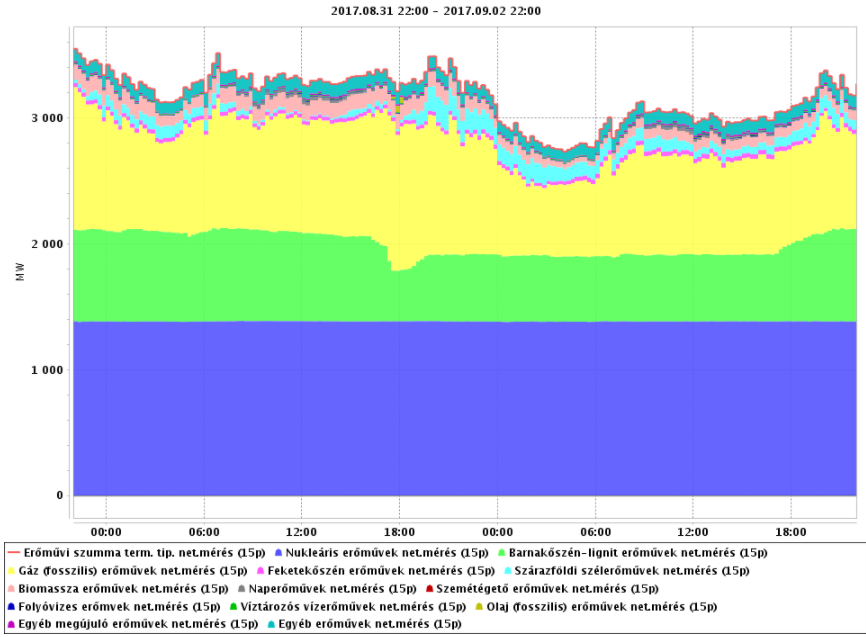
1. ábra



Érdekes hogy július és szeptember hónapban hírtelen csökken az import mennyisége. Illetve 2017 februárja nagyon kiugrott.

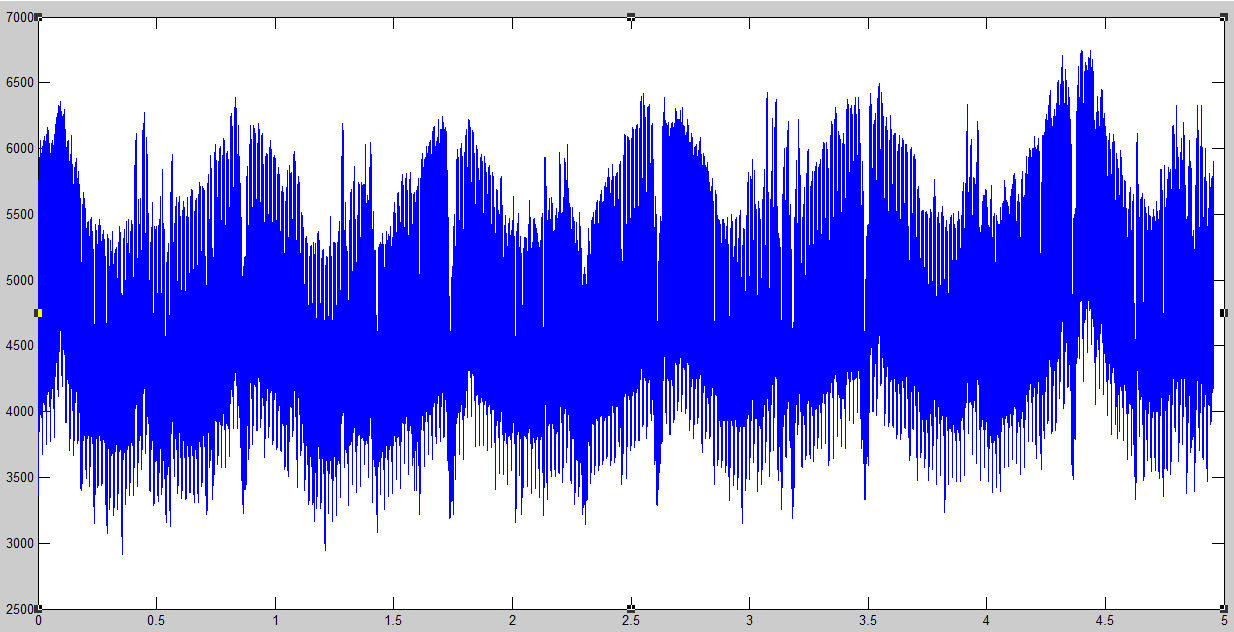
A maradék 70%-ot belső termelés adja. Több fajta erőmű szolgáltatja az áramot Magyarországon, az alábbi diagramon (2. ábra) egy két napos időszak látható. Itt jól látszik, hogy a paksi atomerőmű áram termelése a legmeghatározóbb, ezek után a kőszén, gáz erőművek adják a legtöbb energiát és egy elenyésző rész a zöld energia, a víz, szél és egyéb megújuló energia erőművek.

