

A munkanélküliség előrejelzése Pythonban, ARIMA és MLP modellekkel

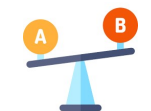
2024

Hallgató: Károlyi Krisztián, Gazdasági Informatika szak
Témavezető: Dr. Madaras Szilárd, egyetemi adjunktus, Üzleti Tudományok Tanszék

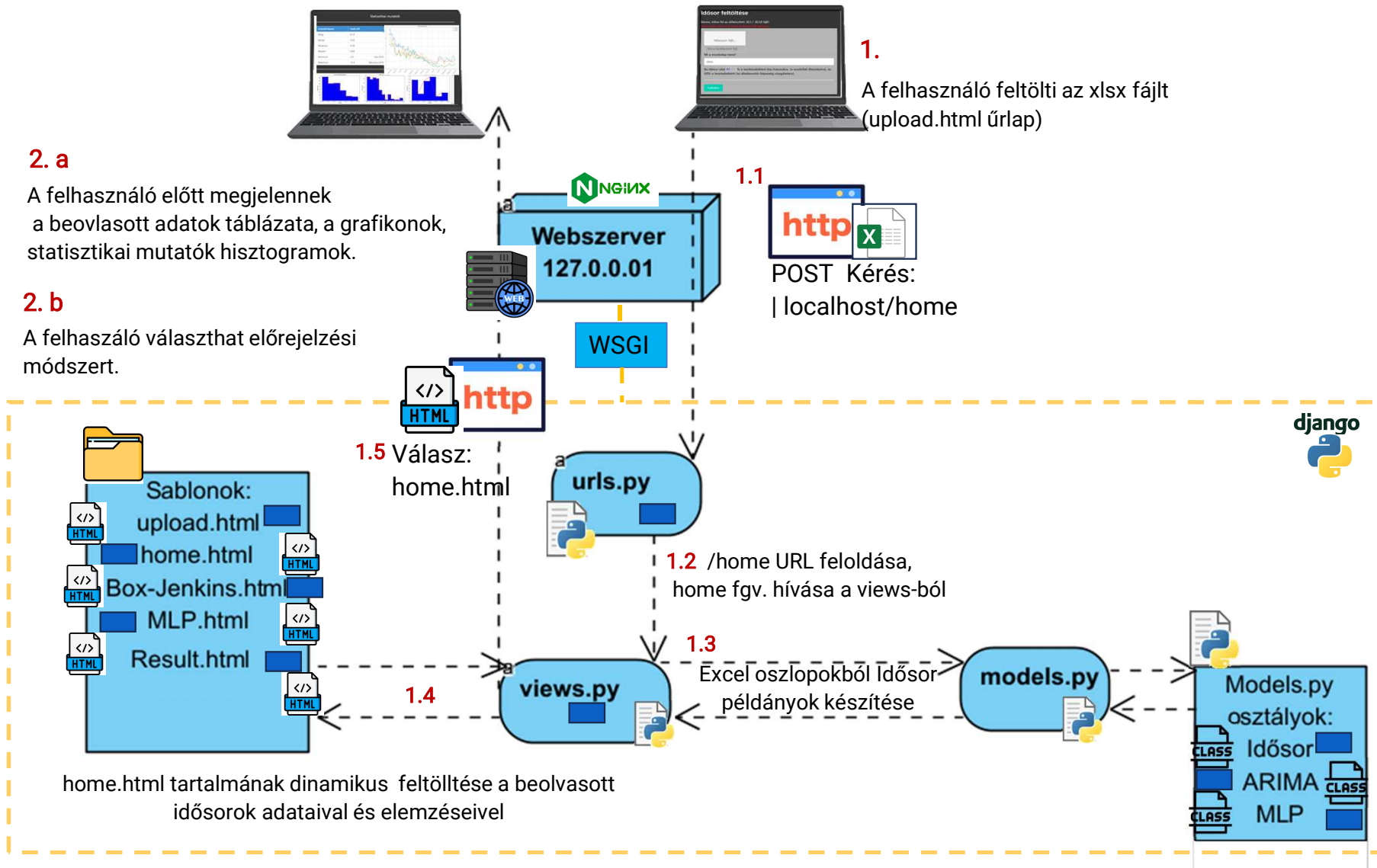
A dolgozat témája

Fő kérdés: A munkanélküliségi rátákat ARIMA vagy MLP (mesterséges neuronháló) modellekkel lehet könnyebben, pontosabban előrejelezni?

- A vizsgált idősorok: Hargita, Kovászna és Maros megye munkanélküliségi rátái (INNSE – Romániai Nemzeti Statisztikai Hivatal adatai)
- A legjobban illeszkedő ARIMA¹ modellek megkeresése (Box-Jenkins eljárás)
- A legjobban illeszkedő és általánosító MLP² regressziós modellek megkeresése
- Az ARIMA és MLP regressziós modellek becslési hibáinak összehasonlítása (MSE, RRMSE, MAPE hibamutatók)
- Előrejelzések készítése
- A modelleket Python-ban implementáltam, Django GUI-val összekötve (Excel állományból feltöltött idősorok elemzését és modellezését támogató program)

