

Implementacija algoritama za predikciju vremenskih serija

Pregled rada

U okviru ovog projektnog zadatka analiziran je i praktično implementiran algoritam opisan u radu Forecasting time series with complex seasonal patterns using exponential smoothing, čiji su autori Alysha M. De Livera, Rob J. Hyndman i Ralph D. Snyder [1]. Rad je objavljen 2011. godine i predstavlja značajan doprinos oblasti predikcije vremenskih serija, posebno u slučajevima kada podaci posjeduju složene i dugotrajne sezonske obrasce. Autori u radu uvode dva nova modela pod nazivom BATS i TBATS kao proširenje klasičnih metoda eksponencijalnog izgladivanja, sa ciljem boljeg modeliranja realnih vremenskih serija kod kojih standardni pristupi pokazuju ograničenja. Sami rad pored što je dao teoretski opis dva nova modela, je predstavio i rezultate tih modela evaluirani na tri različita dataseta. U sklopu ovog projektnog zadatka odabrali smo reproducirati rezultate TBATS modela nad datasetom sedmičnih podataka o potrošnji benzina (u barelima) u Sjedinjenim Američkim Državama u periodu od Februara 1991 do Jula 2005.

Opis TBATS

Algoritam predstavljen u odabranom radu poznat je pod nazivom TBATS, što označava kombinaciju trigonometrijskog modeliranja sezonalnosti, Box–Cox transformacije, ARMA grešaka, trenda i sezonskih komponenti. Osnovna ideja modela zasniva se na proširenju klasičnih ETS modela tako da mogu efikasno obuhvatiti više sezonskih perioda i sezonalnosti velikog reda.

Sezonske komponente u TBATS modelu predstavljene su pomoću trigonometrijskih funkcija, odnosno Fourierovih redova, čime se omogućava fleksibilno modeliranje dugih sezonskih ciklusa bez eksponencijalnog rasta broja parametara. Dodatno, primjena Box–Cox transformacije omogućava stabilizaciju varijanse vremenske serije, što doprinosi robusnosti modela u prisustvu heteroskedastičnosti. Nakon uklanjanja trenda i sezonalnosti, preostala struktura u rezidualima modelira se ARMA procesom, čime se dodatno poboljšava kvalitet predikcije. Formalno model se opisuje sljedećim jednačinama.

$$\begin{aligned} y_t^{(\lambda)} &= \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^S \sum_{k=1}^{K_i} \left[s_{k,t-1}^{(i)} \cos\left(\frac{2\pi k}{m_i}\right) + s_{k,t-1}^{*(i)} \sin\left(\frac{2\pi k}{m_i}\right) \right] \\ &\quad + \sum_{j=1}^p \varphi_j d_{t-j} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \\ \varepsilon_t &\sim \mathcal{N}(0, \sigma^2), \quad y_t^{(\lambda)} = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \log(y_t), & \lambda = 0. \end{cases} \end{aligned} \tag{1}$$

gdje:

- y_t observacija u vremenu t ,
- $y_t^{(\lambda)}$ je Box-Cox transformirana serija sa parametrom λ ,

- ℓ_t predstavlja nivo,
- b_t predstavlja trend,
- $\phi \in [0, 1]$ je *damping* trenda,
- S broj sezonalnih komponenti,
- m_i period i -te sezonalne komponente,
- K_i broj harmonika korišteni za m_i ,
- $s_{k,t}^{(i)}$ i $s_{k,t}^{*(i)}$ su trigonometrijske sezonalne komponente stanja,
- ε_t
- d_t je ARMA proces greške,
- φ_j and θ_j su autoregresivni koeficijenti pomičnog prosjeka,
- p i q su AR i MA redovi,
- σ^2 varijansa.

Zahvaljujući ovoj kombinaciji komponenti, TBATS model je naročito pogodan za vremenske serije koje karakterišu složeni sezonski obrasci, kao što su energetske, ekonomske i transportne podatke.

Opis skupa podataka

Korišten je skup podataka "*US Finished Motor Gasoline Product Supplied*", koji je javno dostupan u R paketu `forecast`.

- **Izvor:** US Energy Information Administration (preuzet sa `robjhyndman.com` [2] što je stranica jednog od kolaboratora na originalnom radu).
- **Frekvencija:** Sedmična (Weekly).
- **Period analize:** Od februara 1991. do jula 2005. godine.
- **Ukupan broj uzoraka:** 745 sedmica.