2. ábra



2012 január 1 és 2017 szeptember 1 közötti adatokat gyűjtöttem be órás bontásban, ami 49559 mérést jelent (valahol elhagytam egy órát?). Meglepően jól látszik (3. ábra) az éves ciklus. Január, február, december több áramfogyasztás, nyáron leesik, bár az évközepén is vannak, kilengések gondolom ezek a hőség rekordos napok, illetve a decemberi hosszú ünnep karácsony és a szilveszter idején is jelentős változás figyelhető meg, de ebben az esetben visszaesés formájában, nem dolgozik senki. A 3. ábra a hitelesített tény rendszer terhelés mutatja, ennyi áramra volt igény az adott órában. Illetve az a trend is jól látszik, hogy évről évre több áramot fogyaszt a magyar nép.

3. ábra



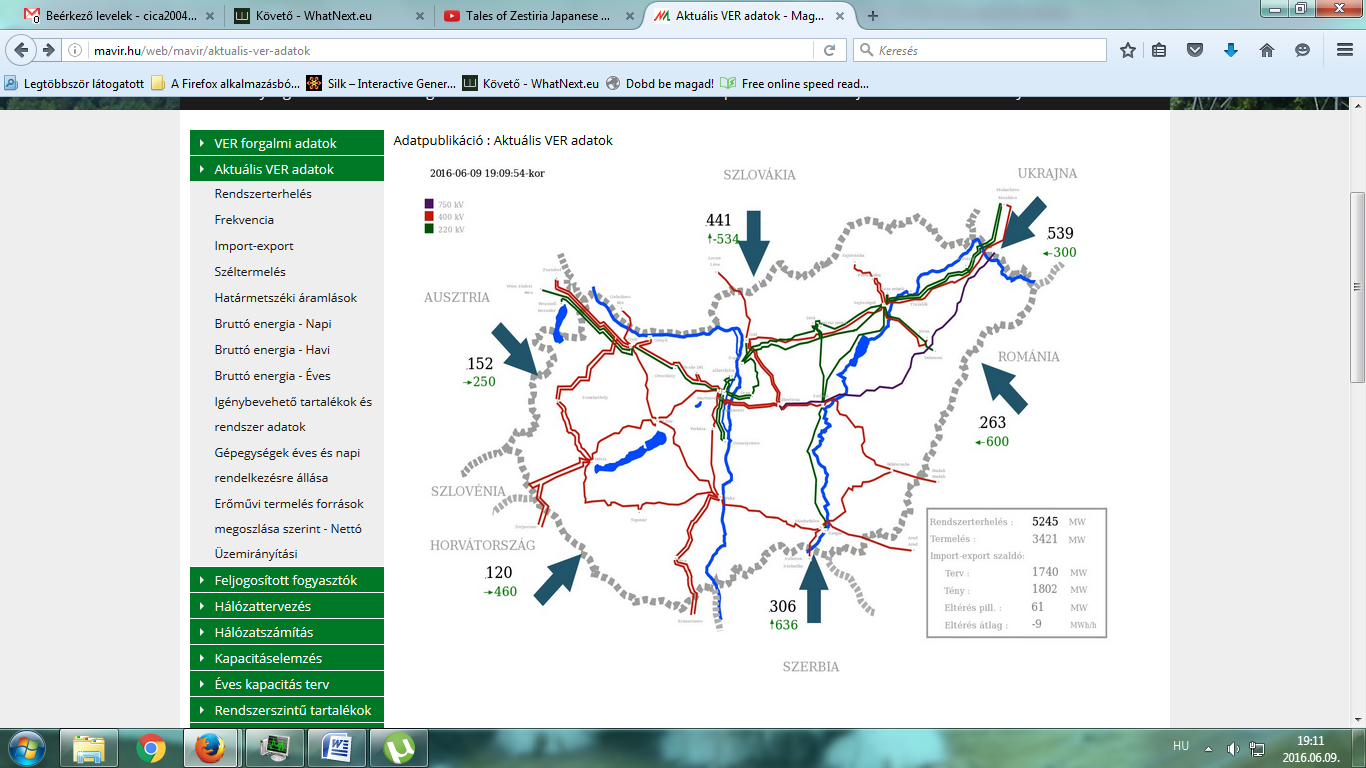
Sajnos nem teljesen hiánytalanok az adatok. Általában mindegyikből van egy terv és egy tényleges adat. A tényleges adatok sokszor üresek ilyenkor a tervezet adatokkal helyettesítem őket.