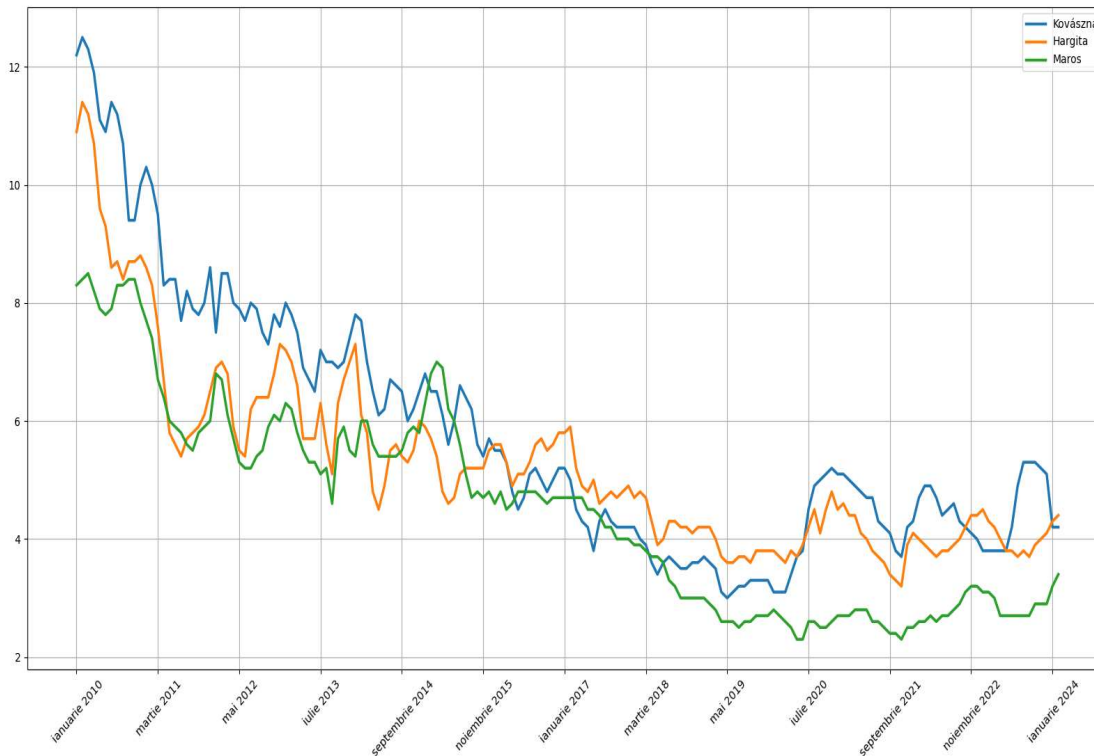


A program felépítése

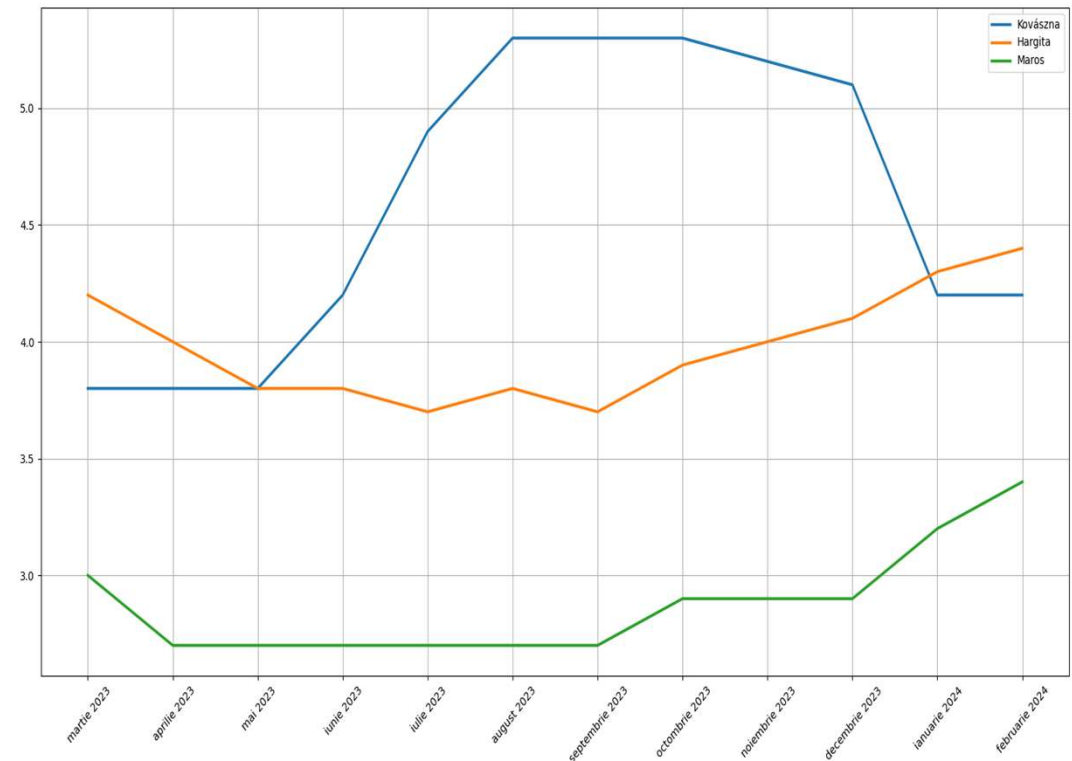


A vizsgált idősorok

Tanító adatok: 2010 január – 2023 február
A modellek illesztésére



Tesztadatok: 2023 március – 2024 február
általánosító képesség kiértékelésére

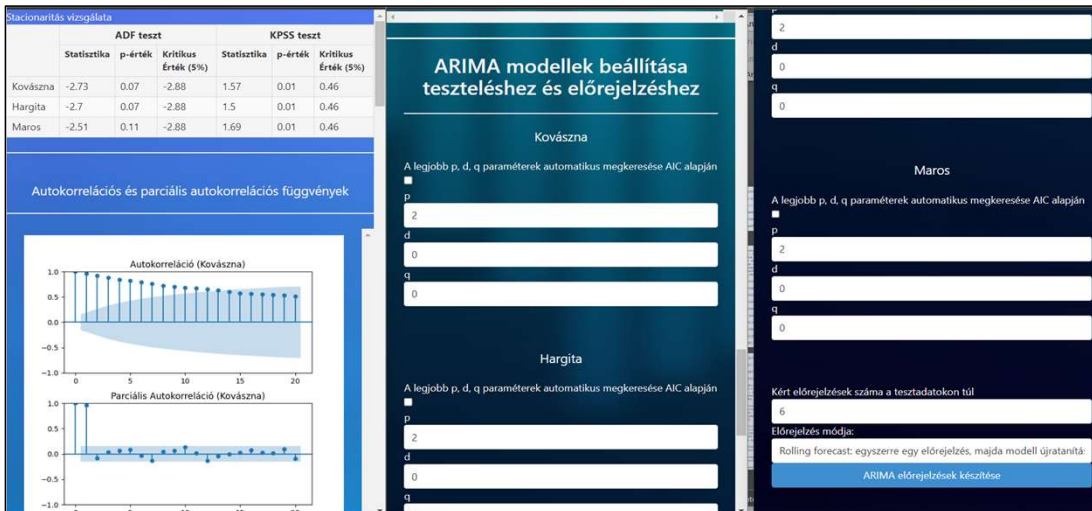


Forrás: saját szerkesztés Python-ben, az INSSE adatai alapján)

ARIMA modellek: illesztés

időszak	Eredeti időszor				Első differencia			
	ADF teszt		KPSS teszt		ADF teszt		KPSS teszt	
	Statistika	p-érték	Statistika	p-érték	Statistika	p-érték	Statistika	p-érték