Podjela podataka je izvršena na identičan način kao i u originalnom radu i to na trening skup koji se sastoji od prve 484 sedmice i testni skup koji se sastoji od preostalih 261 sedmica. Podaci pokazuju snažan uzlazni trend i jasnu godišnju sezonalnost koja evoluirala kroz vrijeme.

Implementacija

Replikacija je izvršena u programskom jeziku **Python**, koristeći biblioteku `tbats` koja predstavlja port originalnog R paketa. Kod je dostupan javno na repozitoriju [3].

1. **Priprema okruženja:** Zbog nekompatibilnosti novijih verzija biblioteke `scikit-learn` sa bibliotekom `tbats`, bilo je potrebno implementirati prepravku za funkciju `check_array`, kako bi se omogućilo pokretanje koda.
2. **Konfiguracija modela:** Originalni rad navodi period sezonalnosti $m_1 \approx 52.179$. U Pythonu je ovo definisano parametrom:

```
estimator = TBATS(seasonal_periods=[365.25/7], ...)
```

3. **Ograničenja optimizacije:** Inicijalna pokretanja algoritma u Pythonu rezultirala su Box-Cox parametrom $\lambda \approx 0.66$, što se značajno razlikovalo od rezultata u radu ($\lambda \approx 0.992$).

Ispod je prikazan ključni dio koda za treniranje modela:

```
# Inicijalizacija TBATS modela sa parametrima iz rada
estimator = TBATS(
    seasonal_periods=[365.25/7],
    use_box_cox=False,          # Forsiranje Lambda ~ 1 (kao u radu)
    use_trend=True,            # Aditivni trend
    use_damped_trend=False,
    use_arma_errors=True,      # Automatska detekcija ARMA(p,q)
    use_parallel=False
)

# Treniranje na prvih 484 uzoraka
model = estimator.fit(y_train)

# Predikcija za narednu 261 sedmicu
y_forecast, conf_int = model.forecast(steps=261)
```

1 Rezultati evaluacije

Evaluacija je izvršena poređenjem procijenjenih koeficijenata i korijena srednje kvadratne greške (RMSE) na testnom skupu.

Tabela 1 prikazuje poređenje parametara dobijenih u Python implementaciji sa onima navedenim u originalnom radu.

Parametar	Originalni Rad (R)	Naša Implementacija (Python)
Box-Cox (λ)	0.9922	0.659921
Alpha (α)	0.0478	≈ 0.093942
Beta (β)	≈ 0.00	0.0047
ARMA Model	MA(1)	MA(1)
Theta (θ_1)	-0.2124	≈ -0.2581

Tabela 1: Poređenje parametara modela

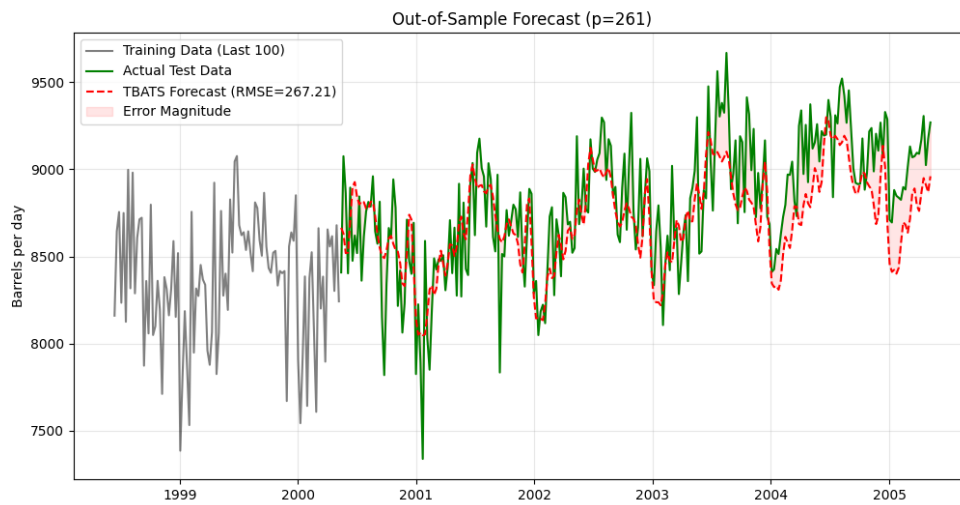
Primjećujemo da je struktura modela (MA(1) greške, mali beta koeficijent) uspješno replicirana. Međutim, vrijednost α (smoothing level) je veća u našoj implementaciji, što znači da je naš model osjetljiviji na šum u podacima.