A másik nagy adat csoport amit le lehet tölteni a mavir-ról az a határmetszéki áramlások. **Határmetszék**: a magyar és a szomszédos villamosenergia-rendszer közötti egyrendszerű vagy kétrendszerű távvezetékek, amelyek szinkron üzemmódban összekapcsolják a MAVIR és a szomszédos [rendszerirányító](http://www.enker.hu/glossary/term/285) szabályozási területeit. A 4. ábra szemlélteti hogy honnan mekkora áram folyik, folyhat be. Jól látszik hogy:

* Magyarország – Ausztria HUAT
* Magyarország – Horvátország HUHR
* Magyarország – Románia HURO
* Magyarország – Szerbia HUSR
* Magyarország – Szlovákia HUSK
* Magyarország – Ukrajna HUUK

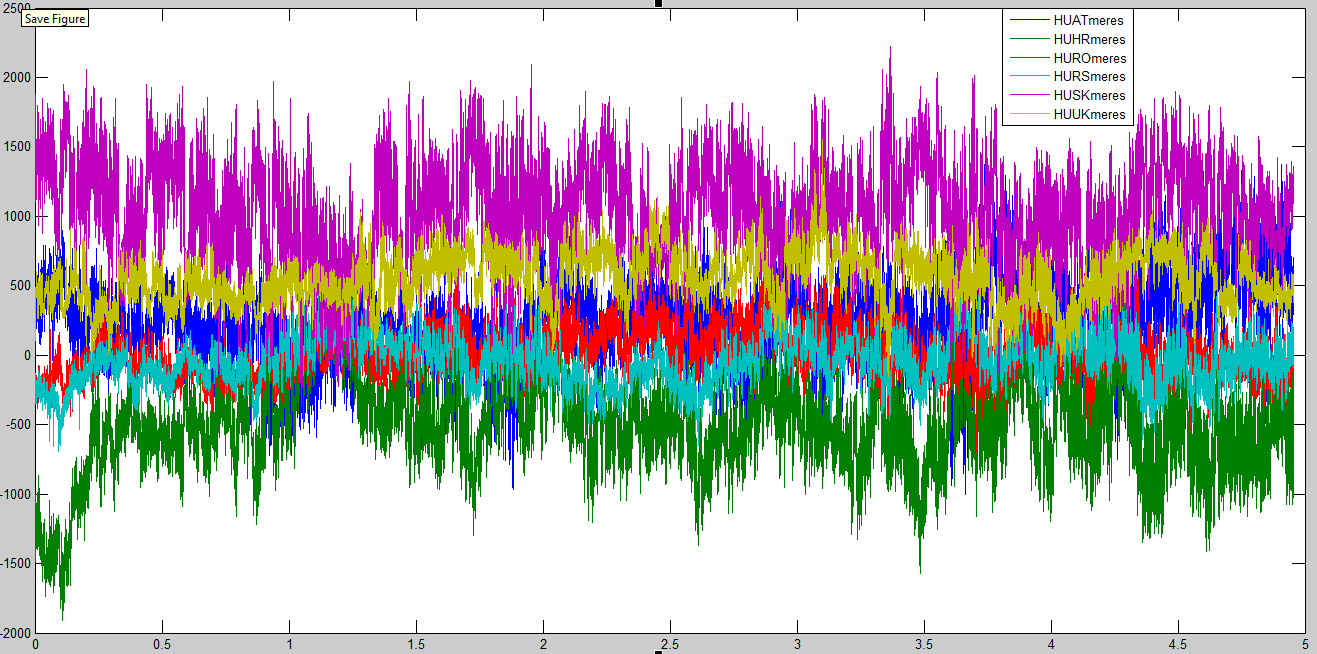
között van határmetszéki áramlás. Későbbiekben a 4 betűs rövidítéseiket fogom használni mikor rájuk hivatkozom. Jelentős határmetszék az HUUK mert egy 750kV egy 400 kV és 2 db 220 kV távvezetéken kapcsolódunk Ukrajnához, illetve jelentős még a Ausztriai hetszék és a Szlovákiai, ezeken a pontokon több 400 kV-os illetve 220 kV-os távvezetéken kapcsolódunk. A következtetésemet alátámasztja még mavir honlapról beszerzett adatok amelyek azt mutatják hogy ezekből az országokból folyik be a legtöbb áram Magyarországra.