	Kritikus érték (5% szig. szint): (-2.88)		Kritikus érték (5% szig. szint): (0.46)		Kritikus érték (5% szig. szint): (-2.88)		Kritikus érték (5% szig. szint): (0.46)	
Kovácsna	-2.73	0.07	1.57	0.01	-3.44	0.01	0.44	0.06
Hargita	-2.7	0.07	1.5	0.01	-2.79	0.06	0.22	0.1
Maros	-2.51	0.11	1.69	0.01	-6.63	0.0	0.12	0.1

■ nem stacionárius
 ■ nem stacionárius (de közel a határértékhez)
 ■ stacionárius

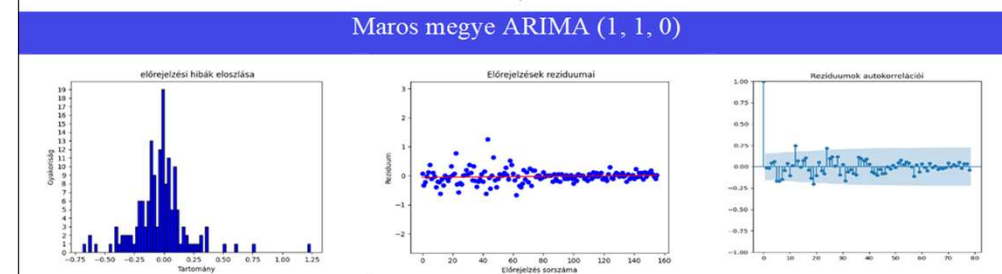
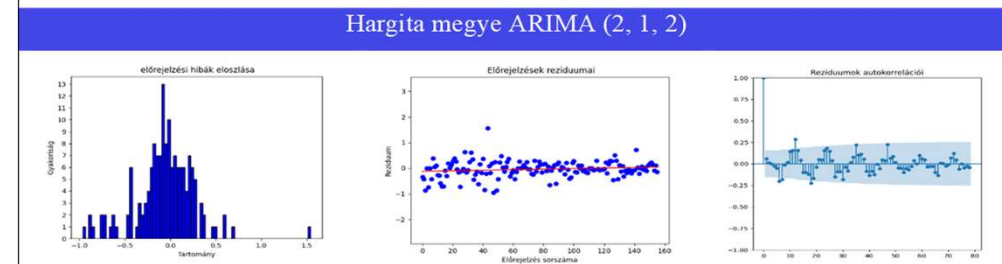
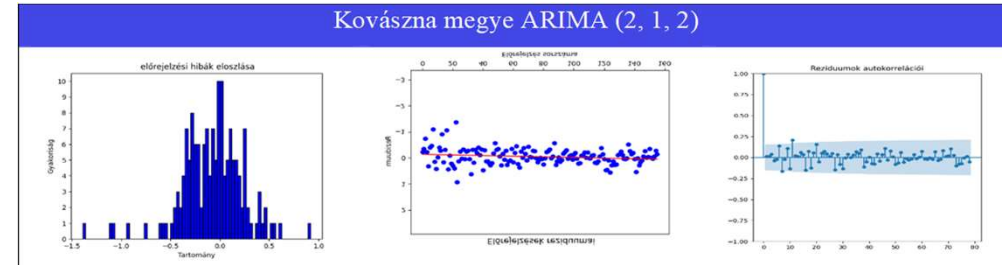


$$y_t = \alpha + \underbrace{\phi_1 y_{t-1} + \phi_2 y_{t-2}}_{\text{AR}(2) \text{ késleltetett értékek}} + \underbrace{\varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}}_{\text{MA}(2) \text{ késleltetett hibák}}$$