Izračunata je RMSE vrijednost na testnom skupu (Out-of-sample) koristeći formulu:

$$RMSE = \sqrt{\frac{1}{h} \sum_{j=1}^h (y_{n+j} - \hat{y}_{n+j})^2}$$

RMSE je značajno veći od onog koji je dobiven u radu (grafički prikaz pokazuje odstupanja u amplitudi sezonalnosti). Uzrok odstupanja je to što iako biblioteka `tbats` u Pythonu prepisana iz R paketa, ona koristi drugačiji optimizacijski engine (`scipy.optimize` vs R `optimizer`). Pronalaženje optimalnih

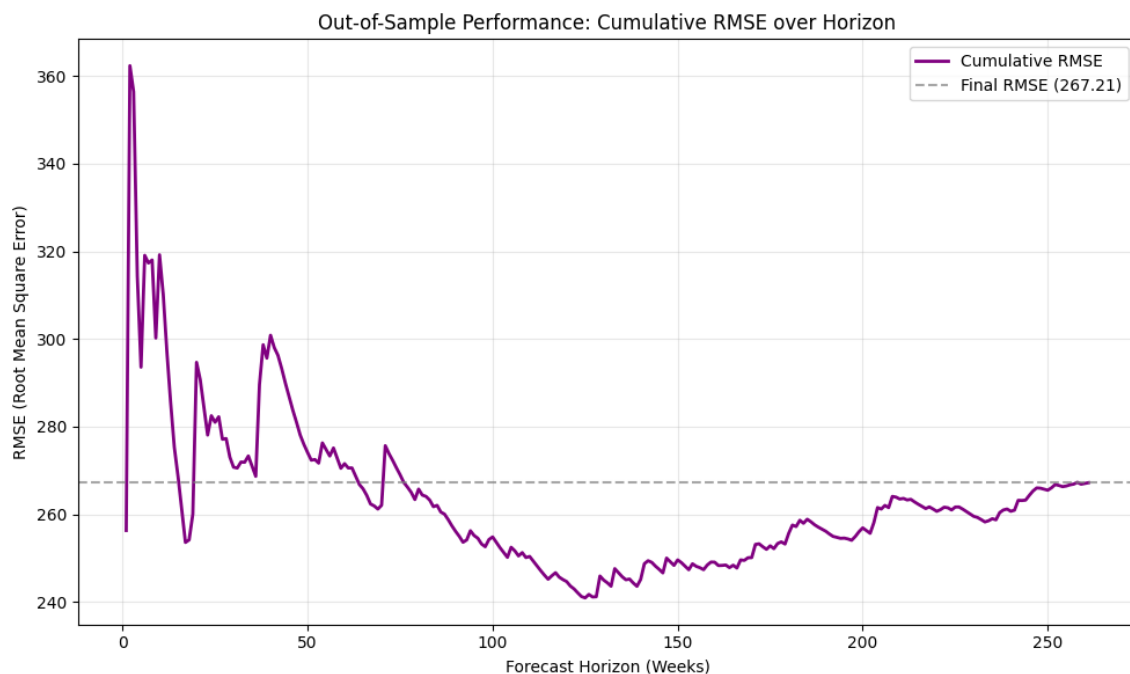
parametara TBATS modela je problem nekonveksne optimizacije. Python implementacija je konvergirala u lokalni minimum koji nije optimalan za ovaj specifičan skup podataka, što je rezultiralo lošijim predikcijama amplituda sezonskih peakova u odnosu na originalni rad.