4. ábra



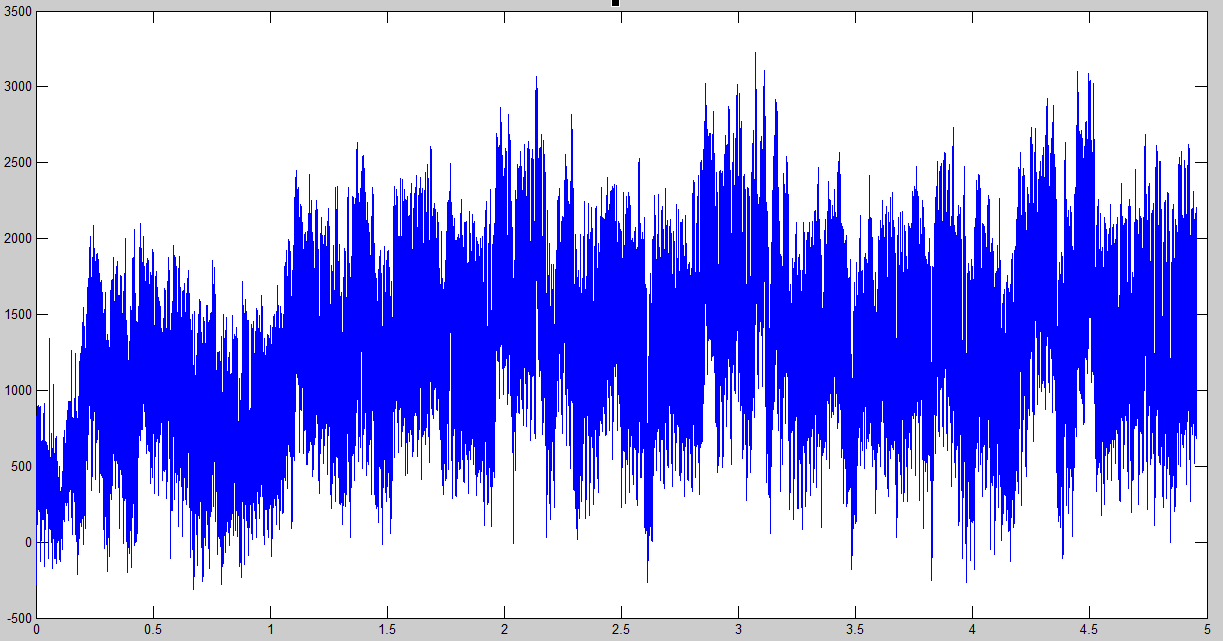
Nagyon fontos megjegyezni hogy nem Magyarország használja fel az összes import áramot, Horvátország, Szerbia rajtunk keresztül részesednek a közép Európában termelt áramból. Az 5. ábra kaotikus színes világa mutatná be az egyes metszékre jellemző exportot, importot.