ϕ autoregresszív együtthatók
 θ mozgóátlag együtthatók
 ε_t hibatag (fehérzaj)

Sándor (2019)

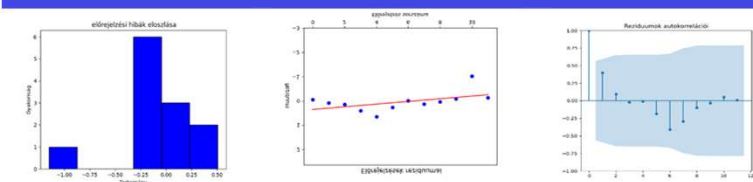
Az illesztett adatok becslése: hibamutatók



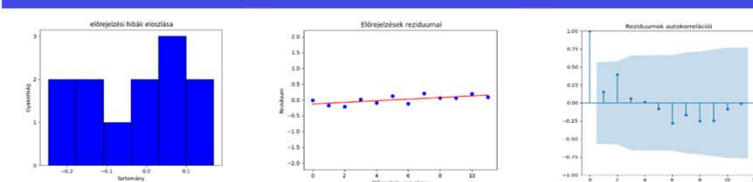
Megye	Modell	AIC	MSE	RRMSE	MAPE	R ²
Kovácsna	ARIMA (2, 1, 2)	91.51	0.1	0.05	3.97	0.98
Hargita	ARIMA (2, 1, 2)	102.17	0.11	0.06	4.23	0.96
Maros	ARIMA (1, 1, 0)	-10.9	0.05	0.05	3.24	0.98

A kiválasztott ARIMA modellek: tesztelés és előrejelzés

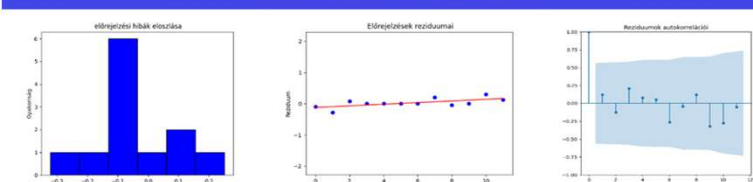
Kovácsna megye ARIMA (2, 1, 2)



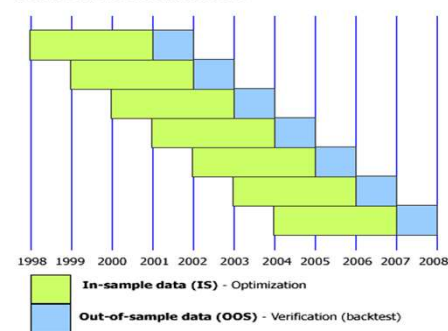
Hargita megye ARIMA (2, 1, 2)



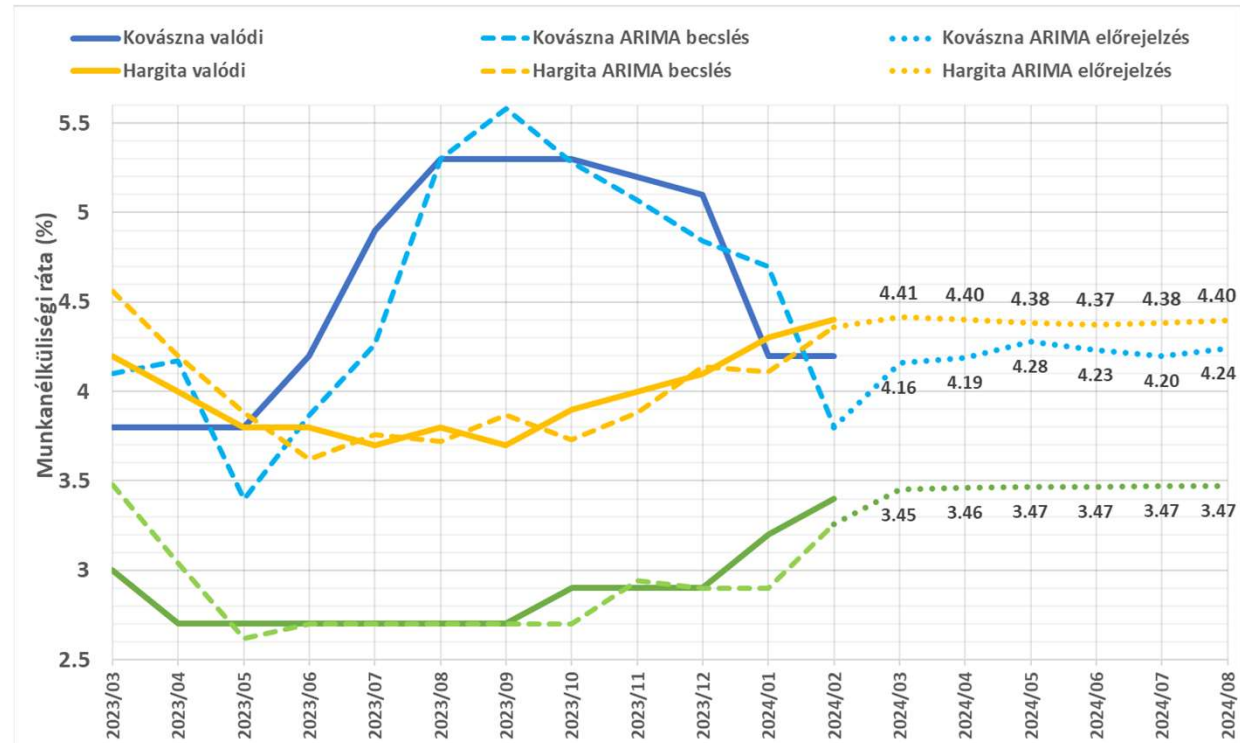
Maros megye ARIMA (1, 1, 0)