Slika 1: Prikaz predikcije vs stvarni podaci

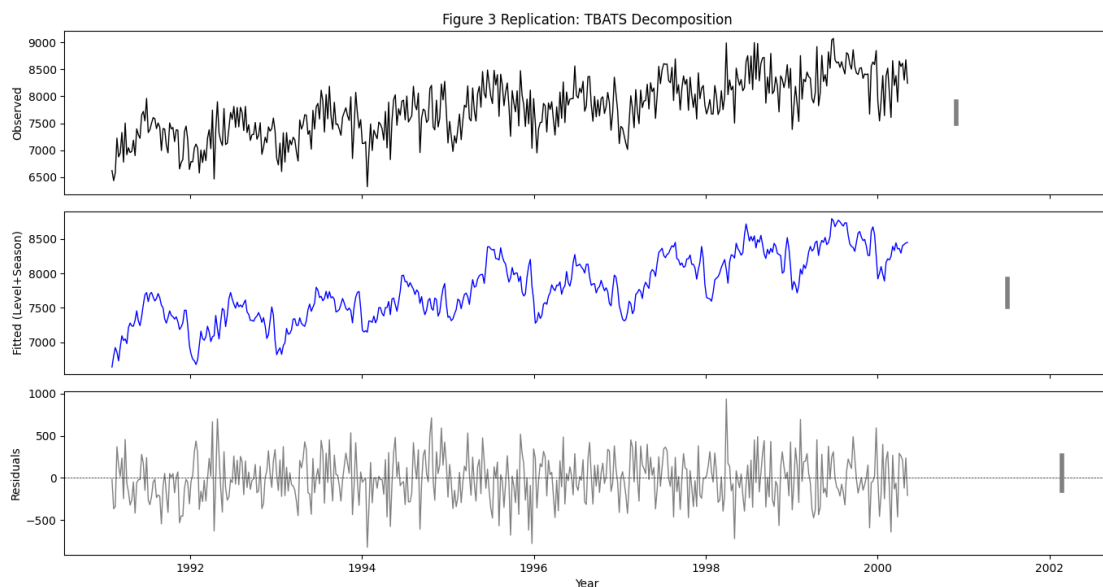
Slika 1 prikazuje predikciju modela u odnosu na stvarne podatke. Vidljivo je da model ispravno hvata fazu sezonalnosti i trend, ali ne uspijeva u potpunosti replicirati magnitudu ekstremnih vrijednosti u ljetnim mjesecima, što dovodi do veće RMSE greške.

Slika 2 prikazuje RMSE grešku sa prolazom kroz testne (out-of-sample) podatke.



Slika 2: Out-of-sample rezultati kako se RMSE mijenja

Trigonometrijska dekompozicija dataseta, dobivena iz pronađenog TBATS modela je prikazana na slici 3. Vertikalni podioci s desne strane predstavljaju jednake visine prikazane drugačijom skalom



Slika 3: Trigonometrijska dekompozicija podataka koristeći model

Zaključak

U okviru ovog projektnog zadatka analiziran je i implementiran TBATS model za predikciju vremenskih serija sa složenim sezonskim obrascima, kako je predstavljen u radu De Livera, Hyndman i Snyder (2011) [1]. Cilj rada bio je reproducirati rezultate originalnog rada nad javno dostupnim skupom sedmičnih podataka o potrošnji benzina u Sjedinjenim Američkim Državama, te ispitati ponašanje modela u drugačijem programskom okruženju. Rezultati su pokazali da je osnovna struktura TBATS modela uspješno replicirana. Identifikovani su isti tip ARMA greške (MA(1)), veoma mali koeficijent trenda, te pravilno modelirana sezonalna struktura sa godišnjim periodom. Model je u stanju da pouzdano uhvati dugoročni trend i fazu sezonalnosti, što potvrđuje njegovu pogodnost za analizu vremenskih serija sa dugim i promjenjivim sezonskim ciklusima. Međutim, uočena su i odstupanja u vrijednostima pojedinih parametara, posebno Box–Cox parametra i koeficijenta izgladivanja nivoa, kao i veća RMSE greška u odnosu na originalni rad. Analizom je utvrđeno da su ova odstupanja posljedica razlika u optimizacijskom procesu između R i Python implementacija. Budući da se procjena parametara TBATS modela svodi na nekonveksan optimizacijski problem, različiti optimizatori mogu konvergirati ka različitim lokalnim minimumima, što direktno utiče na kvalitet predikcije.

References

- [1] A. M. D. Livera, R. J. Hyndman, and R. D. Snyder, “Forecasting time series with complex seasonal patterns using exponential smoothing,” *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1513–1527, 2011. [Online]. Available: <https://doi.org/10.1198/jasa.2011.tm09771>
- [2] R. Handyman, “Gasoline dataset,” 2026. [Online]. Available: <https://robjhyndman.com/data/gasoline.csv>
- [3] T. Hastor, “Github - lilhast1/tbats,” 2025. [Online]. Available: <https://github.com/lilhast1/tbats>