5. ábra



Ezek fényében megfogtam az összes határmetszéki adatot és summáztam őket, 6. ábra lett az eredmény, aminek maximuma 3224 kW lett ami megdöbbentő annak a fényében hogy a tény rendszerterhelés átlagosan 49330 kW. Nagyon szabálytalannak tűnik a rendszer terheléshez képest az export-import.

6. ábra



## Tradicionális statisztikai módszerek

### Regresszió

#### Lineáris regresszió

A lineáris regresszió egyike a leggyakrabban alkalmazott statisztikai eljárásoknak. Ha több folytonos változó lineáris kapcsolatban van egymással, akkor az egyik csoport segítségével (magyarázó változók) előre jelezhetjük a másik csoport értékét (eredmény változók). Szükségünk van a függő és független változó kiválasztására, de ez nem jelent oksági kapcsolatot!

Az összefüggés segítheti a megértését a kapcsolatnak és legfőképp releváns előrejelzéseink lehetnek. Sajnos ez nem ilyen egyszerű.

Nagyon jó kiindulási alap, hiszen általa meg tudjuk határozni hogy a változók függetlenek vagy összefüggnek-e, ha igen menyire vannak hatással egymásra, milyen a kapcsolat erősége, predikció. Ezen okokból nagyon gyakran lehet vele több változós tözsedei előre jelzésekben találkozni.

Kétváltozós eset:

Y az eredmény változó

X a magyarázó változó

és a regressziós együtthatók, míg az a véletlen változó.

A lehető legkisebb hibájú becslés a cél. A hibáról feltételezzük, hogy független a magyarázó változótól és átlaga nulla. A becslés csak tökéletes kapcsolat esetén lenne hibamentes.

Becslési módszerek pl.:

* Legkisebb négyzetek módszere
* Momentumok módszere
* Legnagyobb valószínűségek módszere

##### Legkisebb négyzetek módszere

A **legkisebb négyzetek módszere** a mérések matematikai feldolgozásában használt eljárás. Nevét arról kapta, hogy az eltérések négyzetösszegét igyekszik minimalizálni.

A módszer érzékeny a nagyon kilógó adatokra. Egy kilógó adat az egész eljárás eredményét megváltoztathatja, hamis képet adva az adatsorról. Különböző statisztikai tesztekkel szűrik az adatsort, hogy ne maradjanak benne mérési hibák. A kilógó adatokat elhagyják, vagy a kívülállókra kevésbé érzékeny módszerekkel alternatív becsléseket végeznek. Ilyen például a súlyozott regresszió, amiben a kívülálló adatok súlyát, és ezzel befolyását is csökkentik [6]

##### Elaszticitás

A rugalmasság mérőszáma. Azt fejezi ki, hogy az x magyarázó változó 1% változása hány %-os változást okoz az eredmény változóban.

### Mozgó átlagolás

A mozgóátlagok számítása az idősorok hosszabb távú elemzésének legegyszerűbb módja. Csak annyit tűzünk ki célul, hogy átlagolással kiszűrjük a durva, egészen rövid távú ingadozásokat. Végrehajtása matematikailag igen egyszerű, jól alkalmazkodik az idősor jelleméhez. Hátránya hogy az idősor megrövidül, és jól kell, megválasztani az átlagolandó tagok számát különben torzít. Simítja az idősort, de az extrém értékek erősen befolyásolják. Mindig páratlan számú elemet tudunk átlagolni, mert az adott tagnak az átlagát úgy számítjuk, hogy vesszük az adatott tagot és előtte egy utána ugyan annyi számú további tagot. Ebből pedig már egyértelműen következik, hogy a mozgó átlagoknak van egy igen lényeges tulajdonsága, mégpedig az, hogy nem lehet minden egyes elemhez mozgóátlagot számítani, így a megfigyelt idősor eleje és vége elvész. Gyakran használják negyed éves trendek kiszámítására a devizapiacon vagy a tőzsdén. Egyszerű, nem igényel nagy gazdasági vagy matematikai tudást az alkalmazása. Készítenek egy rövid távú mozgóátlagot és egy hosszú távút, ha rövid távú felülről lefelé halad át a hosszú távún akkor várhatón csökkenő trend fog következni, így érdemes eladni, ha a rövid távú alulról metszik a hosszú távút akkor emelkedésre számítunk, így veszünk.