Walk-Forward Test procedure

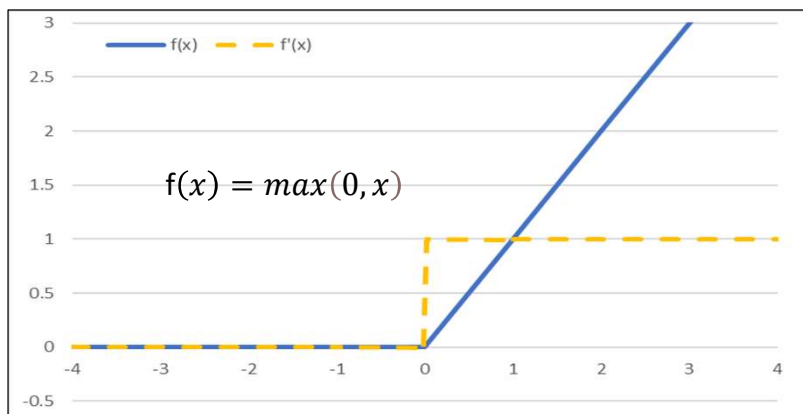
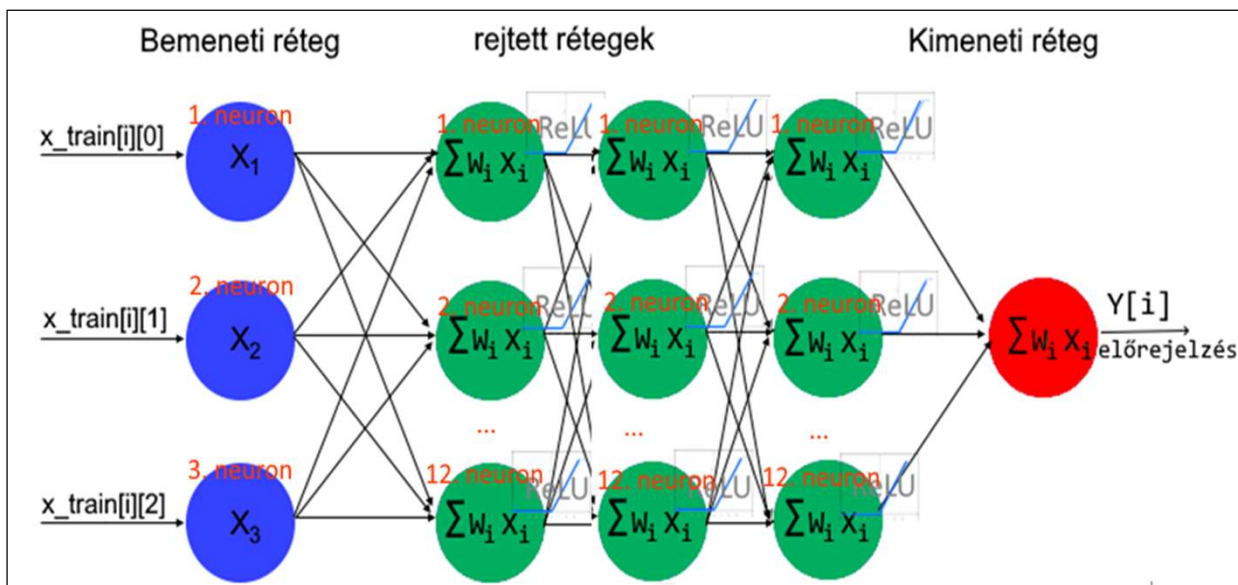


Walk-forward validáció



Megye	Modell	MSE	RRMSE	MAPE	R ²
Kovácsna	MLP (12, 12, 12)	0.06	0.056	4.18	0.83
Hargita	MLP (12, 12, 12)	0.01	0.02	1.92	0.85
Maros	MLP (12, 12, 12)	0.02	0.05	3.14	0.61

A kiválasztott MLP neuronháló szerkezetek



$$x' = \frac{x_i - \bar{x}}{\sigma}$$

- 3 rejtett réteg, 12-12-12 neuron
- Felügyelt tanítás
- MSE veszteségfüggvény
- LBFGS optimalizációs algoritmus
- ReLU aktivációs függvény
(Bamberger és társai, 2023)
- Max. 2000 epoch
(korai leállítás \rightarrow 500-600)
- Standardizált adatok

