**A székelyföldi megyék munkanélküliségi rátáinak összehasonlítása, statisztikai elemzése és előrejelzése. Az ARIMA modell és a mesterséges neuronháló-alapú modellek összehasonlítása.**

# Bevezetés

A dolgozat célja, hogy Hargita, Kovászna és Maros megye elmúlt 10 év munkanélküliségi rátáit elemezzem, majd a feldogozott idősoroknak legmegfelelőbb ARIMA modell paramétereit meghatározzam és ennek segítségével a lehető legjobb előrejelzést adjam. Ezenkívül mesterséges neuronháló módszerével (ANN) is készítsek előrejelzést, végül összehasonlítom a két módszert. Ehhez egy olyan interaktív programot készítettem Python-ben, amely egy webes felüleleten (django) keresztül bekéri az adatsort és a kívánt paramétereket, majd a feldolgozást követően az eredményeket megjeleníti.

## A munkanélküliségi ráta meghatározása Romániában

A dolgozat során felhasznált adatok az INSTITUTUL NATIONAL DE STATISTICA -tól származnak (Romániai Statisztikai Hivatal). Az ő módszertanuk a következőképpen definiálja a munkanélküliséget és a munkanélküli rátát:

A **munkanélküliek** a nemzetközi meghatározás szerint (BIM[[1]](#footnote-1)) szerint azok a 15-74 év közötti személyek, akik egyidejűleg teljesítik a következő három feltételt:

1. munkanélküliek (a mérés pillanatában nincs bejelentett munkahelyük)
2. a következő két héten belül munkába tudnának állni
3. az elmúlt négy hétben aktívan munkát kerestek.

A **munkanélküliségi ráta** a munkanélküliek aránya a munkaerőben.

A gazdaságilag aktív népesség a bázisidőszakban az áruk és szolgáltatások előállítására rendelkezésre álló munkaerőt biztosító valamennyi személyt magában foglalja, beleértve a foglalkoztatottakat és a munkanélkülieket is. (INSSE, 2016)

Tehát a munkanélküliségi ráta megmutatja, hogy a munkaképes lakosság hány százaléka nem rendelkezik a mérés pillanatában munkahellyel, viszont tudna és akarna dolgozni.

## 1.2 Alapvető statisztikai mutatók fogalmai

**Átlag**: „Az egyik legismertebb statisztikai mutató. Jelöljünk n megfigyelést x1, x2, ..., xn -nel. Ekkor az átlaguk:  *”*

**Szórás**: „A szórás egy szóródási mutató, ami azt méri, hogy a megfigyelések mennyire esnek távol az átlagtól. Két minta lehet lényegesen különböző, még ha az átlaguk egyforma is, ugyanis az egyik minta megfigyelései eshetnek jóval közelebb az átlaghoz, mint a másik minta megfigyelései. A szóródás méréséhez az eltérések négyzetét veszi figyelembe, mert ezáltal fejezhető ki az összes megfigyelés távolsága az átlagtól. A szórás kiszámításához négyzetgyököt vonunk az eltérések négyzetének átlagából: ”

**A variancia (szorásnégyzet):** Egy másik fontos szóródási mutató a variancia, ami egyszerűen a szórás négyzete, vagyis az átlagtól valóeltérések négyzeteinek az átlaga:

**Medián:** Jelöljük x1, x2, ..., xn –nel a megfigyeléseket, és jelöljük x(1), x(2), ... , x(n) –nel ugyanezeket a megfigyeléseket növekvő sorrendben. Tehát, x(1) a legkisebb, x(2) a következő, ... és x(n) legnagyobb: x(1) ≤ x(2) ≤ ⋯ ≤ x(n)

A medián a sorrendbe állított x1, x2, ..., xn megfigyelések középső megfigyelése. Ha n páratlan, akkor egészen egyszerű; a medián a (n+1)/2 sorrendű megfigyelés.

Ha n páros, akkor 2 középső megfigyelés van, n/2 és (n/2)+1, tehát a medián:

*(Sándor – Tánczos ,20??)*

# Felhasznált technológiák, módszerek

## Python és a felhasznált kiegészítő csomagok

… AR, MA, ARMA, ARIMA ?

# A program működése

## Beolvasás

## A statisztikai mutatók számítása

* 1. Az idősorok ábrázolása
  2. ACF és PACF

1. Biroul Internaţional al Muncii [↑](#footnote-ref-1)