A mozgóátlagnak a 3 legismertebb fajtája a következő: egyszerű (vagy aritmetikai), exponenciális, és súlyozott. Módszerek közötti különbség a súlyozásban van, az egyszerű mozgóátlag minden egyes elemet egyforma súllyal vesz figyelembe, az exponenciális és a súlyozott pedig a frissebb adatokat nagyobb súllyal értékeli.

**Egyszerű mozgóátlag (Moving Average-Simple Formula)**

Az adatsor egyszerű számtani átlaga, ahol azonban -amint azt már fentebb említettünk-, nem szabad elfelejteni, hogy az egyszerű mozgóátlag késve követi a folyamatokat, így a trend megváltozását is késve jelzi. Mivel miden adatot egyforma súllyal vesz figyelembe, nem számol azzal a ténnyel, hogy a frissebb adatok jelentősége nagyobb.

A napok megválasztását illetően annyit érdemes még megjegyezni, hogy kevesebb nap megválasztása esetén a mozgóátlag gyorsabban és érzékenyebben reagál a változásokra

A tőzsdei gyakorlatban azonban a mozgóátlagokat nem úgy számítják, hogy a kiválasztott adat környezetében végzik az átlagolást, hanem a kiválasztott adat előtti adatokra, így nem veszik el annyi darab mozgóátlagunk, mint a választott periódus fele. Az átlag számítása tehát nem centrikus, hanem visszatekintő.

#### Exponenciális simítás,

Vagy hivatkoznak még rá az irodalomban, mint Exponenciális mozgóátlag angolul Moving Average-Exponential Formula.

Egyszerű modell. Nem feltételezünk az idősoros adatokban sem trend, sem szezonális hatást. Ez is egy fajta mozgó átlagolás azzal a különbséggel, hogy egy adott pont exponenciális simításának értékéhez elegendő a múltban közvetlenül előtte levő értékeket ismerni. Minden korábbi pont visszafelé haladva egyre kisebb súllyal számít. A súly értéke 0 és 1 között lehet. Az 1hez közeli értékek nagy súlyt kapnak az aktuális pont kiszámításában, kevésbé simítanak, míg 0 közeli súlyok erős simítást végeznek.

##### Egyszeres exponenciális simítás

Az exponenciális simítás módszer alapváltozata. Feltételezi, hogy a megfigyelt érték egy állandó körül ingadozik. A simítást α -val jelöljük. Az alfa azt mondja ki hogy mennyire szeretnénk simítani, azaz mennyire vesszük figyelembe az elkövetet hibát. Az alfa helyes megválasztása kulcsfontosságú.

Az egyszeres exponenciális simítás egyenlete: ahol:

0 ≤ α ≥ 1

a t időszakra vonatkozó trend szerinti érték

a t időszakra vonatkozó tényleges megfigyelés

Az egyszeres exponenciális simítás csak egy időszakra ad érdemi előrejelzést. Feltétele hogy az idősorban ne legyen tartós tendencia. Nagyobb alfa esetén jobban követi a tényleges adatokat, míg kisebb alfával hosszabb távú ingadozást lehet meghatározni. Ezért mikor egy idősor feltehetőleg lineáris trendet követ, nem célszerű ezt a módszert használni.