A felügyelt gépi tanulás adatbázisának elkészítése

Késleltetett értékek:

3

Normalizálás:

✓ Elvárt kimenetek (y_{train} és y_{test}) normalizálása, majd visszaalakítása

Standardizáció

Standardizáció

Min-Max

Robosztus

Logaritmizálás

Nincs

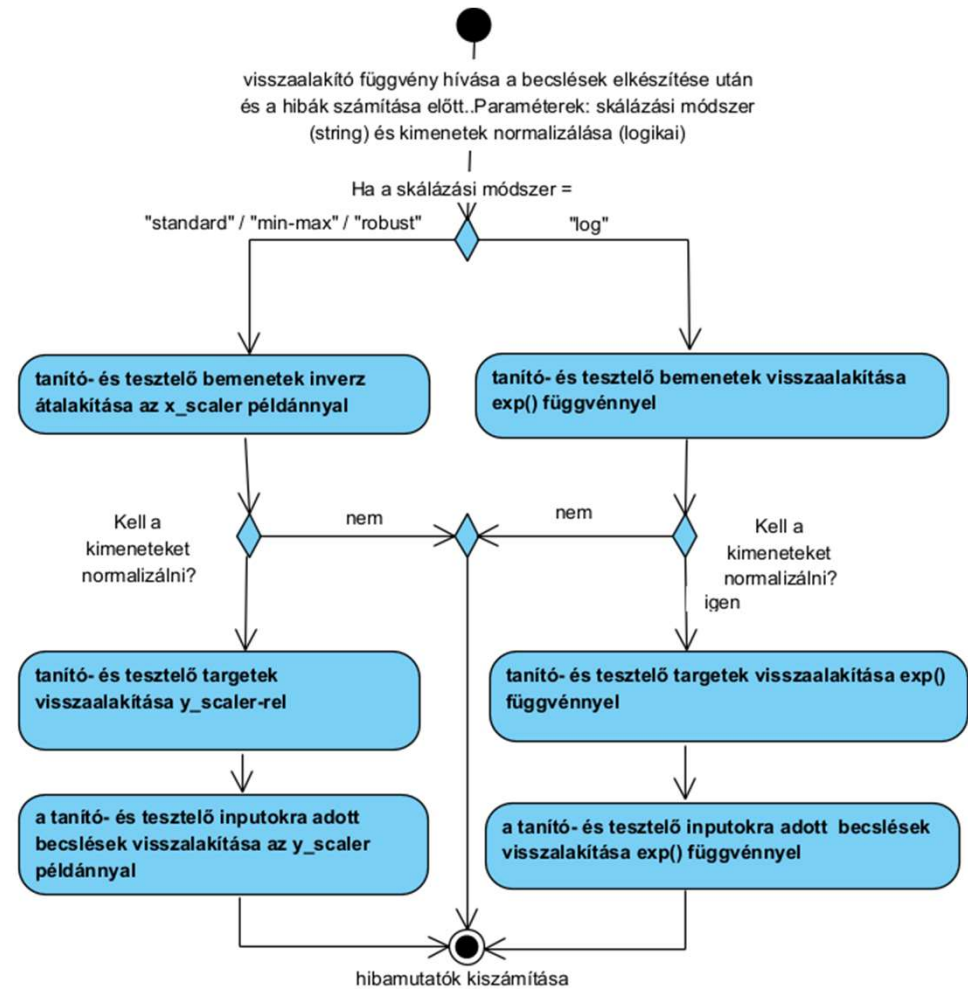
Az adatok normalizálása segít a modellnek könnyebben felismerni a mintázatokat és hamarabb konvergálni.

A standardizáció képlete:
$$Z = \frac{X - \mu}{\sigma}$$

A Min-Max skálázás képlete:
$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

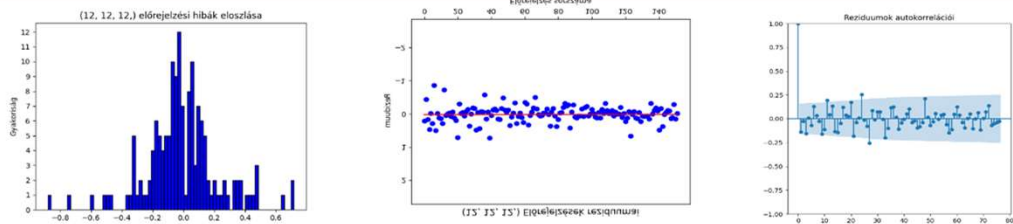
A Robust skálázás képlete:
$$X_{robust} = \frac{X - Q_1(X)}{Q_3(X) - Q_1(X)}$$

```
time_series = [8.3, 8.4, 8.5, 8.2, 7.9]
lag = 3
```

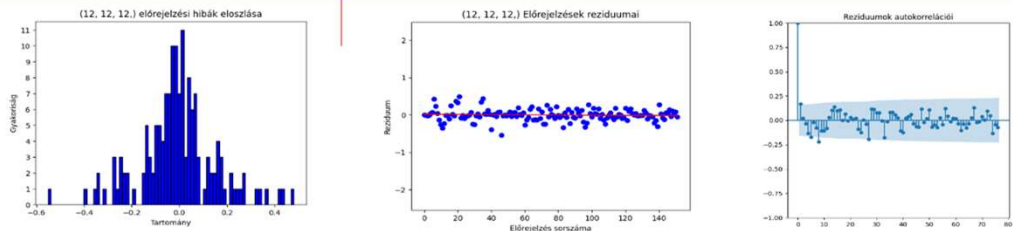


A kiválasztott MLP neuronhálók: tanulás

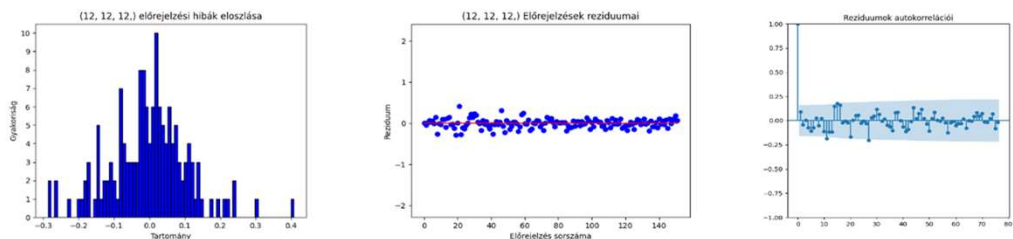
Kovácsna megye



Hargita megye



Maros megye



```
if(str.upper(solver) == "ADAM" or str.upper(solver) == "SGD"):
    n = max(n, 200)
    if(not earlyStop):
        n = self.max_iters
```

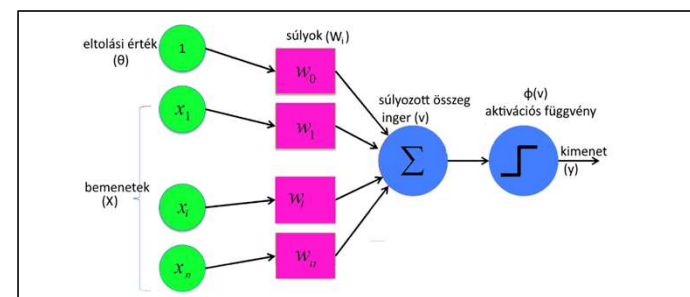
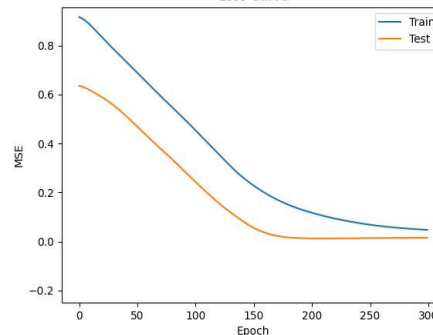
```
for i in range(int(n)):
    self.model = self.model.partial_fit(x_train, y_train)
    predictions = self.model.predict(x_train)
    train_loss = MSE(predictions, y_train)
    self.trainLossCurve.append(train_loss)
```

```
predictions = self.model.predict(self.x_test)
test_loss = MSE(predictions, self.y_test)
self.testLossCurve.append(test_loss)
```

```
else:
    self.model = self.model.fit(x_train, y_train)
```

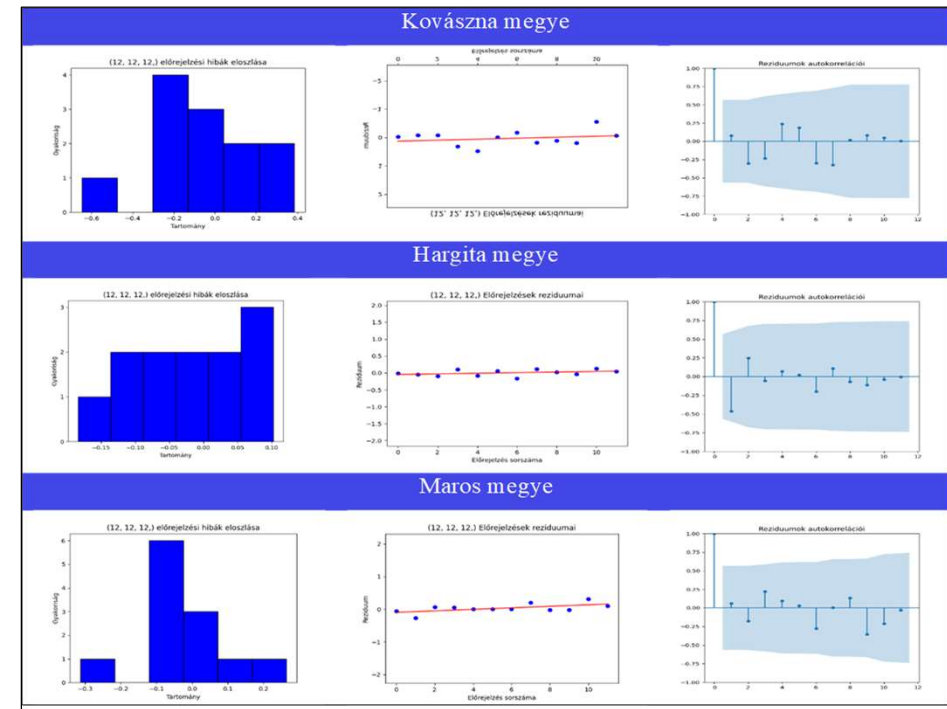
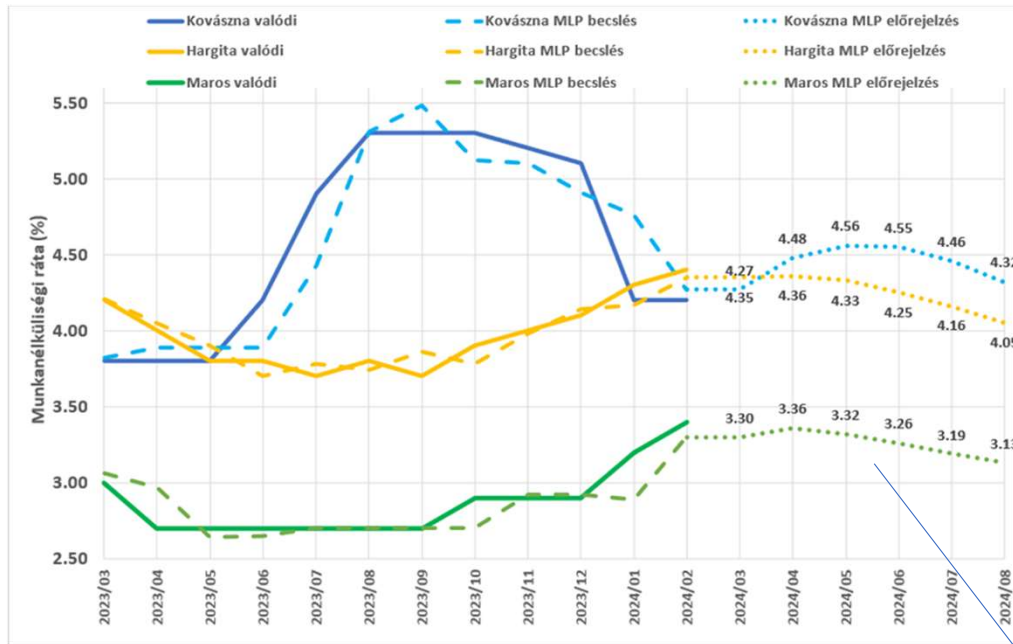
```
self.weights = [layer_weights for layer_weights in self.model.coefs_]
learning_pred = self.model.predict(x_train)
self.learning_pred = learning_pred
```