##### Brown féle kettős exponenciális simítás

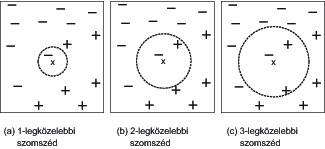
A kétszeres simítás nem más, mint az egyszeresen simított sor ismételt egyszeres simítása, így figyelembe tudjuk venni a trend értékét is. A fentebbi képletben a simított és az előre jelzett értékek megegyeztek, most viszont a kétszeres simítás miatt nem ugyan azok. S lesz a simított érték és felső indexbe teszem a simítás számosságát. Ennek megfelelően az egyszeres simítás képlete:

A kétszeres simítás egyenlete pedig:

A kettős simításnak hála alfa = 1 esetében a torzítás teljesen megszűnik, viszont ez a módszer add teljesen jó megoldást a kis alfák túl simítására. Ezzel a módszerrel tetszőleges hosszú és számú időszakra tudunk előrejelzést készíteni. Már kevés adattal is használható.

### Legközelebbi szomszéd

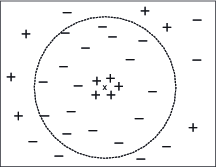
A legközelebbi szomszédok használatának indoklását legjobban a következő mondás szemlélteti: *" Ha valami úgy totyog, mint egy kacsa, úgy hápog, mint egy kacsa és úgy néz ki, mint egy kacsa, akkor az valószínűleg egy kacsa.''* A legközelebbi szomszéd osztályozó minden egyes esetet egy adatpontként reprezentál egy http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1435.png -dimenziós térben, ahol http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1436.png az attribútumok száma. Egy adott teszteset esetén a meghatározott szomszédsági mértékek valamelyikével kiszámítjuk annak közelségét a tanulóhalmaz összes többi adatpontjához. Egy adott http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1437.png eset http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1438.png -legközelebbi szomszédja azt a http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1439.png pontot jelenti, amelyek a legközelebb vannak http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1440.png -hez.



Az ábra a körök középpontjában lévő adatpont 1-, 2- és 3- legközelebbi szomszédját szemlélteti. Egy adatpontot a szomszédainak osztálycímkéje alapján osztályozunk. Abban az esetben, ha a szomszédoknak egynél több címkéje van, az adatpontot a legközelebbi szomszédok többségi osztályához rendeljük hozzá. Az ábrán az adatpont 1-legközelebbi szomszédja egy negatív eset. Ezért az adatpontot a negatív osztályhoz rendeljük hozzá. Ha három legközelebbi szomszéd van az ábrán látható módon, akkor a szomszédság két pozitív és egy negatív esetet tartalmaz. A többségi szavazási sémával az adatpontot a pozitív osztályhoz rendeljük hozzá. Holtverseny esetén az adatpont osztályozásához véletlenszerűen választhatjuk valamelyik osztályt.

A fent leírtak http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1441.png helyes megválasztásának fontosságát hangsúlyozzák. Ha http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1442.png túl kicsi, akkor a legközelebbi szomszéd osztályozó a tanulóadatokban jelenlevő zaj miatt hajlamos lehet a túlillesztésre. Másrészt viszont, ha http://www.tankonyvtar.hu/hu/tartalom/tamop425/0046_adatbanyaszat/math/eq_1443.pngtúl nagy, akkor a legközelebbi szomszéd osztályozó rosszul osztályozhatja a tesztpéldányt, mivel a legközelebbi szomszédok listája a szomszédságtól messzi adatpontokat is tartalmazhat

k -legközelebbi szomszéd osztályozás nagy k esetén



### ARIMA modell

AutoReggressieve Integrated Moving Average angol kifejezés rövidítése. Gyakran használják idősorok elemzéséhez és előrejelzéséhez.

## Modellek összehasonlítása:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Módszer | Főbb jellemzői | Előnyei | Hátrányai | Milyen időtávra |
| Lineáris reggreszió |  | Egyszerű | Pontatlan ha szthohasztikus az adat |  |
| Mozgó átlagolás | Az idősor dinamikus átlagát állítja elő | Egyszerű  Kevés adattal is működik | Matematikailag pontatlan | max negyedév |
| Exponenciális kiegyenlítés | Különböző súllyal veszi a vizsgált időszak résztendenciáit | Rövid távon megbízható | Nehéz meghatározni a súlyokat | Rövidtáv, középtáv |
| Legközelebbi szomszéd | Hasonlóság vizsgálat | Már 2 nap után tud működni | Hasonlóság mértékét nehéz definiálni, a jó működéshez sok adat kell |  |

Lineáris regresszió vizsgálata az adatainkon

## Neurális hálók

Az informatikában a 2010es évekre a GPU-k és a CPU-k exponenciális fejlődése lehetővé tette, hogy eddig már ismert, de túl elterjed matematikai eljárásokat, elveket elővegyenek és alkalmazzanak. Az egyik ilyen fejlődési irány a gépi tanulás, neuron hálók világa. Lassan nem lesz olyan ágazat, ami ne használná valamire a neuron hálót. Képet, hangot ismertetünk fel vele, jó osztályozásra, előre jelzésre. Olyan nagy cégek, mint az Nvidia vagy az Oracel ölnek időt és pénzt a neuron hálók fejlesztésébe. Egyre több programozási nyelvhez érhető el neuron hálós kiegészítés.

## Irodalom jegyzék:

1. <https://www.mateking.hu/statisztika-2/regresszioszamitas/>
2. <https://hu.wikipedia.org/wiki/Line%C3%A1ris_regresszi%C3%B3>
3. <http://psycho.unideb.hu/munkatarsak/balazs_katalin/stat1/stat1ora4.pdf>
4. <https://hu.wikipedia.org/wiki/Legkisebb_n%C3%A9gyzetek_m%C3%B3dszere>
5. <http://www.cs.bme.hu/nagyadat/bartok-ferenc.pdf>
6. <http://ilias.gdf.hu/data/ilias-ha/lm_data/lm_9370/index.html>
7. <http://gtk.uni-miskolc.hu/files/8449/Exponenciális+kisimítás.pptx>
8. <http://statisztikus.hu/fuggelek/>
9. <http://elib.kkf.hu/okt_publ/szf_11_02.pdf>
10. <https://www.mavir.hu>
11. <https://www.mavir.hu/documents/10258/107815/Sz%C3%A9lkihaszn%C3%A1lts%C3%A1g+tanulm%C3%A1ny+2010.pdf/153d2d78-1c6f-4d54-858e-5bc46f56c352>
12. <http://www.portfolio.hu/vallalatok/technikai-elemzes/technikai-elemzes-a-mozgoatlag-hasznalata.17774.html>
13. <https://www.elemzeskozpont.hu/mozgoatlagok>
14. **A predictive pan-European economic and production dispatch model for the energy transition in the electricity sector  
    Published in:** [PowerTech, 2017 IEEE Manchester](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7964294) **Date:** 18-22 June 2017 **Authors**: [Laurent Pagnier](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Laurent%20Pagnier.QT.&newsearch=true), [Philippe Jacquod](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Philippe%20Jacquod.QT.&newsearch=true)  
    <http://ieeexplore.ieee.org/document/7980982/>
15. Modeling and simulation inspired by quantum methods of the Polish Electricity Stock Exchange **Published in:** [Progress in Applied Electrical Engineering (PAEE), 2017](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8001596) **Date: 25-30 June 2017  
    Authors**: [Jerzy Tchorzewski](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Jerzy%20Tchorzewski.QT.&newsearch=true), [Dariusz Rucinski](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Dariusz%20Rucinski.QT.&newsearch=true)<http://ieeexplore.ieee.org/document/8008983/>
16. ANN approach for predicting economic trends based on electric energy consumption during natural disaster period  
    **Published in:** [Knowledge, Information and Creativity Support Systems (KICSS), 2016 11th International Conference on](http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=7945834) **Date:** 10-12 Nov. 2016 **Authors**: [Akanit Kwangkaew](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Akanit%20Kwangkaew.QT.&newsearch=true), [Virach Sornlertlamvanich](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Virach%20Sornlertlamvanich.QT.&newsearch=true), [Itsuo Kumazawa](http://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Itsuo%20Kumazawa.QT.&newsearch=true)  
    <http://ieeexplore.ieee.org/document/7951405/>
17. <http://www.mavir.hu/c/document_library/get_file?uuid=81fd9f45-12cf-44e3-ad74-9492504a42ef&groupId=10258> a határmetszék definiciója innen van