Loss Curve



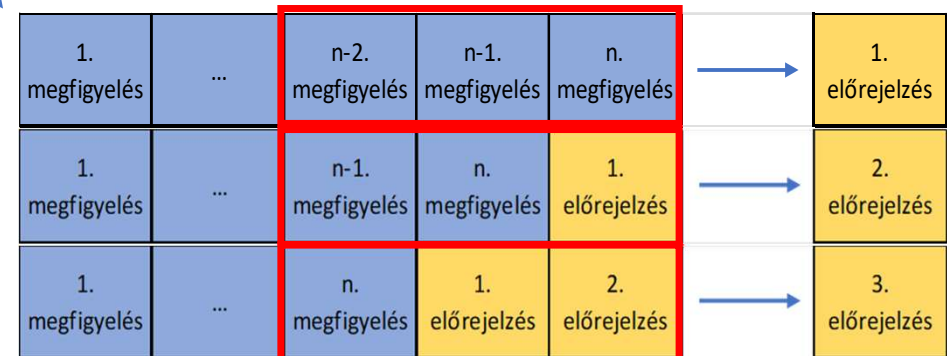
Megye	Modell	MSE	RRMSE	MAPE	R ²
Kovácsna	MLP (12, 12, 12)	0.06	0.04	3.01	0.99
Hargita	MLP (12, 12, 12)	0.03	0.03	2.22	0.99
Maros	MLP (12, 12, 12)	0.01	0.02	1.99	0.99

Az átalánosítás vizsgálata és rekurzív előrejelzések az MLP modellekkel



```
predictions = self.model.predict(self.x_test)
test_loss = MSE(predictions, self.y_test)
```

Megye	Modell	MSE	RRMSE	MAPE	R ²
Kovászna	MLP (12, 12, 12)	0.06	0.056	4.18	0.83
Hargita	MLP (12, 12, 12)	0.01	0.02	1.92	0.85
Maros	MLP (12, 12, 12)	0.02	0.05	3.14	0.61



Következtetések

- ✓ Ajoodha és Mulaudzi (2020); Tufaner és Sözen (2021):
az **MLP** modellek **jobban teljesítettek** a munkanélküliségi ráta előrejelzésében, mint az **ARIMA** modellek.
- ✓ Sasu (2013):
az **MLP** **jobban** általánosított a munkanélküliségi ráták idősorán, mint az **ARIMA**
- ✓ LBFGS, ReLU mellett teljesítettek a legjobban az MLP modellek
(Paszkiewicz és társai, 2023; Pardede és társai, 2022)
- ✓ Python-ben könnyű ANN és ARIMA modelleket készíteni, valamint jól lehet vizualizálni az adatokat
egy **felhasználóbarát webes felület** segítségével → **további kutatásokhoz is alkalmas**
- ❖ Az MLP szerint mindhárom megye esetében emelkedés várható a nyár elejéig, majd a nyár végig fokozatos csökkenés.
- ❖ Továbbfejlesztési lehetőség: felhasználói fiókok, idősorok, modellek és előrejelzések tárolása (MongoDB)

KÖSZÖNÖM A FIGYELMET!